NATIONAL CENTER FOR EDUCATION STATISTICS

Working Paper Series

The Working Paper Series was initiated to promote the sharing of the valuable work experience and knowledge reflected in these preliminary reports. These reports are viewed as works in progress, and have not undergone a rigorous review for consistency with NCES Statistical Standards prior to inclusion in the Working Paper Series.

U. S. Department of Education
Office of Educational Research and Improvement

NATIONAL CENTER FOR EDUCATION STATISTICS

Working Paper Series

A Study of Imputation Algorithms

Working Paper No. 2001–17 September 2001

Contact: Ralph Lee
Statistical Standards Program
E-mail: ralph.lee@ed.gov

U. S. Department of Education
Office of Educational Research and Improvement

U.S. Department of Education
Rod Paige
Secretary

Office of Educational Research and Improvement
Grover J. Whitehurst
Assistant Secretary

National Center for Education Statistics
Gary W. Phillips
Acting Commissioner

The National Center for Education Statistics (NCES) is the primary federal entity for collecting, analyzing, and reporting data related to education in the United States and other nations. It fulfills a congressional mandate to collect, collate, analyze, and report full and complete statistics on the condition of education in the United States; conduct and publish reports and specialized analyses of the meaning and significance of such statistics; assist state and local education agencies in improving their statistical systems; and review and report on education activities in foreign countries.

NCES activities are designed to address high priority education data needs; provide consistent, reliable, complete, and accurate indicators of education status and trends; and report timely, useful, and high quality data to the U.S. Department of Education, the Congress, the states, other education policymakers, practitioners, data users, and the general public.

We strive to make our products available in a variety of formats and in language that is appropriate to a variety of audiences. You, as our customer, are the best judge of our success in communicating information effectively. If you have any comments or suggestions about this or any other NCES product or report, we would like to hear from you. Please direct your comments to:

National Center for Education Statistics
Office of Educational Research and Improvement
U.S. Department of Education
1990 K Street NW
Washington, DC 20006

September 2001

The NCES World Wide Web Home Page is
http://nces.ed.gov

Suggested Citation

U.S. Department of Education. National Center for Education Statistics. *A Study of Imputation Algorithms*. Working Paper No. 2001–17, by Ming-xiu Hu and Sameena Salvucci. Project Officer, Ralph Lee. Washington, DC: 2001.

Foreword

In addition to official NCES publications, NCES staff and individuals commissioned by NCES produce preliminary research reports that include analyses of survey results, and presentations of technical, methodological, and statistical evaluation issues.

The ***Working Paper Series*** was initiated to promote the sharing of the valuable work experience and knowledge reflected in these preliminary reports. These reports are viewed as works in progress, and have not undergone a rigorous review for consistency with NCES Statistical Standards prior to inclusion in the Working Paper Series.

Copies of Working Papers can be downloaded as pdf files from the NCES Electronic Catalog (http://nces.ed.gov/pubsearch/), or contact Sheilah Jupiter at (202) 502–7444, e-mail: sheilah_jupiter@ed.gov, or mail: U.S. Department of Education, Office of Educational Research and Improvement, National Center for Education Statistics, 1990 K Street NW, Room 9048, Washington, DC 20006.

Marilyn M. Seastrom
Chief Mathematical Statistician
Statistical Standards Program

Ralph Lee
Mathematical Statistician
Statistical Standards Program

This page intentionally left blank.

A Study of Imputation Algorithms

Prepared by:

Ming-xiu Hu
Sameena Salvucci
Synectics for Management Decisions, Inc.

Prepared for:

U.S. Department of Education
Office of Educational Research and Improvement
National Center for Education Statistics

September 2001

(originally delivered 1998)

This page intentionally left blank.

Table of Contents

List of Tables

Introduction

No matter how well a survey questionnaire is designed and no matter how efficient a data collection procedure is employed, missing values almost always exist in survey data. There are two main reasons for missing values, *survey (or unit) nonresponse* and *item nonresponse*. Examples of survey nonresponse include when sampled subjects are unable to be contacted; when sampled subjects refuse to respond altogether; when sampled subjects are found to be out-of-scope. Examples of item nonresponse include when sampled subjects refuse to answer certain questions; when sampled subjects are unable to answer certain questions; when interviewers fail to ask the question or fail to record the answer; when an inconsistent response is deleted in data editing.

One of the most common methods to compensate for survey nonresponse is through weighting adjustments; that is, to reassign the weights of the nonrespondents to the respondents. However, there are some problems with the use of weighting adjustments for dealing with unit nonresponse (Rubin 1996):

- Even in the simplest case of unit nonresponse, where the shared data base of respondents is fully observed (i.e, there is no item nonresponse), many ultimate users' complete-data analyses do not allow for sampling weights.

- Even with complete-data analyses that can deal with sampling weights, the construction of intervals and p-values that validly account for the fact that nonresponse adjustments in the weights are estimated from data are not immediate from complete-data analyses.

- With general patterns of nonresponse, special analysis methods need to be developed and special software needs to be written.

- Weighting adjustments are focused on unbiased estimation and are essentially blind to efficiency concerns.

Given these problems with using weighting adjustments, *imputation* has become one of the most popular tools used to solve missing value problems in survey data analyses. The use of imputation to create complete data can have the following advantages:

- Data collectors usually have more inside knowledge about the reasons for the missing values. This inside knowledge can be used in imputation;

- Missing values complicate the data structure, so that more sophisticated statistical tools are required to conduct analyses. Imputation may ease this difficulty;

- Imputation can prevent the loss of information due to deletion of incomplete records if the statistical methods used (e.g., regression) require complete records;

- Imputation can reduce nonresponse bias in some situations;
- Pairwise correlation matrices computed from incomplete data may not be positive definite. Imputation can avoid this problem.

The basic objective of imputation 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 such as "complete-case analysis," "available-case analysis," and "fill-in with means", satisfy this basic objective and so have a certain appeal. However, it is certainly not enough to just achieve this basic objective. Another desirable objective is statistical validity: assuming that the ultimate user's complete-data analysis is statistically valid for a scientific estimand, the answer that results from applying the same analysis method to an incomplete-data remains statistically valid for the same scientific estimand assuming the truth of the database constructor's posited model for missing data. This goal can be achieved through some imputation methods, but cannot be achieved through others.

It is probably a popular misunderstanding 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. Here is a counterexample given by Rubin (1996): suppose we have a coin that, in truth, is biased .6 heads and .4 tails. This known truth is model A, whereas model B asserts that the coin has two heads. Using model A for creating imputations (i.e., future predictions) yields a hit rate (agreements between predictions and outcomes) of $.6 \times .6 + .4 \times .4 = .52$, whereas using model B for predictions yields a hit rate of .6. This does not mean that model B is better than model A for handling missing values. Filling in missing values using model B yields the invalid statistical inference that in the future all coin tosses will be heads, clearly inconsistent for the estimand Q = fraction of tosses that are heads, whereas using model A yields consistent estimates for all such scientific estimands.

Many imputation techniques and imputation software packages have been developed over the years. Different methods may work well under different circumstances. It is advisable to conduct a sensitivity analysis when choosing an imputation method for a particular survey.

This task reviewed about thirty imputation methods and five imputation software packages. Eleven of the most popular imputation methods were evaluated through a Monte Carlo simulation study.

This report consists of five chapters. The first four chapters are on methodology discussions based on our review of numerous papers and books. Chapter 1 describes about thirty most commonly used imputation methods with brief discussions of their strengths and weaknesses. The imputation methods used across the national surveys conducted by the National Center for Education Statistics (NCES) are also summarized in this chapter. Chapter 2 discusses five imputation software packages. Nonresponse bias correction via imputation is addressed in chapter 3. Variance estimation with imputed data and multiple imputation inference is discussed in chapter 4. Chapter 5 reports the results of the simulation study, which evaluates 11 imputation methods according to eight evaluation criteria for four types of distributions, five types of missing mechanisms and four types of missing rates.

Chapter 1 Imputation Algorithms

Imputation methods are generally classified into two categories: *random* (also called *stochastic*) and *deterministic*. A deterministic imputation method determines one and only one possible value for imputing each missing case. Once the imputation scheme is set up, the imputation result is unique. On the other hand, a random imputation method draws imputation values randomly either from the observed data or from the predicted distribution. Multiple sets of imputations can be created to capture the uncertainty between imputations via any random imputation method. Generally, a random imputation method adds more variability to the statistics computed from an imputed data set than a deterministic imputation method.

However, in this chapter, we will discuss imputation techniques under five categories:

- Simple deterministic imputation
- Simple random imputation
- Model-based deterministic imputation
- Model-based random imputation
- Bayesian-related imputation methods

It is easy to see that these five categories are not mutually exclusive; we are using them mainly for convenience of discussion.

1.1 Simple deterministic imputation method

1.1.1 Deductive imputation

This method deduces missing values from available information, such as similar items in previous surveys, related items in current surveys, etc. To apply this method, the user needs to find some deterministic relationship between the missing item and items from other resources. Cold deck is one deductive imputation method that uses information from previous similar surveys. Generally, it is impossible to find enough information to impute all missing items in a survey using deductive imputation, but this method can be used to impute some of the missing variables. Whenever possible, deductive imputation should be used before any other imputation method because it provides accurate or approximately accurate imputations for missing cases. However, the performance of a deductive imputation method completely depends on the available sources.

1.1.2 Overall or cell mean imputation (also called adjusted mean imputation or substitution method)

This is the simplest but least attractive imputation method. Overall mean imputation uses the overall sample mean to replace all missing values in the data set. This method can provide unbiased estimates for the population means or totals only if the missing values are missing completely at random (MCAR). Cell mean imputation first uses some auxiliary variables to form imputation cells, and then replaces missing values in each cell with its sample mean. The method can give unbiased estimates for the population mean or total if the missing values only depend on the auxiliary variables which are used to construct the imputation cells. However, the distribution of the data will be distorted substantially and the concentration of all imputed values at the cell means creates spikes in the distribution. Therefore, quartile estimates will be biased, and the variances materially underestimated.

If the mean imputation method is used, it is advisable to calculate the variance-covariance estimates using a denominator of *n-m-1* instead of *n-1*, where *n* is the sample size and *m* is the number of cases missing one or both variables for pairwise covariance estimate calculation. We will call this strategy *the adjusted mean imputation* (or *substitution*) *method* in this report.

Cohen (1996) suggested another way to adjust variance estimates by imputing more diversified values for the missing cases. For example, instead of imputing the mean for all the missing values, Cohen suggested imputing half of the missing values with $\bar{y}_r + \sqrt{\frac{n+r-1}{r-1}} D_r$ and the other half with $\bar{y}_r - \sqrt{\frac{n+r-1}{r-1}} D_r$, where r is the number of response values, $\bar{y}_r$ is the mean of observed values, and $D_r^2 = \frac{1}{r}\sum_1^r (y_i - \bar{y}_r)^2$. This type of adjustment will retain the first and second moments as observed.

1.1.3 Deterministic hot deck imputation

Hot deck imputation is one of the most popular imputation methods because it is simple and intuitively makes sense to many practitioners who do not have a strong statistical background. Hot deck imputation does not employ any explicit statistical model. Its major disadvantage is that it can not recover typical values for objects with certain characteristics if no such subject responds to a survey. Hot deck imputation employs many methods. The following are the most popular deterministic hot deck imputation methods.

(1) *Sequential nearest neighbor hot deck imputation*. This method is also called *traditional hot deck imputation*. The first step in this method is to use some auxiliary variables to specify imputation classes. Second, within each imputation class, a single value such as the class mean or some pre-specified value is assigned as a starting point. Then the records in the data file are treated sequentially. If a record has a response for the target variable, that value replaces the previously stored value for its imputation class. If a record has a missing value for the target variable, it is assigned the value currently stored for its imputation class.

A major attraction of this method is its computing economy, since all imputations are made in a single pass through the data file. A disadvantage is that this method may easily give rise to multiple use of donors, a feature which leads to a loss of precision for survey estimators (Kalton and Kasprzyk 1982).

(2) *Multivariate matching*. In this method, donors and donees are matched on several predetermined auxiliary variables. For each missing case in each matched class, the nearest donor is chosen for imputation. If no donor is found in a matched class, the class is combined with other classes to obtain donors.

While this method is not convenient to implement using computer programs, an approximately equivalent imputation algorithm may be used to replace it. The algorithm first sorts the data file with the same auxiliary variables, and then imputes the nearest response value for each missing case. This alternative method is very easy to implement. The donor and donee will match on all auxiliary variables if such donors are available. Otherwise, it will automatically find a donor matched on some of the auxiliary variables, which is equivalent to collapsing the matched classes.

(3) *Distance function matching*. This method imputes the nearest response value for each missing case according to some univariate distance function of auxiliary variables, such as the norm in the multi-dimensional Euclidean space, Mahalanobis distance, the difference between the predicted values from a regression model, etc.

1.2 Simple random imputation methods

1.2.1 Overall or cell mean imputation with random disturbance

To overcome the underestimated variance typical of the mean imputation method (see section 1.1.2), we may add a small disturbance drawn from a distribution with a mean zero and variance-covariance matrix equal to the observed variance-covariance matrix. Most often a normal distribution is used to draw the random disturbance.

1.2.2 Random hot deck method

Random hot deck imputation is one of the most popular methods in practice. It generally consists of three steps: (1) determine auxiliary variables on which donors and donees will match; (2) randomly draw imputations from observed data according to the observed frequency (weighted or unweighted) within each matched class; (3) if a matched class does not have any observed value, combine that class with other classes and perform imputation based on the combined imputation classes.

1.2.3 Overall random imputation

Overall random imputation generally refers to drawing imputation values randomly from observed data using different sampling schemes. The most frequently used scheme is resampling with or without replacement. It is one of the easiest methods to implement, because it does not use any auxiliary variables and will not be able to reduce nonresponse biases.

1.2.4 Approximate Bayesian Bootstrap (ABB)

The ABB method first randomly draws *r* values with replacement from the *r* observed values $Y_1,\ldots,Y_r$ to create Y_{obs}^*, and then randomly draws *m* values with replacement from Y_{obs}^* as imputed values for the *m* missing values in the target variable *Y*. The ABB method draws imputations from a resample of the observed data instead of drawing directly from the observed data. This extra step introduces additional variation, which makes the ABB method approximately "proper" for multiple imputation according to Rubin's theory (1987). (This method is called approximately Bayesian Bootstrap because it is approximately equivalent to the Bayesian Bootstrap described below.)

Similarly to the overall random imputation method, when ABB imputation is performed for the overall sample, it will not be able to reduce nonresponse biases because it does not use any auxiliary information. ABB imputation may work well for within-class imputations if the missing mechanism only depends on the variables used to construct the imputation classes.

1.2.5 Bayesian Bootstrap (BB)

BB imputation consists of two steps: (1) draw *r*-1 uniform random numbers between 0 and 1, and let their ordered values be $a_1,\ldots,a_{r-1}$; also let a_0=0 and a_r=1, where *r* is the number of observed values; (2) draw each of the *m* missing values from $Y_1,\ldots,Y_r$ with probabilities $(a_1-a_0),(a_2-a_1),\ldots,(1-a_{r-1})$; that is, independently *m* times, draw a uniform random number *u*, and impute Y_i if $a_{i-1} < u \le a_i$ (*i*=1, 2, …, *r*).

Rubin (1981) showed that the Bayesian Bootstrap is equivalent to assuming that the prior distribution of p is the (improper) distribution

$$\Pr(p) = \prod_{k=1}^{K} p_k^{-1} ,$$

where $\mathrm{p} = (p_1,\ldots,p_K)$ is the vector of probabilities $\Pr(Y_i = d_k) = p_k , \sum p_k = 1$ and $d_1,\ldots,d_K$ are all possible distinct values in $Y_1,\ldots,Y_r$. The posterior distribution of p is

$$\Pr(p \mid Y_{obs}) \propto \prod_{k=1}^{K} p_k^{r_k - 1} ,$$

where r_k is the number of y_i that equals d_k, and $\sum_{k=1}^{K} r_k = r$. The posterior distribution is a (*k*-1) dimensional Dirichlet distribution. The BB method first draws a value p***** of p from this posterior distribution, then independently draw imputations for missing values from among $d_1, \ldots, d_K$ using the probabilities in p *****.

The difference between ABB and BB is that the underlying parameter of the data, which gives the probabilities of each component in Y_{obs}, is being drawn from a scaled multinomial with the ABB rather than from a Dirichlet distribution. Both distributions have the same means and correlations, but the variances for the ABB method are (1+1/*r*) times the variances for the BB method (Rubin 1981).

1.2.6 Within-class random imputation

Random hot deck is a specific within-class random imputation method. Two factors may vary from one method to another in the within-class random imputation methods: how to form the imputation classes and how to draw imputations within each class. The three most commonly used methods for constructing imputation classes are as follows:

(i) Imputation classes are formed using multiple auxiliary variables. Cases matching on selected auxiliary variables are classified into the same imputation class. The disadvantage of this method is that, as the number of auxiliary variables increase, the number of imputation classes can quickly become enormous. This may limit the use of auxiliary information in the imputation.

(ii) Imputation classes are constructed using regression predicted values from a multivariate regression model. Cases with close predicted values are classified into the same imputation class. The use of auxiliary variables is unlimited (at least theoretically so) with this classification method. This method was used by imputation software PROC IMPUTE (version 2.0, Wise & McLaughlin, 1992).

(iii) Imputation classes are constructed using the propensity score method (Rosenbaum and Rubin 1983, 1984). In brief, the idea is to find a single valued function $b(X)$ of the covariates X, with the property that the desirable properties of classification on X are inherited by classifying on $b(X)$. As shown by Rosenbaum and Rubin, the best such score is the function $e(X)$, the propensity given X, defined as the conditional probability of observing the target variables Y given X. Then, the property that the missing mechanism is independent of Y given X, carries over to independence given the propensity score $e(X)$, so that the imputation is unbiased. The propensity scores can be estimated through logistic regression.

ABB and BB (described in sections 1.2.4 and 1.2.5) have already been shown to draw imputations within each imputation class. The following methods also do so (Gimotty & Brown 1990).

(i) *Resampling using simple random sampling with replacement*: Within *k-th* imputation class, the imputed value is selected randomly with replacement from a multinomial distribution with parameter vector $\underline{p}_k$, the observed proportions of all possible categories. Then, given the observed data, the conditional expected value and conditional variance of $\underline{\hat{p}}_k^*$, the proportion estimates of all possible categories based on the imputed values only, are

$$E[\underline{\hat{p}}_k^* | data] = \underline{p}_k , \qquad Cov[\underline{\hat{p}}_k^* | data] = \frac{1}{m_k}\left(diag(\underline{p}_k \underline{p}_k^T) - \underline{p}_k \underline{p}_k^T \right) ,$$

where m_k is the number of missing values in *k-th* imputation class.

(ii) *Resampling using simple random sampling without replacement*: Within *k-th* imputation class, each observed value is used only once as an imputed value. However, when $m_k > r_k$, all observed values are used as many times as possible and then a simple random sample is taken from the observed values without replacement and those values are used as imputed values for the remainder of the nonrespondents. Here, we only consider $m_k \le r_k$. In this case, the distribution of the frequencies of the imputed values in each category is hypergeometric. The conditional expectation given the data is the same as in (i), whereas the conditional variance-covariance matrix is given by

$$Cov[\underline{\hat{p}}_k^* | data] = \frac{r_k - m_k}{(r_k - 1) m_k}\left(diag(\underline{p}_k \underline{p}_k^T) - \underline{p}_k \underline{p}_k^T \right) .$$

(iii) *Randomized strategy using maximum likelihood estimates*: Let the proportion estimate based on the observed data be $\underline{p}_k = (p_{1k}, \ldots, p_{jk}, \ldots, p_{Ik})^T$, then the estimated frequency is $m_k \underline{p}_k = (m_k p_{1k}, \ldots, m_k p_{jk}, \ldots, m_k p_{Ik})^T$. Then category j is assigned as the imputed value to $[m_k p_{jk}]$ missing cases, which leaves $c_k = \sum_{j=1}^{I} m_k p_{ik} - [m_k p_{jk}] \equiv \sum_{j=1}^{I} c_{jk}$ missing values un-imputed in the *k-th* imputation class, where $[m_k p_{jk}]$ is the largest integer which is smaller than $m_k p_{jk}$. The imputed values for these remaining missing values are independently selected from multinomial distribution with parameter vector $\underline{c}_k^*$ where $c_{jk}^* = c_{jk} / c_k$. The conditional expectation of the imputed proportion is the same as before, but the conditional variance-covariance matrix is given by

$$Cov[\underline{\hat{p}}_k^* | data] = \frac{c_k}{m_k^2}\left(diag(\underline{c}_k^* (\underline{c}_k^*)^T) - \underline{c}_k^* (\underline{c}_k^*)^T \right) .$$

Method (i) is strictly stochastic and acts to increase the variability of statistics computed from an imputed data set compared to a deterministic method. Both method (ii) and method (iii) may be deterministic. Method (ii) is deterministic when the number of observations equals the number of missing values. Method (iii) is deterministic when $m_k p_{jk}$ are integers for each imputation class. However, in general, method (ii) adds more variability than method (iii) and method (i) adds more variability than method (ii). However, all of them add less variability than the ABB and the BB imputation methods.

1.3 Model-based deterministic imputation methods

Generally, "correctly" modeling missing data must be the data constructor's responsibility because he/she typically knows more about reasons for nonresponse and has access to confidential and detailed information not released for public use. Model-based approaches will produce more accurate imputations than randomization-based approaches if the model assumptions are satisfied. But the difficulty with model-based approaches is that those assumptions are usually unverifiable in practice and therefore it may not be easy to choose an appropriate model-based imputation approach for a typical survey. A good model-based approach would work well for a wide range of underlying data distributions and missing mechanisms.

1.3.1 Ratio imputation

Suppose that an auxiliary variable x closely related to the target variable y is observed on all sample units. Ratio imputation uses $y_{hi}^{*} = \frac{\bar{y}_{rh}}{\bar{x}_{rh}} x_{hi}$ as imputed values for the $i\text{-}th$ nonrespondent in $h\text{-}th$ imputation class. This method can be motivated by the fact that y_{hi}^{*} is the best predictor under the following "ratio" superpopulation model:

$$E(y_{hi}) = b_h x_{hi}, \quad V(y_{hi}) = s_h^2 x_{hi}, \quad Cov(y_{hi}, y_{hj}) = 0,$$

provided that the model holds for both the respondents and nonrespondents.

The ratio imputation method may provide very accurate imputations if the missingness of y mainly depends on a highly correlated auxiliary variable x. But this is a very restrictive assumption. In practice, missing values are more likely to depend on several auxiliary variables. Since ratio imputation can use only one auxiliary variable, it is not fully efficient in many situations. One way around this is to use some auxiliary variables as classification variables, but this is still not a satisfactory solution to the limitation on the efficient use of auxiliary variables. As the number of classification variables increase, the number of imputation classes quickly becomes enormous and then some imputation classes may not have sufficient samples to obtain fairly accurate ratio estimates.

1.3.2 Predicted regression imputation

This method uses the predicted values from a regression model as imputations for all missing cases. The predicted value $\hat{y}_i$ is the best predictor of the *i-th* unobserved value y_i under the following super-population model:

$$E(y_i) = a + b'x_i, \quad V(y_i) = s^2, \quad Cov(y_i, y_j) = 0$$

provided that the model holds for both the respondents and the nonrespondents. Predicted regression imputation may also be performed within each imputation class. The disadvantage of this method is the shrinkage to the mean phenomenon.

1.3.3 EM algorithm

The EM algorithm (Dempster, Laird, and Rubin 1977) consists of two steps: the *E*-step calculates the expectation of the complete data sufficient statistics given the observed data and current parameter estimates, and the *M*-step updates the parameter estimates through the maximum likelihood approach based on the current values of the complete sufficient statistics. The algorithm then proceeds in an iterative manner until the difference between the last two consecutive parameter estimates converges to a specified criterion. The final *E*-step calculates the expectation of each missing value given the final parameter estimates and the observed data; this will be used as the imputation value.

Although the EM algorithm can be used to impute each individual missing value, it is more often used to directly obtain estimates for population parameters. Assuming a normal distribution for the data, both the expectations of the sufficient statistics in the *E*-step and the maximum likelihood estimates of the parameters in the *M*-step are easy to derive. But it may not be easy to do so with other distributions. Convergence may be slow and not guaranteed with the EM algorithm especially with sparse data. If each *M*-step also requires an iterative process to obtain the maximum likelihood estimates, the convergence process will further be slowed down. This method also suffers the shrinkage to the mean phenomenon. The advantage of the *EM* algorithm is its stable convergence; that is, iterations always increase the likelihood.

1.3.4 Dear's principal component method (DPC)

The imputation strategy using the principal component method consists of three steps:

(D1) Let $R = \{r_{ij}\}$ be an n ´ p missingness indicator matrix for variables $X_1 \ldots X_p$ with n observations, i.e., $r_{ij} = 0$ or 1 according to whether x_{ij} is missing or observed. Use all available cases to calculate the sample mean and variance for each variable, and then standardize X to Z. Next, use the case-wise-deletion method (delete the whole case if one variable has a missing value on that case) to obtain the correlation matrix, S.

(D2) Calculate the largest eigenvalue of S, l_1, and its associated eigenvector $h_1 = (h_{11}, \cdots, h_{1p})$.

(D3) Let the first principal component for the ith case be

$$g_i = \sum_{j=1}^{p} h_{1j} z_{ij} r_{ij} ,$$

so that the points on the first principal component line that are closest to the i-th case replace the missing variables:

$$z_{ij}^* = \begin{cases} z_{ij}, & if \quad r_{ij} = 1 \\ h_{1j} g_i & if \quad r_{ij} = 0 \end{cases}.$$

Repeat (D3) for all cases with missing variables and convert Z^* back to X^*.

One desirable property of principal component analysis is that it does not require any distributional assumptions for its use. However, since the case-wise-deletion method is used to obtain the correlation matrix S, *DPC* works poorly for data sets with only a few complete cases.

1.3.5 General iterative principal (GIP) component method

To avoid the problems mentioned above and make DPC a general purpose method, the following refinements have been introduced.

(G1) Use all-available-data method to calculate S. If S is non-positive definite, modify it with the algorithm provided by Huseby, Schwertman, and Allen (1980); or replace all missing values by the mean and use n-m-1 instead of n-1 as the denominator in the variance-covariance calculations to obtain S.

(G2) Perform D2 and D3 with S obtained from $G1$.

(G3) Recalculate S from the imputed data matrix and repeat $G2$.

(G4) Cycle iteratively through $G3$ and $G2$ until successive imputed values do not change materially.

1.3.6 Singular value decomposition (SVD) method

Singular value decomposition (SVD) can be used in a simple way to impute data to missing values (Krzanowski 1988). The method is easy to compute and a description of the steps for one missing value x_{ij} in X followed:

(S1) Omit the *ith* case (row) from X and calculate the SVD of the remaining $(n\text{-}1)\times p$ data matrix, denoted by $X^{-i} = \overline{U}\overline{D}\overline{V}'$ with $\overline{U} = \{\overline{u}_{st}\}$, $\overline{V} = \{\overline{v}_{st}\}$ and $\overline{D} = diag\{\overline{d}_1, \cdots, \overline{d}_p\}$, where $\overline{U}$ and $\overline{V}$ are orthonormal matrices (i.e., $\overline{U}'\overline{U} = \overline{U}\overline{U}' = I$).

(S2) Omit the *jth* variable (column) from X and calculate the SVD of the remaining $n\times(p\text{-}1)$ data matrix, denoted by $X_{-j} = \tilde{U}\tilde{D}\tilde{V}$ with $\tilde{U} = \{\tilde{u}_{st}\}$, $\tilde{V} = \{\tilde{v}_{st}\}$ and $\tilde{D} = diag\{\tilde{d}_1, \cdots, \tilde{d}_{p-1}\}$.

(S3) Impute for *(i, j)th* missing case with

$$x_{ij}^* = \sum_{t=1}^{p-1} \left(\tilde{u}_{it}\tilde{d}_t^{1/2}\right)\left(\overline{u}_{jt}\overline{d}_t^{1/2}\right).$$

In the case where there is more than one missing value, an iterative scheme can be conducted as follows: start with any initial imputed values such as the mean, and update each initial imputed value in turn using *S*3. The process is then iterated until stability is achieved in the imputed values.

1.3.7 A comparison of ASM, EM, DPC, GIP, and SVD

Bello (1993) conducted a simulation study to compare the five deterministic imputation methods: the adjusted mean substitution (AMS), EM algorithm, DPC, GIP, and SVD. In the study, Bello's two simulation populations are multivariate normal $N_p(m, \Sigma)$ and *t*-distribution with 4 degrees of freedom, $T_p(4, m, \Sigma)$, where $m=0$ and $\Sigma=V\Lambda V'$. V is a randomly generated orthogonal matrix and $\Lambda=diag\{\lambda_1, \ldots, \lambda_p\}$, $l_i = wv^{i-1} + 0.1$ as used by Bendel (1978), where

$$w = \begin{cases} (c - 0.1p)(1 - v)/(1 - v^p) & 0 < v < 1 \\ c/p - 0.1 & v = 1 \end{cases}$$

and c is the trace of Σ. Evidently, values of v represent a continuum such that the interdependence among the variables increases as v decreases from 1 to 0. The variables are independent when $v=1$.

Other varying factors are sample size (n), dimensionality (p), interdependence among the variables (v), and missing rate (γ). Missing data are created randomly, which actually results in the ideal missing mechanism, missing completely at random. The number of Monte Carlo simulations for each combination of n, p, v, and γ was fixed at 100. The mean square error (Euclidean norm) of the estimators of Σ over the 100 simulations are used as the main comparison criterion. For the estimator of the mean vector, all the imputation methods are similar since the data are missing completely at random.

The primary findings of Bello's study are as followed:

For multivariate normal distributions:

- When the variables are nearly independent (v=0.7) and p<10, the AMS outperforms the other four regression-like imputation methods. EM algorithm is the second best, followed by DPC, SVD, and GIP. This is not surprising since the mean imputations are obtained under the pretext that the variables are uncorrelated.

- For p>2, as v≤0.3, the regression-like imputation techniques show appreciable superiority over the adjusted mean imputation method.

- When the missing rate r≥0.10 and n becomes large (≥100), EM is, on the average, the best technique followed by GIP, SVD, ASM, and DPC.

For multivariate t-distributions:

- Although the principal component and singular value decomposition method can be presumed to be distributional-assumption-free, this does not mean that DPC, GIP, and SVD are robust to structures in data.

- When v=0.7, the imputation methods behave similarly to their normal counterparts.

- EM—which depends on a normality assumption—is running neck-and-neck with the distributional-free techniques—DPC, GIP, and SVD. When n is sufficiently large (200) and the variables are strongly dependent (v<0.3) with moderate dimensionality (p=5), EM outperforms the other imputation techniques. On the other hand, when p=2 and v=0.3, for any n value, GIP is the most efficient method.

- When p increases, n increases, and v decreases, the regression-like methods become better and better than ASM.

- There is insufficient evidence to discredit the use of EM when the data are markedly deviate from normality especially when p>2 and reasonably moderate-to-high interdependence exists among the variables. This remark implicitly suggests that whatever is known to affect EM—for example, outliers—may also affect other imputation techniques as well.

Regarding the computer-time used by these imputation techniques, ASM and DPC are non-iterative techniques and no special computer-time is required. Among the three iterative methods, the convergence rate of EM was observed to be the slowest, followed by SVD, and then GIP.

Although the performances of the methods are compared based on the artificial assumption, MCAR, these results can still be used as references.

1.4 Model-based random imputation methods

1.4.1 Draw imputations from predicted distributions

If some information about the type of data distribution is available, imputations can be drawn from a predicted distribution. This method assumes a distribution for the data and uses 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. Rubin's example (Rubin 1978) can illustrate this. Suppose a sample of 1000 units with 500 respondents and 500 nonrespondents. The 500 respondents look like a half-normal. If we learn from other sources that the population is approximately normal, then we can use the data of the 500 respondents to obtain the mean and variance estimates, and draw imputations from the normal distribution with the estimated mean and variance. This makes it possible to recover the other half of the normal distribution. Although this is an extremely artificial example, it is possible in real applications that data of some specific categories are totally or mostly missing. In those cases, methods that draw imputations from observed data will not be able to recover missing values for those categories, while drawing imputations from a predicted distribution may be able to recover them. The disadvantage of this method is that it requires information in order to develop an appropriate distribution assumption.

1.4.2 Random regression imputation

As stated in section 1.3.2, predicted regression imputation suffers from shrinkage to the mean phenomenon. Small random disturbances can be added to the predicted values as imputations to increase variability. The small random disturbance may be drawn using the following methods:

(1) draw a random disturbance from a distribution such as $N(0,\hat{s})$ with mean 0 and variance $\hat{s}$ obtained from observed data;
(2) draw a random disturbance from respondents' residuals of the regression model;
(3) draw a random disturbance from residuals of those respondents which have similar values on some selected auxiliary variables to protect against non-linearity and non-additivity in regression models.

1.4.3 Ratio with random disturbance imputation

We can add a small random disturbance to the imputed values obtained from a ratio imputation model (see section 1.3.1) as was done above to the predicted regression imputation. The random disturbance can be drawn using three methods parallel to those described above.

1.4.4 Modeling non-ignorable missing mechanism

Most imputation methods 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 conditional auxiliary variables. Greenless, Reece, and Zieschang (1982) try to model both the target variable and its missing indicator variable for a non-ignorable missing mechanism which allows the missingness to depend on the target variable itself.

Let *Y* be the target variable with missing values, *X* be the auxiliary variables for predicting *Y*, *R* be the response indicator, and *Z* be the auxiliary variables for predicting *R*. *X* and *Z* may overlap. Then the imputation model employed is:

$$Y_i = X_i b + e_i \qquad e_i \sim N(0, s^2)$$
$$P(R_i = 1 | Y_i, Z_i) = 1/[1 + \exp(-a - gY_i - dZ_i)].$$

The later equation indicates that the response probability of *Y* depends on *Y* itself. Then the likelihood for *i-th* respondent is given by

$$L_i = \frac{1}{1 + \exp(-a - gY_i - dZ_i)} \cdot \frac{1}{s} \cdot f\left(\frac{Y_i - X_i b}{s}\right),$$

and the likelihood for *i-th* nonrespondent is given by

$$L_i = \int_{-\infty}^{+\infty} \left(1 - \frac{1}{1 + \exp(-a - gY - dZ_i)}\right) \cdot \frac{1}{s} \cdot f\left(\frac{Y - X_i b}{s}\right) dY .$$

The maximum likelihood estimates for a, b, g, d, and s are obtained by maximizing the whole sample likelihood $L = \prod_{i=1}^{n} L_i$. The solution to this maximizing problem may be found through the generalized Gauss-Newton algorithm.

We may impute the missing values using the mean of the distribution of *Y* conditional on nonresponse, the values of *X* and *Z*, and the parameter estimates $\hat{a}, \hat{b}, \hat{g}, \hat{d}$, and $\hat{d}$. This mean can be calculated in a straightforward way using numerical integration:

$$E(Y_i | X_i, Z_i, R_i = 0) = \frac{\int_{-\infty}^{+\infty} Y \left(1 - \frac{1}{1+\exp(-\hat{a} - \hat{g}Y - \hat{d}Z_i)}\right) \cdot \frac{1}{\hat{s}} \cdot f\left(\frac{Y - X_i\hat{b}}{\hat{s}}\right) dY}{\int_{-\infty}^{+\infty} \left(1 - \frac{1}{1+\exp(-\hat{a} - \hat{g}Y - \hat{d}Z_i)}\right) \cdot \frac{1}{\hat{s}} \cdot f\left(\frac{Y - X_i\hat{b}}{\hat{s}}\right) dY}$$

Alternatively, to avoid the shrinkage to the mean phenomenon, we may use the following imputation scheme.

(1) Draw e_i from $N(0,1)$ and a uniform random number η from $U[0, 1]$.

(2) Calculate $\hat{Y}_i = X_i\hat{b} + \hat{s}e_i$ and $\Pr(R_i = 0 | \hat{Y}_i, Z_i) = 1 - \frac{1}{1+\exp(-\hat{a} - \hat{g}\hat{Y}_i - \hat{d}\hat{Z}_i)}$.

(3) If $\Pr(R_i = 0 | \hat{Y}_i, Z_i) > h$, impute $\hat{Y}_i$ for the *i-th* missing case; otherwise re-do (1) and (2).

If the model of the missing indicator variable is approximately satisfied, this method should give better imputations than usual imputation methods. However, that is an unverifiable assumption in real applications and the extra model makes it less robust for general imputation purposes. This method may not be recommended if there is no strong evidence to show that the missing mechanism is confounded, that is, the missingness of *Y* depends on *Y* itself.

1.5 Imputation methods related to Bayesian theories

1.5.1 Data augmentation

This Bayesian iterative method was proposed by Tanner and Wong (1987). It assumes two distributions: the distribution of the data and the prior distribution of the parameters. Similar to the EM algorithm, it consists of two steps: (1) *I*-step (imputation step) draws imputations for the missing values from the predicted distribution of the data, using current parameter estimates; (2) *P*-step (parameter estimation step) draws parameter estimates from their posterior distribution, using both the observed and imputed data. To start this iterative process, we may use the EM algorithm to obtain initial parameter estimates for the first *I*-step.

Schafer's software (Schafer 1997) implements this method using models for continuous data, categorical data, and mixed continuous and categorical data.

- For continuous data, this software assumes a multivariate normal distribution for the data, and a normal prior for the mean parameters and a normal-inverted Wishart for the variance-covariance parameters.

- For categorical data, this software assumes a multinomial distribution for the data and a Dirichlet prior distribution for the parameters. In cases where the number of parameters becomes enormous, the software imposes loglinear constraints (Bishop, Fienberg, and Holland 1975) on the parameters.

- For mixed continuous and categorical data, the software employs a general location model (Olkin & Tate 1961). It assumes multinomial distribution for the categories defined by the categorical variables. Within each category, the continuous variables are assumed to have multivariate normal distribution. The prior for the parameters in the multinomial distribution is Direchlet and that for the parameters in the multivariate normal distribution is Jeffrey's non-informative prior. To reduce the parameters, a loglinear constraint can be imposed on the multinomial parameters and a linear constraint on the mean parameters of the multivariate normal distribution.

The data augmentation procedure approximates the actual posterior distribution of the parameter vector by a mixture of complete data posteriors. Their method of constructing the complete data sets is closely related to the Gibbs sampler (Geman and Geman 1984). This method efficiently uses relationships 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. Under simple random sampling, the data augmentation method provides "proper" multiple imputations in the sense of Rubin (1987). The disadvantage of the data augmentation method is that it requires iterations and, similar to the EM algorithm, convergence can be slow.

1.5.2 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 Schenker (1986) adjusted the normal model implemented in Schafer's software as follows. First, the parameters m^* and s^{*2} are obtained in the same way as in the data augmentation method. Second, the components of m-dimensional vector $X = (X_1, \ldots, X_m)$ are drawn with replacement from the observed data Y_{obs}. Under repeated draws from Y_{obs}, the standardized variable

$$Z_i = (X_i - \bar{y}_r) / \sqrt{(r-1)s_r^2 / r}$$

has expected value 0 and variance 1. Finally, the m missing values Y_{mis} are imputed using $m^* + s^* Z_i$, i=1, 2, …, m.

1.5.3 Sequential imputation method

Kong, Liu and Wong (1994) propose a sequential imputation procedure that involves imputing the missing data sequentially. According to the authors, in many applications the sequential imputation method can work well without the need for iterations.

To describe the method, let q be the parameter vector of interest and Y be the complete data. Suppose the complete-data posterior distribution $p(q \mid Y)$ is simple. Suppose the real data Y can be decomposed into

$$Y = \begin{pmatrix} Y_1 \\ \vdots \\ Y_t \\ \vdots \\ Y_n \end{pmatrix} = \begin{pmatrix} (Y_{r1}, Y_{m1}) \\ \vdots \\ (Y_{rt}, Y_{mt}) \\ \vdots \\ (Y_{rn}, Y_{mn}) \end{pmatrix} = (Y_r, Y_m),$$

where Y_{rt} and Y_{mt} (t=1, 2, …, n) are the response and nonresponse variables in the t-th observation. The missing variables may be different for different observations. The main goal is to find the posterior distribution $p(\theta \mid Y_r)$:

$$p(q \big| Y_r) = \int p(q \big| Y)\, p(Y_m \big| Y_r) dY_m = E_{Y_m|Y_r}\left[p(q \big| Y) \right].$$

If we can draw M independent copies of Y_m's from the conditional distribution $p(Y_m \mid Y_r)$, then we can approximate the posterior distribution $p(\theta \mid Y_r)$ by $\frac{1}{M}\sum_{j=1}^{M} p(q|Y(j))$, where $Y(j) = (Y_r, Y_m(j))$ and Y_m(j) is the j-th imputations for the missing part Y_m. However, drawing imputations directly from conditional distribution $p(Y_m \mid Y_r)$ is usually difficult. The Gibbs sampler or the data augmentation procedure do this approximately by iterations.

The sequential imputation method achieves something similar by imputing the Y_{mt}'s sequentially and using importance sampling weights to avoid iterations. The sequential imputation starts by drawing Y_{m1}^* from $p(Y_{m1} \mid Y_{r1})$ and computing $w_1 = p(Y_{r1})$. Then for t=2, …, n, the following two steps are done sequentially,

(1) Draw Y_{mt}^* from the conditional distribution $p(Y_{mt} | Y_{r1}, Y_{m1}^*, \ldots, Y_{r,t-1}, Y_{m,t-1}^*, Y_{rt})$;

(2) Compute the predictive probabilities $p(Y_{rt} | Y_{r1}, Y_{m1}^*, \ldots, Y_{r,t-1}, Y_{m,t-1}^*)$ and

$$w_t = w_{t-1} \cdot p(Y_{rt} | Y_{r1}, Y_{m1}^*, \ldots, Y_{r,t-1}, Y_{m,t-1}^*). \qquad (1.1)$$

Let $w = w_n$, so that

$$w = p(Y_{r1}) \cdot \prod_{t=2}^{n} p(Y_{rt} \mid Y_{r1}, Y_{m1}^*, \ldots, Y_{r,t-1}, Y_{m,t-1}^*) .$$

Both steps are required to be computationally simple, which is often the case if the predictive distributions $p(Y_1)$ and $p(Y_t \mid Y_1, \ldots, Y_{t-1})$ are simple. This is the key to the feasibility of sequential imputation.

We can independently repeat the above process M times to draw M sets of imputations and weights, denoted as $Y_m^*(j)$ and $w(j)$ respectively (j=1, 2, …, M). Then the posterior distribution $p(\theta \mid Y_r)$ is estimated by

$$\frac{1}{W} \sum_{j=1}^{M} w(j)\, p\left(q \middle| Y_r, Y_m^*(j)\right), \qquad (1.2)$$

which is easy to compute under the assumption that the complete-data posterior is simple, where $W = \sum w(j)$.

To understand why (1.2) is the appropriate approximation, we note that each independent imputation $Y_m^*(j)$ is not drawn from the actual conditional distribution $p(Y_m \mid Y_r)$, but from the "trial density"

$$p^*(Y_m^* \mid Y_r) = p(Y_{m1}^* \mid Y_{r1}) \prod_{t=2}^{n} p(Y_{mt}^* \mid Y_{r1}, Y_{m1}^*, \ldots, Y_{r,t-1}, Y_{m,t-1}^*, Y_{rt})$$

Using standard results from importance sampling, we should use weights

$$w^*(j) = \frac{p(Y_m^*(j) \mid Y_r)}{p^*(Y_m^*(j) \mid Y_r)} = \frac{p(Y_m^*(j), Y_r)}{p(Y_r)} \cdot \frac{p(Y_{r1})}{p(Y_{r1}, Y_{m1}^*(j))} \prod_{t=2}^{n} \frac{p\left(Y_{r1}, \ldots, Y_{rt}, Y_{m1}^*(j), \ldots, Y_{m,t-1}^*(j)\right)}{p\left(Y_{r1}, \ldots, Y_{rt}, Y_{m1}^*(j), \ldots, Y_{mt}^*(j)\right)}$$

$$= \frac{p(Y_m^*(j), Y_r)}{p(Y_r)} \cdot \frac{p(Y_{r1})}{p(Y_{r1}, \ldots, Y_m, Y_{m1}^*(j), \ldots, Y_{mn}^*(j))} \prod_{t=2}^{n} \frac{p\left(Y_{r1}, \ldots, Y_{rt}, Y_{m1}^*(j), \ldots, Y_{m,t-1}^*(j)\right)}{p\left(Y_{r1}, \ldots, Y_{r,t-1}, Y_{m1}^*(j), \ldots, Y_{m,t-1}^*(j)\right)}$$

$$= \frac{p(Y_{r1})}{p(Y_r)} \prod_{t=2}^{n} p\left(Y_{rt} \mid Y_{r1}, \ldots, Y_{r,t-1}, Y_{m1}^*(j), \ldots, Y_{m,t-1}^*(j)\right) = \frac{w(j)}{p(Y_r)},$$

which is proportional to $w(j)$ since $p(Y_r)$ is the same for all M imputations. This implies $w(j)$ (j=1, …, M) are correct weights and (1.2) is an appropriate approximation.

In sequential imputation, it is generally desirable to have the trial distribution $p^*(Y_m \mid Y_r)$ as close to the true distribution $p(Y_m \mid Y_r)$ as possible. This usually means that the complete cases should be processed first, and the other cases should be processed in order of increasing missingness so that missing values are imputed conditioned on as many of Y_r as possible. One advantage of sequential imputation is that this method can impute data sequentially even when the data are collected at different times, for example, in medical studies.

In situations where we want to compare models, it will be important to get the likelihood of different models. For a particular model *H* the likelihood of *H* given incomplete data Y_r is

$$p_H(Y_r) = \int p_H(Y_r|q)p_H(q)dq \, .$$

Suppose that we have applied sequential imputation based on model *H*. Then for all *j* we have

$$1 = E_{p^*}[w^*(j)] = E_{p^*}[w(j) / p(Y_r)] \, ,$$

which implies $E_{p^*}[w(j)] = p(Y_r)$. Therefore,

$$\hat{p}(Y_r) = \frac{1}{M}\sum_{j=1}^{M} w(j)$$

is an unbiased estimate of the likelihood $p(Y_r)$ for the imputation model.

In summary, 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. However, it requires that $p(Y_1)$, $p(Y_t|Y_1,\ldots,Y_{t-1})$, and $p(q \mid Y)$ are all simple. Otherwise, it may be not feasible to implement the sequential imputation method. This is a very restrictive condition.

1.6 Imputation practice across NCES surveys

The following surveys conducted by the National Center for Education Statistics over the years used some method to impute for item nonresponse:

Universe Surveys

(1) Common Core of Data (CCD, conducted annually)
(2) Private School Universe Survey (PSS, conducted biennially)
(3) Integrated Postsecondary Education Data System (IPEDS):
Institutional Characteristics (IPEDS-IC, conducted annually)
Fall Enrollment (IPEDS-EF, conducted annually)
Completions (IPEDS-C, conducted annually)
Financial Statistics (IPEDS-F, conducted annually)
Salaries, Tenure and Fringe Benefits of Full-Time Instructional Faculty (IPEDS-SA, conducted annually)

Fall Staff (IPEDS-S, conducted biennially)
Academic Libraries (IPEDS-L, conducted biennially)

Sample Surveys

(1) Schools and Staffing Survey (SASS, conducted in 1987–88, 1990–91, 1993–94)
(2) SASS Teacher Follow-up Survey (SASS-TFS, conducted in 1988–89, 1991–92, 1994–95)
(3) National Household Education Survey (NHES, conducted in 1991, 1993, 1995, 1996)
(4) Recent College Graduates Survey (RCG, conducted in 1976, 1978, 1981, 1985, 1987, 1991)
(5) National Study of Postsecondary Faculty (NSOPF, conducted in 1988 and 1993)
(6) National Assessment of Education Progress (NAEP, conducted biennially since 1980 and annually from 1969 to 1980)
(7) Third International Mathematics and Science Study (TIMSS, conducted in 1995)
(8) National Postsecondary Student Aid Study (NPSAS, conducted at 3-year intervals since 1986–87)

Fast Response Surveys

(1) Fast Response Survey System (FRSS; "College-Level Remedial Education in the Fall of 1989," conducted in 1990)
(2) Postsecondary Education Quick Information System (PEQIS; "Deaf and Hard of Hearing Students in Postsecondary Education," conducted in 1993)

Imputation methods used across these surveys are presented in table 1.6.1.

Table 1.6.1—Imputation methods used across NCES surveys

Survey	Imputation Methods Used
CCD	Ratio imputation and adjustment
PSS	Sequential hot deck, ratio adjustment, deductive imputation
IPEDS-IC	Ratio imputation, mean imputation
IPEDS-EF	Ratio imputation, mean imputation, raking method
IPEDS-C	Cold deck imputation, ratio imputation, raking method, mean imputation
IPEDS-SA	Within-class ratio imputation, within-class mean imputation
IPEDS-F	Ratio adjusted cold deck imputation, sequential hot deck imputation
IPEDS-S	Ratio adjustment cold deck imputation, hot deck imputation
IPEDS-L	Logical imputation, ratio adjustment
IPEDS-ALS	Cold deck imputation, ratio imputation
NSOPF	PROC IMPUTE, sequential hot deck
SASS	Sequential hot deck, deductive imputation
SASS-TFS	Sequential hot deck, deductive imputation
RCG	Hot deck, within-class random imputation, deductive imputation
NHES	Hot deck, manual imputation
NPSAS	Hot deck, regression imputation, deductive
NAEP	Multiple imputation based on Bayesian models*
TIMSS	Multiple imputation based on Bayesian models*
FRSS	Sequential hot deck imputation, mean imputation, and median imputation
PEQIS	Sequential hot deck imputation, ratio adjustment

* Multiple imputation techniques were applied to create plausible values for performance scores based on Item Response Theory.

Chapter 2 Imputation Software Products

2.1 PROC IMPUTE (See 1.2.6 Within-class random imputation)

PROC IMPUTE is an advanced imputation software created by American Institutes for Research (AIR) under a contract with NCES. It is a stand-alone FORTRAN program and only works with ASCII data files. The software is in the public domain and users can obtain a copy through NCES.

PROC IMPUTE is a regression-based distributional estimation procedure that is believed to be more general and to produce more accurate results than a standard hot deck procedure (AIR, 1980). It considers each variable on the file in turn as a “target” variable whose missing values are to be filled in, and it uses information on other variables to minimize the error in imputing each target variable. PROC IMPUTE uses three steps that are similar to those used in hot deck procedure to impute each target variable:

(1) It uses stepwise regression analysis to find the best combination of predictors for each target variable;

(2) It creates homogeneous cells (imputation classes) of records which have close predicted regression values;

(3) It imputes each missing record in a given cell with a weighted average of two donors which are drawn from its own cell and its adjacent cell, respectively, with probability proportional to the observed frequencies within the two cells. The weighted average value is rounded to an integer if the integer flag is set for the target variable.

The software also automatically creates missing data flags for each variable with a value of “I” for imputed values, “R” for reported values, and “A” for skip missing values.

Since PROC IMPUTE involves ordinary multivariate regression analysis, it only works for continuous and dichotomous variables. Polytomous variables need to be recoded into dichotomous variables before running PROC IMPUTE.

PROC IMPUTE can incorporate about 30 variables in one imputation model. A large data set needs to be divided into several subsets and each subset is imputed via a separate imputation model. Some key variables may be included in all imputation models. Note that PROC IMPUTE does not need to be run multiple times to impute a large data set because of the batch run feature of PROC IMPUTE: one batch run can handle all the data no matter how large the data set is.

PROC IMPUTE has two other important features. First, it can create as many as nine sets of imputations. Although it is not "proper" according to Rubin's multiple imputation theory (Rubin 1987), results of our simulation study (described in chapter 5) show that, in many situations, PROC IMPUTE provides better multiple imputation variance estimates than some "proper" methods. Second, it can perform within-class imputations through a "BY" statement which is parallel to a SAS "BY" statement. This feature is useful for stratified data where the user may want to perform imputations within each stratum. It is also convenient for Monte Carlo simulations where multiple data sets need to be generated so that the average performance over replications can be assessed. Using a "BY" statement with a data set identification variable, all data sets can be imputed through one run of PROC IMPUTE.

2.2 Schafer's imputation software (See 1.5.1 Data augmentation under Imputation methods related to Bayesian theories)

Dr. Joseph Schafer of Pennsylvania State University developed this public domain software. The original version was written using S-PLUS functions and FORTRAN subroutines and ran under an S-PLUS environment. The current menu-driven version for Windows was written in FORTAN 90. It only works with ASCII data files in which a numeric value is used to represent a missing value. It will not work if a "." is used as a missing value in the ASCII files.

Schafer's imputation software (Schafer 1997) applies the data augmentation method. Like the EM algorithm, it consists of two steps: (1) the *I*-step (imputation step) draws imputations for the missing values from the predicted distribution of the data given current parameter estimates; (2) the *P*-step (parameter estimation step) draws parameter estimates from their posterior distributions given both the observed and imputed data. To start this iterative process, the EM algorithm or ECM algorithm (Meng and Rubin 1991) may be used to obtain initial parameter estimates for the first *I*-step.

The software consists of three modules using different statistical models for continuous data, for categorical data, and for mixed continuous and categorical data.

(1) For continuous data, the software assumes a multivariate normal distribution for the data, and a normal prior for the mean parameters and a normal-inverted Wishart for the variance-covariance parameters. Under these assumptions, the posterior distributions of the mean parameter and the variance-covariance parameters are multivariate normal and normal-inverted Wishart, respectively. Therefore, *P*-steps draw parameter estimates from these posterior distributions and *I*-steps draw

imputations for missing values from their predictive normal distribution with updated parameter estimates obtained in the *P*-steps.

(2) For categorical data, the software assumes a multinomial distribution for the data and a Dirichlet prior distribution for the parameters. Under this saturated multinomial model, the posterior distribution of the parameters—the cell probabilities—is also a Dirichlet distribution. However, as the number of categorical variables increase, the number of cells formed by the variables quickly becomes enormous. In these cases, the software imposes loglinear constraints (Bishop, Fienberg and Holland 1975) to reduce the number of parameters for estimation. For these constrained loglinear models, a Bayesian Iterative Proportion Fitting algorithm (Gelman, Rubin, Carlin and Stern 1995) is used to simulate the posterior distributions for the parameters.

(3) For mixed continuous and categorical data, the software employs a general location model (Olkin and Tate 1961). It assumes multinomial distribution for the categories defined by the categorical variables. Within each category, the continuous variables are assumed to have multivariate normal distribution. The prior for the parameters in the multinomial distribution is a Direchlet distribution and that for the parameters in the multivariate normal distribution is Jeffrey's non-informative prior. In cases where the number of parameters becomes enormous, a loglinear constraint can be imposed on the multinomial parameters and a linear constraint on the mean parameters of the multivariate normal distribution.

2.3 IRMA

Imputation Run Manager (IRMA) is a public domain software developed by Synectics for Management Decisions, Inc., under a contract with NCES. User permission can be obtained through NCES.

IRMA is designed to supply a variety of imputation techniques to the users. The current version of IRMA was built using Microsoft Visual Basic and includes two imputation techniques: 1) PROC IMPUTE and 2) Schafer's Imputation Software. IRMA preserves all the nice features of PROC IMPUTE and Schafer's Imputation Software and provides some enhanced features. For instance, while PROC IMPUTE and Schafer's Imputation Software only work with ASCII files, IRMA works with SAS, SPSS, and ASCII data files. Another enhancement allows the unimputed input data file and the imputed output data file to be of different types. For example, the input file can be a SAS file, but the user can require IRMA to output the imputed file in SPSS format, or in both SPSS and SAS formats. More imputation methods will be added to a future version of IRMA.

2.4 GEIS and GES

Generalized Edit and Imputation System (GEIS) and Generalized Estimation System (GES) were developed by Statistics Canada. GEIS performs data editing and imputation functions while GES constructs point estimates and variance estimates using a number of different estimation modules. The software is a SAS-based application which runs under a SAS environment. Data must be either in SAS format or in ASCII format with fixed field positions. A site license for GEIS and GES costs $20,000 (CDN), and there is a $2,000 yearly maintenance fee.

The imputation methods used in this software are nearest neighbor hot deck, current ratio, current mean, previous value, previous mean, and auxiliary trend, which are the key methods used by Statistics Canada for imputation of survey missing data. All of these are single and deterministic imputation methods and therefore suffer the disadvantage of deflating the variance estimates.

2.5 SOLAS for Missing Data Analysis 1.0

This commercial product was developed by Statistical Solutions Limited. A single user license costs $995 for commercial purposes and $795 for academic purposes.

Imputation methods used in this software include: (1) Group Mean Imputation, which replaces missing values with the cell means of the sample; (2) Last Value Carried Forward (Sequential Hot Deck), in which the last observed value is used to fill in missing values at a later point in the study; and (3) Nearest Neighbor Hot Deck Imputation, in which missing values are replaced with values taken from the closest matching respondents. Multiple imputations can also be created by this software. These imputation methods are not very attractive for the purpose of statistical inference. Any statistician with some programming skill can easily implement these imputation algorithms. However, SOLAS can do more than imputation. It can also perform many standard statistical analyses based on imputed data, including descriptive analysis, cross-tabulation, statistical tests (*t* and non-parametric), ANOVA, regression, BMDP survival analysis.

Chapter 3 Nonresponse Bias

Nonresponse bias is the bias of a survey estimate due to the difference between respondents and nonrespondents. It is one of the most important issues concerning survey data analysts. It is desirable to eliminate nonresponse bias through imputation and/or estimation methods. One way is to construct a so-called *restoring estimator*, defined by Rancourt, Lee, and S@rndal (1994) as:

> Given the sample *S*, if the conditional expectation of the difference between an imputation estimator $\hat{\bar{y}}^*$ and the complete data estimator $\bar{y}_S$ equals to 0, i.e., $E(\hat{\bar{y}}^* - \bar{y}_S | S) = 0$, where the expectation is over the response mechanism and the imputation model, then $\hat{\bar{y}}^*$ is called a *restoring estimator*.

This actually is equivalent to the "first order proper" estimator defined by Rubin (1996).

If missing values occur completely at random (MCAR)—that is, the survey has uniform response—, then the respondents represent the population well and survey nonresponse causes no bias. However, this ideal missing mechanism rarely exists in real applications.

The most commonly assumed missing mechanism is *missing at random* (MAR), which may more appropriately be called *missing conditionally at random*. MAR requires that respondents and nonrespondents have no systematic differences given some observed auxiliary variables (called *conditioning variables* in imputation literature). One simple example of MAR is that respondents and nonrespondents within each imputation class formed by some predictive auxiliary variables both represent random samples from the subpopulation. In this case, estimates within each imputation class will have no nonresponse biases, and thus the combined overall estimates will have no nonresponse bias. Therefore, with a missing mechanism MAR, nonresponse bias can be corrected through imputation by conditioning on the auxiliary variables that are related to the missing mechanism of the target variable. In real applications, we usually do not know which auxiliary variables are responsible for the missing values of the target variable. Thus many imputation pioneers such as Rubin and Little advocate using as many auxiliary variables as possible to make the missing mechanism as close to MAR as possible.

Different imputation methods use conditioning variables in different ways. Some ways are more effective than others depending upon the circumstances. Hot deck method uses conditioning variables as classification or matching variables; regression-type imputation uses conditioning variables as predictors through a regression model; and the data augmentation method uses the association between the target variable and auxiliary variables through a Bayesian model. These are the three most popular ways to use conditioning variables. Generally, hot deck method is the simplest and most intuitive way; therefore it has been used the most often in past surveys. However, it may be the least effective way of using auxiliary variables. Due to the efforts of

Rubin and many of his followers, the data augmentation method is becoming more and more popular.

The most serious nonresponse bias situation is with *confounded* missing mechanisms; that is, the probability that a datum is missing depends on the target variable itself. More formally, confounded and unconfounded missing mechanisms may be defined as:

> Let R be the set of the respondents and S be the whole sample. A response mechanism $q(\cdot \mid S)$ is said to be *unconfounded* if it is of the form $q(R \mid S) = q(R \mid X_S)$; that is, it depends on the auxiliary variables only, and the response probabilities satisfy $P(k \in R \mid S)$ for all units $k \in S$. If it depends on y-values as well, then it is called *confounded*.

An unconfounded missing mechanism will become MAR if all auxiliary variables related to response probabilities are used as conditioning variables. A confounded missing mechanism can never become MAR.

With a confounded missing mechanism, it is generally impossible to completely eliminate nonresponse biases unless the confounded missing mechanism is known. Unfortunately, the missing mechanism is never known in real applications.

Rancourt, Lee, and S@rndal (1994) discussed several estimators designed to correct nonresponse biases for data imputed via a ratio imputation method. These estimates along with the ratio estimator and the observed-data-based estimator are compared via a simulation study in terms of bias, MSE (mean square error) and coverage rate for a variety of missing mechanisms. Their results are summarized as follows.

Suppose that the data have been imputed via the ratio imputation method. The target variable is y and the fully observed auxiliary variable x is used to impute y. The whole sample S consists of n units with r respondents and $m = n\text{-}r$ nonrespondents. The estimate of the population mean based on the observed values only is

$$\bar{y}_r = \frac{1}{r}\sum_{k=1}^{r} y_k \ .$$

The standard ratio imputation estimate is given by

$$\bar{y}_{rimp} = \frac{1}{n}\left(\sum_{k=1}^{r} y_k + \sum_{j=1}^{m} y_j^*\right) = \frac{\bar{y}_r}{\bar{x}_r}\bar{x}_S \ ,$$

where y_j^* represents the imputed value for the j-th missing case, and $\bar{x}_S$ is the mean of x over the whole sample S.

Under the ideal missing mechanism MCAR, $\bar{y}_r$ is unbiased and $\bar{y}_{rimp}$ is approximately unbiased. Under unconfounded missing mechanisms where missing probabilities only depend on x , $\bar{y}_r$ is generally biased but $\bar{y}_{rimp}$ is unbiased. If the missing mechanism is confounded, both $\bar{y}_r$ and $\bar{y}_{rimp}$ are generally biased. Rancourt, Lee, and S@rndal suggest using

$$\bar{y}_{crimp} = \bar{y}_r \left[1 + (1 - \frac{r}{n})(C \frac{\bar{x}_m}{\bar{x}_r} - 1) \right]$$

to correct the biases for the ratio imputation estimator when the response mechanism is confounded. When C=1, $\bar{y}_{crimp}$ becomes the ratio imputation estimator $\bar{y}_{rimp}$. With correction factor $C = \frac{\bar{y}_m}{\bar{x}_m} \Big/ \frac{\bar{y}_r}{\bar{x}_r}$, it becomes unbiased, but it is obviously unestimable since $\bar{y}_m$ is not known. The eight correction factors C were considered by Rancourt, Lee, and S@rndal (1994):

$$C_1 = \frac{\bar{x}_m}{\bar{x}_r},\ C_2 = \frac{\bar{x}_m}{\bar{x}_S},\ C_3 = \frac{\bar{w}_m}{\bar{w}_r},\ C_4 = \frac{\bar{w}_m}{\bar{w}_S},$$

and

$$K_i = 1 - (C_i^2 - 1)(\hat{R}_{xy}^2 - 1),\ i=1, 2, 3, 4,$$

where w_k corresponding to the rank of x_k. The K_i takes into account the correlation between x and y. The correction factors C_1, C_3, K_1, and K_3 are based on the observed data only, while the correction factors C_2, C_4, K_2, and K_4 are based on the whole sample S. Therefore, for the convenience of description, $\bar{y}_{crimp}$ with C_1, C_3, K_1, or K_3 was called the *r-corrected* estimate, while $\bar{y}_{crimp}$ with C_2, C_4, K_2, and K_4 was called the *S-corrected* estimate.

In their simulation study, Rancourt, Lee, and S@rndal chose

$$y_k = a + bx_k + cx_k^2 + e_k,\ E(e_k) = 0, \quad V(e_k) = d^2 x_k$$

as simulation populations. Different types of populations are formed by setting the constants a, b, and c to different values:

(1) RATIO: a=0, c=0;
(2) CONCAVE: a=0, c<0 (c= -0.01 in the simulation);
(3) CONVEX: a=0, c>0 (c=0.01 in the simulation);
(4) NONRATIO: $a \neq 0$, b>0, c=0.

Three correlation levels r_{xy} = 0.7, 0.8, and 0.9 were obtained by a suitable value of d. Therefore, a total of 12 populations were considered: three RATIO, three CONCAVE, three CONVEX, and three NONRATIO with correlation levels 0.7, 0.8, and 0.9, respectively.

Five missing mechanisms were used in the simulation study:

(M1) Uniform response (MCAR);

(M2) The nonresponse probability is a decreasing function of x_k specified as $\exp(-gx_k)$. This is an unconfounded mechanism.

(M3) The nonresponse probability is an increasing function of x_k specified as $1-\exp(-gx_k)$. This is also an unconfounded mechanism.

(M4) The nonresponse probability is a decreasing function of y_k specified as $\exp(-gy_k)$. This is a confounded mechanism.

(M5) The nonresponse probability is an increasing function of y_k specified as $1-\exp(-gy_k)$. This is also a confounded mechanism.

The smaller units will be underrepresented in the response set R for (M2) and (M4), while the larger units will be underrepresented in the response set R for (M3) and (M5). The constant g is determined such that the average nonresponse rate is equal to one of the values 10 percent, 20 percent, 30 percent, and 40 percent.

The ten estimates were compared in terms of bias, mean square error, and coverage rate of the 95 percent confidence intervals. The primary findings are:

(1) The r-corrected estimators (using C_1, C_3, K_1, K_3) performed very poorly since the correction only used the observed data for x;

(2) For uniform response mechanism (M1), both uncorrected estimators $\bar{y}_r$ and $\bar{y}_{rimp}$ have better performance than the corrected estimators. But the loss is not very severe by mistakenly using the correction when it is not necessary for uniform nonresponse;

(3) For unconfounded missing mechanisms (M2) and (M3), the ratio imputation estimator $\bar{y}_{rimp}$ has the best performances for RATIO, CONCAVE and NONRATIO populations, while the S-corrected estimators have the best performances for the CONVEX population;

(4) For confounded mechanisms (M4) and (M5), $\bar{y}_{rimp}$ is better than the S-corrected estimators for CONCAVE and NONRATIO populations, but the S-corrected estimators are better than $\bar{y}_{rimp}$ for RATIO and CONVEX populations;

(5) The observed-data based estimator $\bar{y}_r$ performs poorly for all nonuniform response mechanisms. All estimators perform poorly for CONVEX populations with the (M5) response mechanism.

All in all, the correction to the ratio imputation estimator is not a great success in this study. Correction with observed data of x (r-corrected estimators) should never be recommended. We will generally benefit from the S-corrected estimators with CONVEX populations.

Chapter 4 Variance Estimation and Multiple Imputation

One of the most common criticisms on the use of imputation for missing data is that it leads to underestimated variances. Generally, deterministic single imputation more seriously underestimates variances than random single imputation does. Rubin (1987) sees it as a disadvantage of single imputation that "… the one imputed value cannot in itself represent uncertainty about which value to impute: If one value were really adequate, then that value was never missing. Hence, analyses that treat imputed values just like observed values generally systematically underestimate uncertainty, even assuming the precise reasons for nonresponse are known." In Rubin's opinion, multiple imputation is needed to obtain "proper" variance estimates.

However, Rao (1996) cites some disadvantages of multiple imputations:

- significantly higher costs of storage and processing of multiple data sets;
- general ABB methods for generating proper imputations that accommodate issues of clustering, stratification, and weighting to compensate for unequal probabilities of selection are not currently available;
- a small number of imputations, m, may result in a low level of precision for the multiple imputation variance estimator since the between imputation variance based on m-1 degrees of freedom may be poorly estimated.

This chapter summarizes and discusses three types of variance estimation methods for imputed survey data. Section 4.1 discusses the method proposed by S@rndal (1992) which attempts to add imputation variances to the overall variance estimates without performing multiple imputation. Section 4.2 describes the application of jackknife variance estimation methods for imputed data (Rao1996; Fay 1996). Inference based on multiply imputed data is discussed in section 4.3.

4.1 Add imputation variance without multiple imputation

S@rndal (1992) tries to correct underestimated variances by adding the component of imputation variance to the sample variance for data imputed via a single imputation procedure.

Suppose U is the population (N units), S is the sample (n units), and R is the respondents (r units). Denote the true value of the total by t, the estimate based on the complete data by $\hat{t}$, and the estimate based the imputed data by $\hat{t}_\bullet$ (obtained via the same formula as $\hat{t}$). Our interest is the variance of $\hat{t}_\bullet$ since $\hat{t}_\bullet$ is the actual estimate used in the inference.

The total error of $\hat{t}_\bullet$ can be decomposed as

$$\hat{t}_\bullet - t = (\hat{t}_\bullet - \hat{t}) + (\hat{t} - t) = \text{imputation error + sampling error.}$$

We define the imputation residual as $e_k = y_k - y_k^*$, which can not be observed for a unit $k \in S - R$. Then the imputation error becomes $\hat{t}_\bullet - \hat{t} = -\sum_{k \in S-R} w_k e_k$.

The model-assisted approach considers three different distributions, one is "with respect to the imputation model" (indicated by ξ), the second one is "with respect to the sampling design" (indicated by *S*), the third one is "with respect to the response mechanism, given *S*" (indicated by *R*). The estimator $\hat{t}_\bullet$ is overall unbiased in the sense that $E_x E_S E_R(\hat{t}_\bullet - t) = 0$ if two conditions hold:

(a) order of the expectations can be changed: $E_x E_S E_R(\cdot) = E_S E_R E_x[\cdot|S, R]$;
(b) imputation residuals have zero model expectation: $E_x(e_k) = 0$.

Condition (a) is satisfied if the response mechanism is one that may depend on S and on auxiliary data, but not on the *y*-values.

The overall variance of an unbiased estimator $\hat{t}_\bullet$ is

$$V_{tot} = E_x E_S E_R[(\hat{t} - t) + (\hat{t}_\bullet - \hat{t})]^2 = E_x V_p + E_S E_R V_x \equiv V_{sam} + V_{imp},$$

where $V_p = E_S(\hat{t} - t)^2$ is the design-based variance of $\hat{t}$, and $V_x = E_x[(\hat{t}_\bullet - \hat{t})^2|S,R]$ is the conditional model-based imputation variance. In the above equation, we ignore the cross-product term. The argument for obtaining the sample variance $\hat{V}_{sam}$ and the imputation variance $\hat{V}_{imp}$ is as follows:

(i) $\hat{V}_{sam}$: Let $\hat{V}_p$ be the standard estimator of the design variance for a complete data set, and $\hat{V}_{\bullet p}$ is the quantity obtained via the same formula for $\hat{V}_p$ using the imputed data. Evaluate the conditional expectation $E_x(\hat{V}_p - \hat{V}_{\bullet p}|S, R) = V_{dif}$, and find a model unbiased estimator $\hat{V}_{dif}$ for V_{dif}, which will usually require the estimation of certain parameters of the model ξ.
(ii) $\hat{V}_{imp}$: Find a model unbiased estimator $\hat{V}_x$ for V_x , which may again require the estimation of unknown parameters of the model ξ. Then $\hat{V}_x$ is overall unbiased for the imputation variance V_{imp}.

Note that the role of $\hat{V}_{dif}$ is to correct for the fact that the data after imputation may display "less than natural" variation. This often happens when the imputed values equal the predicted value from a fitted regression, that is, "the value on the line". The variation around the line is not reflected in the predicted value. As shown for the ratio imputation method, if residuals and predicted values are used as imputed values, $\hat{V}_{dif}$ is no longer needed to be added to the sample variance estimator.

Here is a simple example. Suppose the sample S is drawn with SRSWOR and the response mean $\bar{y}_R$ is imputed for all missing values. The corresponding imputation model ξ states that $y_k = b + e_k$, where the e_k are uncorrelated error terms with $E_x(e_k) = 0$, $V_x(e_k) = s^2$. Then

$$\hat{t}_\bullet = N\bar{y}_R,$$

$$\hat{V}_p = N^2(1/n - 1/N)\sum_S (y_k - \bar{y}_S)^2/(n-1) \equiv N^2(1/n - 1/N)S_{yS}^2$$

$$\hat{V}_{\bullet p} = N^2\left(\frac{1}{n} - \frac{1}{N}\right)\sum_R (y_k - \bar{y}_R)^2/(n-1) \equiv N^2\left(\frac{1}{n} - \frac{1}{N}\right)\frac{r-1}{n-1}S_{yR}^2$$

Since $E_x S_{yS}^2 = E_x S_{yR}^2$, $E_x(\hat{V}_p - \hat{V}_{\bullet p}|S,R) = N^2(1/n - 1/N)(n-r)/(n-1)E_x S_{yR}^2$.
Therefore, $\hat{V}_{dif} = N^2(1/n - 1/N)(n-m)/(n-1)S_{yR}^2$ is a model unbiased estimator for V_{dif}, which gives

$$\hat{V}_{sam} = \hat{V}_{\bullet p} + \hat{V}_{dif} = N^2(1/n - 1/N)S_{yR}^2,$$

Since

$$\begin{aligned}
V_{imp} = E_x(\hat{t}_\bullet - \hat{t})^2 &= \left(\frac{N}{n}\right)^2 E_x\left(\sum_{S-R}(y_k - \bar{y}_R)\right)^2 = \left(\frac{N}{n}\right)^2 (n-r)^2 E_x(\bar{y}_{S-R} - \bar{y}_R)^2 \\
&= \left(\frac{N}{n}\right)^2 (n-r)^2\left[E_x\bar{y}_{S-R}^2 + E_x\bar{y}_R^2 - E_x[\bar{y}_{S-R}\bar{y}_R]\right] = \left(\frac{N}{n}\right)^2 (n-r)^2\left[\frac{s^2}{n-r} + \frac{s^2}{r}\right], \\
&= N^2(1/r - 1/n)s^2
\end{aligned}$$

we have $\hat{V}_{imp} = N^2(1/r - 1/n)S_{yR}^2$. Therefore, $\hat{V}_{tot} = \hat{V}_{sam} + \hat{V}_{imp} = N^2(1/r - 1/N)S_{yR}^2$.

The following table shows the contribution of each variance component to the total variance for SRSWOR using the mean imputation method for three different missingness rates. Note when the missing rate is 30 percent, the variance based on the imputed value only accounts for 49 percent of the total variance, while the variance due to imputation accounts for another 30 percent. Thus 21 percent of the total variance needs to be added to the sampling variance.

Table 4.1.1—Contribution of each variance component to the total variance for the SRSWOR sampling with the mean imputation method

Missing rate in percentage	Contribution (in percentage) to $\hat{V}_{tot}$		
100(1-*r/n*)	$\hat{V}_{\bullet p}$	$\hat{V}_{dif}$	$\hat{V}_{imp}$
10	81	9	10
20	64	16	20
30	49	21	30

The analytical formulas for SRSWOR sampling with the ratio imputation method has also been derived in S@rndal (1992).

As a comment on this approach, it is very convenient that imputation variance can be estimated without performing multiple imputation and , therefore, there is no need for a great deal of storage space and processing time which multiple imputation requires. The variance estimates obtained through this method may be more accurate than those obtained through a small number of multiple imputations since a small number of multiple imputations may lead to poor between-imputation variance estimation. However, S@rndal (1992) only derived analytical formulas for two simple cases: SRSWOR sampling with the mean and ratio imputation. For a more complex survey design and/or more complicated imputation algorithms, the derivation is not trivial and may be impossible. It will be even more difficult to apply the method to nonlinear statistics such as median, quartile, ratio, etc. Furthermore, this method only takes care of variance estimates. It seems arduous to adjust for covariance via this method.

To make this method more attractive, random imputation methods should be used instead of deterministic imputation methods, because deterministic imputation methods not only distort the distribution of data, but also require extra effort to estimate V_{dif}.

4.2 Jackknife variance estimation with imputed data

Rao (1996) and Fay (1996) extended the jackknife variance estimation method to imputed survey data. Rao (1996) discussed the jackknife method for imputed survey data for two situations: (1) stratified random sampling with ratio imputation and regression imputation; (2) stratified multistage sampling with cell mean imputation and weighted hot deck imputation. Fay (1996) applied the jackknife method to imputed data via fractionally weighted imputation.

4.2.1 Jackknife variance estimation with imputed data for stratified random sampling

Rao (1996) expanded the jackknife variance estimation method to imputed survey data collected with a stratified random sampling design. Let n_h be the sample size and N_h be the population size for the h-th stratum (h=1, 2, …L). In case of complete data, a design-unbiased

(p-unbiased) estimator of population mean is given by $\bar{y} = \sum_{h=1}^{L} W_h \bar{y}_h$, where $W_h = N_h / \sum N_h$ is the weight for stratum h and $\bar{y}_h$ is the h-th stratum sample mean. The jackknife variance estimator is given by

$$v_J(\bar{y}) = \sum_{h=1}^{L} \frac{n_h - 1}{n_h} \left(1 - \frac{n_h}{N_h}\right) \sum_{j=1}^{n_h} (\bar{y}^{h(-j)} - \bar{y})^2 ,$$

where $\bar{y}^{h(-j)}$ is the jackknife sample mean obtained by deleting the j-th observation from the h-th stratum.

In presence of nonresponses, let A_{rh} and A_{mh} be the sample of respondents and nonrespondents in that stratum. The jackknife sample mean $\bar{y}^{h(-j)a}$ can be adjusted in the following way: (1) under deterministic imputation, if a respondent is left out, all the imputed values should be adjusted by the amount $y_{hi}^{*(-j)} - y_{hi}^{*}$, where $y_{hi}^{*(-j)}$ is the value that one would impute for the i-th nonrespondent if the j-th respondent is deleted in the h-th stratum; (2) under stochastic imputation, if a respondent is excluded, each of the imputed values in stratum h should be adjusted by an average amount $E_*^{(-j)} y_{hi}^* - E_* y_{hi}^*$, where E_* denotes expectation with respect to the imputation procedure given the donor set and $E_*^{(-j)}$ is the expectation with respect to the imputation procedure when the donor set is modified by excluding unit j. Then the jackknife variance estimator with imputed data is given by

$$v_J(\bar{y}) = \sum_{h=1}^{L} \frac{n_h - 1}{n_h} \left(1 - \frac{n_h}{N_h}\right) \sum_{j=1}^{n_h} (\bar{y}^{h(-j)a} - \bar{y}_I)^2 ,$$

where $\bar{y}_I = \sum_{h=1}^{L} W_h \left(\sum_{A_{rh}} y_{hi} + \sum_{A_{mh}} y_{hi}^* \right) / n_h$ is the overall sample mean with imputed data.

The following two examples apply this technique to ratio imputation and regression imputation.

Example 1 (ratio imputation). Suppose that an auxiliary variable x closely related to an item y is observed on all sample units. Ratio imputation uses $y_{hi}^* = \frac{\bar{y}_{rh}}{\bar{x}_{rh}} x_{hi}$ as imputed values for the i-th nonrespondent in the h-th stratum. Under this deterministic imputation procedure, if j-th respondent is excluded in the jackknife variance estimation, the imputed value will be $y_{hi}^{*(-j)} = \left(\bar{y}_{rh}^{(-j)} / \bar{x}_{rh}^{(-j)}\right) x_{hi}$.

A stochastic counterpart of ratio imputation adds the donors' residuals to the above ratio imputed values. Under this imputation approach, $E_* y_{hi}^* = \left(\bar{y}_{rh} / \bar{x}_{rh}\right) x_{hi}$ and $E_*^{(-j)} y_{hi}^* = \left(\bar{y}_{rh}^{(-j)} / \bar{x}_{rh}^{(-j)}\right) x_{hi}$. Thus the adjusted imputed values are given by $y_{hi}^* + \left(\bar{y}_{rh}^{(-j)} / \bar{x}_{rh}^{(-j)}\right) x_{hi} - \left(\bar{y}_{rh} / \bar{x}_{rh}\right) x_{hi}$.

Example 2 (regression imputation). Again assume that x is observed on all sample units. Linear regression imputation uses $y_{hi}^{*} = \bar{y}_{rh} + \hat{b}_{rh}\left(x_{hi} - \bar{x}_{rh}\right)$, where $\hat{b}_{rh}$ is the ordinary least square regression coefficient based on the respondents in stratum h. Under this deterministic imputation procedure, when the j-th respondent is deleted in the jackknife variance estimation, the imputed values will be $y_{hi}^{*(-j)} = \bar{y}_{rh}^{(-j)} + \hat{b}_{rh}^{(-j)}\left(x_{hi} - \bar{x}_{rh}^{(-j)}\right)$, where $\hat{b}_{rh}^{(-j)}$ is the least squares regression coefficient when the j-th respondent is deleted.

A stochastic counterpart of regression imputation adds a donor's residual to the above imputations, where the donor is selected through a simple random sampling. Under this approach, we have $E_{*}^{(-j)} y_{hi}^{*} = \hat{y}_{hi}$ and $E_{*}^{(-j)} y_{hi}^{*} = \hat{y}_{rh}^{(-j)} = \bar{y}_{rh}^{(-j)} + \hat{b}_{rh}^{(-j)}\left(x_{hi} - \bar{x}_{rh}^{(-j)}\right)$. Thus the adjusted imputed values are given by $y_{hi}^{*} + \hat{y}_{rh}^{(-j)} - \hat{y}_{rh}$ if the j-th respondent is deleted and remain unchanged if the j-th non-respondent is deleted.

In these two examples, the imputed estimators of mean are approximately design-unbiased under uniform response within each stratum, as well as design model unbiased under their super-population models (defined in sections 1.3.1 and 1.3.2). The jackknife variance estimators are p-consistent, as well as approximately design model unbiased under their super-population models.

Rao (1996) also discussed jackknife variance estimation for stratified multistage sampling design with missing data imputed by the class mean imputation method and the weighted within-class hot deck method. We omit them here because they are parallel to the two examples given above. Linearized versions of the jackknife variance estimators, which are useful with computer programs that use the linearization method of variance estimation (e.g., SUDAAN), are also provided in that paper.

However, as Judkins (1996) pointed out, this jackknife method is essentially a univariate tool with well behaved extensions only for variables that are either never missing or are missing or present in whole blocks. It has only been applied to simple statistics such as total, mean or functions of total or mean under marginal imputation. For more complex statistics, such as regression and correlation coefficients, marginal imputation often attenuates the association between variables. Joint imputation from the same donor, called *common donor hot deck*, may be used sometimes to alleviate this problem with marginal imputation when a record has several missing related values. This method preserves bivariate relationships only when both variables are missing; that is, when there are no partial nonrespondents with respect to the two variables.

4.2.2 Jackknife variance estimation with fractionally weighted imputation

Fay (1996) discussed the application of the jackknife variance estimation method to survey data imputed through the *fractionally weighted imputation* (FWI) method. FWI creates one set of imputations by fractionally weighting m sets of imputations. In general, FWI assigns a weight

1/m to each of the *m* imputations. If the original analysis is weighted, then the *m* imputed values each receive *1/m* times the original weight.

Let A_r and A_{nr} be the sample of respondents and nonrespondents, respectively, *n* be the total sample size, and *r* be the number of respondents. For any data imputed via a single imputation method, the mean may be estimated by

$$\bar{y} = \left(\frac{r}{n}\right)\bar{y}_r + \left[1 - \frac{r}{n}\right]\bar{y}^*_{nr},$$

where $\bar{y}_r$ and $\bar{y}^*_{nr}$ are the mean of the reported values of the respondents and the mean of imputed values for the nonrespondents respectively. The standard jackknife variance estimator is

$$v_J = \frac{n-1}{n}\sum_{j=1}^{n}\left(\bar{y}^{(-j)} - \bar{y}\right)^2$$

where

$$\bar{y}^{(-j)} = \begin{cases} 1/(n-1)[n\bar{y} - y_j] & if \quad j \in A_r \\ 1/(n-1)[n\bar{y} - y^*_j] & if \quad j \in A_{nr} \end{cases}.$$

This naïve jackknife variance estimate treats the imputed values as true observed values. Rao and Shao (1992) modified this jackknife mean by

$$\bar{y}^{(-j)a} = \begin{cases} \dfrac{1}{n-1}[r\bar{y}_r - y_j + \sum_{i \in A_{nr}}(y^*_i + \bar{y}^{(-j)}_r - \bar{y}_r] & if \quad j \in A_r \\ 1/(n-1)[n\bar{y} - y^*_j] & if \quad j \in A_{nr} \end{cases},$$

where $\bar{y}^{(-j)}_r = (r\bar{y}_r - y_j)/(r-1)$ is the mean of the $(r$-1) respondents without jth observation. This formula reflects that, when a respondent is deleted, each imputed value y^*_i need to be adjusted by the amount of $(\bar{y}^{(-j)}_r - \bar{y}_r)$ since we only have r-1 respondents for imputation when jth respondent is left out. For example, for the mean imputation method, the originally imputed values $y^*_i = \bar{y}_r$ for all $i \in A_{nr}$, and then the adjusted imputed value is $\bar{y}^{(-j)}_r$ when jth respondent is left out.

For fractionally weighted imputations, the mean may be estimated by

$$\bar{y}_{(FWI)} = \frac{1}{n}\left[r\bar{y}_r + \sum_{j \in A_{nr}}\sum_{l=1}^{m}\frac{1}{m}y^*_{jl}\right],$$

where y^*_{jl} is lth imputation for jth missing value. The Rao-Shao type jackknife variance estimate may be constructed by replacing $\bar{y}^{(-j)a}$ with

$$
\bar{y}_{FWI}^{(-j)a} = \begin{cases} \frac{1}{n-1}[r\bar{y}_r - y_j + \sum_{i \in A_{nr}} \sum_{l=1}^{m} \frac{1}{m}(y_{il}^* + \bar{y}_r^{(-j)} - \bar{y}_r)] & if \quad j \in A_r \\ \frac{1}{n-1}[n\bar{y}_{(FWI)} - \sum_{l=1}^{m} \frac{1}{m} y_{jl}^*] & if \quad j \in A_{nr} \end{cases}.
$$

Fay (1996) claimed that "unlike MI (multiple imputation), the RS (Rao-Shao type) variance estimator does not use variation among the *m* different imputed sets.…Because the effect of missing data is incorporated in the variance calculation as a whole, instead of isolated…for MI, it is generally unnecessary to reference a t distribution to obtain adequate approximation for construction of confidence intervals" (p. 492).

In some situations, Rubin's multiple imputation (non-proper MI) inference may have inconsistent variance estimates. A modified version of Rao-Shao type jackknife variance estimate may be used:

$$
v_{J(MI)} = \frac{n-1}{n}\left[\sum_{j=1}^{n}(\bar{y}_{(MI)}^{(-j)a} - \bar{y}_{(MI)})^2 + \sum_{j=1}^{n}\frac{1}{m}(\bar{y}_{(MI)}^{(-j)m} - \bar{y}_{(MI)})^2\right],
$$

where

$$
\bar{y}_{MI}^{(-j)a} = \begin{cases} \frac{1}{n-1}[r\bar{y}_r - y_j + \sum_{i \in A_{nr}} \sum_{l=1}^{m} \frac{1}{m}(y_{il}^* + \bar{y}_r^{(-j)} - \bar{y}_r)] & if \quad j \in A_r \\ \frac{1}{n-1}[n\bar{y}_{(MI)} - \sum_{l=1}^{m} \frac{1}{m} y_{jl}^*] & if \quad j \in A_{nr} \end{cases}
$$

and

$$
\bar{y}_{MI}^{(-j)m} = \begin{cases} [\bar{y}_{(MI)} + \frac{1}{n-1}\sum_{i \in A_{nr}} \bar{y}_r^{(-j)} - \bar{y}_r] & if \quad j \in A_r \\ \bar{y}_{(MI)} & if \quad j \in A_{nr} \end{cases}.
$$

In this jackknife variance estimate, the first sum of squares are usual jackknife terms, and the second sum of squares are designed to capture the variations usually added by the proper multiple imputations.

Fay (1996) points out, "FWI resembles MI but may be distinguished by (a) the manner in which the imputations are made, (b) the procedures to obtain the estimates from the data set, and (c) the variance estimation and analysis of the resulting data set" (p. 492).

Some anomalies given by Fay demonstrate that MI does not address effectively for some relatively simple situations. This is not surprising because, as Judkins (1996) pointed out that " Fay's fractionally weighted imputation (FWI) can be expected to yield true variance no larger than multiple imputation with the same number of replicates" (p. 508). Based on his finding, Fay suggests that researchers implement Monte Carlo studies to examine the performance characteristics of MI to develop a body of systematic evidence before applying it to specific problems.

However, Fay’s FWI method is subject to the same limitation as the Rao’s jackknife described in the preceding section; that is, it is basically a univariate tool and hard to extend to the multivariate case. Rubin (1996) further criticizes the limitation of Fay’s FWI method: “Fay’s approach is essentially constrained to the special situation where (a) there is the simplest pattern of nonresponse (i.e., there are respondents with no missing data and nonrespondents with all outcome variables missing), (b) hot-deck draws (possibly weighted) are made from each adjustment cell to impute donor values to nonrespondents, (c) there are effectively an unlimited number of respondent donors in each adjustment cell, and (d) the adjustment cell classification and design weights are assumed to control adequately for nonresponse biases for all estimands of interest. Since hot-deck classification is based on observed variables, Fay’s approach implicitly assumes an ignorable nonresponse mechanism, because otherwise (d) is violated” (p. 515).

4.3 Multiple imputation inference

The discussion in this section is based on Rubin (1996).

4.3.1 Objectives of imputations

The basic objective of imputation is to allow ultimate data users to apply their existing analysis tools to any dataset with missing values using the same command structure and output standards as if there were no missing data. Certain ad hoc methods of handling missing data, such as “complete-case analysis,” “available-case analysis,” and “fill-in with means” satisfy this basic objective and so have a certain appeal.

The ideal supplemental objective of imputation is that each complete-data statistical tool can be applied to each incomplete dataset to obtain the same inference as if the dataset had no missing values. This objective is obviously unachievable no matter what imputation method is used. It is analogous to saying that the objective of a survey is to obtain the same answer as a complete census.

A less-ideal achievable supplemental objective could be as follows. Assuming that the ultimate user’s complete-data analysis is statistically valid for a scientific estimand, the answer that results from applying the same analysis method to an incomplete-data remains statistically valid for the same scientific estimand assuming the truth of the database constructor’s posited model for missing data. This supplemental goal can be achieved through some imputation methods, but can not be achieved through others.

Before we discuss multiple imputation inference, let’s first clarify the meanings of *scientific estimands* and *statistical validity.*

Scientific Estimands: Quantities of scientific interest that can be calculated in the population and do not change its value depending on the data collection design used to measure them (i.e., they does not vary with sample size and survey design, or the number of nonrespondents or follow-up efforts). For example, scientific estimands include population means, variances, correlations, factor loadings, regression coefficients, but exclude the sampling variance of a sample mean under a particular sampling plan and the expectation of the complete-data sample mean when missing values are filled in with zero or the observed sample means.

Statistically Validity: This must be a frequency concept, averaging over randomization distributions generated by known sampling mechanisms and posited distribution for the response mechanisms. Bayesian validity is also important, but is far more difficult to achieve in this context because it requires far more compatibility between the database constructor and the analyst.

First and foremost, to achieve statistical validity for scientific estimands, point estimation must be approximately unbiased for the scientific estimands, averaging over the sampling and the posited nonresponse mechanisms. Second, interval estimation and hypothesis testing must be valid in the sense that nominal levels describe operating characteristics over sampling and posited response mechanisms. There are two versions of frequentist validity for nominal levels: *randomization validity* and *confidence validity*. Randomization validity means that, for interval estimates, the actual interval coverage equals the nominal interval coverage, and for tests of hypotheses, the actual rejection rate equals the nominal rejection rate. Confidence validity means that, for interval estimates, the actual coverage rate is greater than or equal to the nominal coverage rate, and for tests of hypotheses, the actual rejection rate is less than or equal to the nominal rejection rate. Confidence validity is a more generally achievable objective.

To express the concepts in mathematical equations, let X be the array of all background information fully observed in a population and Y be the array of outcome information in the population that is to be sampled in the survey. $Q = Q(X, Y)$ is a scientific estimand. Suppose $\hat{Q}$ is a complete-data estimate of Q with sampling variance consistently estimated by the statistic U. Then randomization validity with complete-data is equivalent to

$$E(\hat{Q} \mid X, Y) \cong Q \quad \text{(unbiasedness of point estimate)}$$

and

$$E(U \mid X, Y) \cong Var(\hat{Q} \mid X, Y) \quad \text{(unbiasedness of variance estimate).}$$

For confidence validity with complete data, the second condition is replaced by

$$E(U \mid X, Y) \geq Var(\hat{Q} \mid X, Y).$$

4.3.2 Multiple imputation inference

The goal of multiple imputation (sometimes also called *repeated imputation*) is to provide statistically valid inference in the difficult real-world situation where (1) ultimate users and database constructors are distinct entities with different analyses, models, and capabilities, and (2) there typically is no one accepted reason for the missing data.

Multiple imputation was designed to satisfy both the achievable basic objective and the achievable supplemental objective stated in preceding sub-section by using Bayesian and frequentist paradigms in complementary ways: the Bayesian model-based approach to **create** procedures, and the frequentist (randomization-based approach to **evaluate** procedures.

Multiple imputation is based on the following Bayesian results:

$$P(Q|Y_{obs}) = \int P(Q|Y_{obs}, Y_{mis})P(Y_{mis}|Y_{obs})dY_{mis},$$

or in words

(Actual posterior distribution of *Q*) = *AVE* (complete-data posterior distribution of *Q*),

where *AVE* (complete-data posterior distribution of *Q*) refers to the average over the repeated imputations, which are draws from $P(Y_{\mathrm{mis}} \mid Y_{\mathrm{obs}})$, which is the posterior predictive distribution of missing data given the observed data. About the first two moments, we have:

$$E(Q|Y_{obs}) = E[E(Q|Y_{obs}, Y_{mis})|Y_{obs}]$$

or in words

(Posterior mean of *Q*) = *AVE* (repeated complete-data posterior means of *Q*)

$$V(Q|Y_{obs}) = E[V(Q|Y_{obs}, Y_{mis})|Y_{obs}] + V[E(Q|Y_{obs}, Y_{mis})|Y_{obs}].$$

Suppose that we have *m* sets of repeated imputations, and the *lth* (*l*=1, 2, …, *m*) point estimate and its corresponding variance-covariance estimate based on the *lth* set of imputed data using standard formulas are (Q_{*l}, U_{*l}). Then the repeated-imputation estimate of *Q* is:

$$\overline{Q}_m = \sum_1^m Q_{*l} / m.$$

The associated variance-covariance of $\overline{Q}_m$ is:

$$T_m = \sum_1^m U_{*l} / m + \frac{m+1}{m} B_m,$$

where $\overline{U}_m = \sum_1^m U_{*l} / m$ is the within-imputation variability, and

$$B_m = \frac{1}{m-1}\sum_{l=1}^{m}(Q_{*l} - \overline{Q}_m)(Q_{*l} - \overline{Q}_m)'$$

is the between-imputation variability. We expect:

$$(Q - \overline{Q}_\infty) \sim N(0, T_\infty),$$

where $\overline{Q}_\infty = \lim_{m \to \infty} \overline{Q}_m$ and $T_\infty = \lim_{m \to \infty} T_m$.

A "proper" multiple imputation procedure treats (*X*, *Y*) and the intended sample (as indicated by *I*) as fixed, and deals with the fixed but unknown values of the complete-data statistics $(\hat{Q}, U)$ in the sample as if they were estimands. That is, the randomization distribution critically involved in the definition of proper multiple imputation is generated by the response mechanism, in which *X*, *Y*, and *I* are fixed , and the response indicator *R* is the random variable. That means a proper imputation must satisfy the followings:

$$E(\overline{Q}_\infty | X, Y, I) \cong \hat{Q} \tag{4.1}$$
$$E(\overline{U}_\infty | X, Y, I) \cong U \tag{4.2}$$
$$E(B_\infty | X, Y, I) \cong Var(\overline{Q}_\infty | X, Y, I) \tag{4.3}$$

The definition of proper concerns the situation where "population" equals complete-data sample, "estimands" equals complete-data statistics $(\hat{Q}, U)$, and "survey design" equals the posited response mechanism. The criterion is valid frequency inference, and the method for creating inferences is Bayesian predictive inference using simulated values.

It follows from (4.1)–(4.3) that, if the complete-data inference is randomization-valid and the multiple imputation procedure is proper, the infinite-*m* repeated imputation inference is randomization-valid under the posited response mechanism.

Rubin (1987, chapter 4) presented analytic results, simulation evaluations, and many examples of proper and improper multiple imputation methods, where the evaluations were all from the random-response randomization-based frequentist perspective. The trick in many of the examples of proper imputation was to get the variance condition (4.3) correct, and it was shown that when drawing imputations to approximate repetitions from a sensible Bayesian model, conditions (4.1)–(4.3) typically followed automatically. The more straightforward conditions, (4.1) and (4.2), typically were simple properties of any intelligent imputation scheme that tried to track the data. An example of a method that does not track the data is "fill in the mean," which, although it may satisfy (4.1) for $\hat{Q} = \hat{y}$, fails to do so for $\hat{Q} = s^2$ or for the 25th percentile, or to satisfy (4.2) for $U = s^2/n$, etc. Hot deck (bootstrap) and random-draw regression methods tend to satisfy (4.1) and (4.2) but fail to satisfy (4.3) until a Bayesian, systematic between-imputation component of variability is added (e.g., via the Bayesian Bootstrap), to reflect uncertainty in the estimation of population parameters.

A multiple imputation procedure is *strongly superefficient* for the complete-data statistic $\hat{Q}$ if, first, $\overline{Q}_\infty$ and $\hat{Q}$ estimate the same estimand, that is, the procedure is "first-moment proper" for $\hat{Q}$: $E(\overline{Q}_\infty | X, Y) = E(\hat{Q} | X, Y)$, and second $\overline{Q}_\infty$ has no larger variance than the complete-data estimate itself: $Var(\overline{Q}_\infty | X,Y) \le Var(\hat{Q} | X,Y)$. If the second condition is replaced by $Cov(\overline{Q}_\infty, \hat{Q} | X,Y) \le Var(\hat{Q} | X,Y)$, then it is called *superefficient imputation*. Strongly superefficient imputation implies superefficient imputation.

A multiple imputation procedure is *confidence-proper* for the complete-data statistics $(\hat{Q},U)$ if the imputations are "first-moment proper" for $(\hat{Q},U)$ and

$$E(\overline{U}_\infty | X,Y) = E(U | X,Y)$$

and if B_∞ conservatively estimates the "excess variance" of $\overline{Q}_\infty$ over $\hat{Q}$:

$$E(B_\infty | X,Y) \ge Var(\overline{Q}_\infty | X,Y) - Var(\hat{Q} | X,Y)$$

If a multiple imputation procedure is proper for $(\hat{Q},U)$ it is confidence proper for $(\hat{Q},U)$. If the complete-data inference based on $(\hat{Q},U)$ is confidence valid and the multiple imputation procedure is confidence proper for $(\hat{Q},U)$, then the repeated-imputation inference is confidence valid no matter how complex the survey design.

According to Rubin (1996), any imputation method that satisfies the validity objective in generality must not only reflect the underlying response mechanism but must also be a random draw method. Nonrandom draw methods can be applied in special cases but require special analysis techniques. Of course, the development of user-friendly appropriate software for creating multiple imputations and analyzing multiply-imputed data is still badly needed.

Rubin (1996) also advises including all variables in a multiple imputation model to make it proper in general. If *X* is correlated with *Y* but not used to multiply-impute *Y*, then the multiply-imputed dataset will yield estimates of the (*X*, *Y*) correlation biased towards zero. Thus, the danger with an imputer's model is generally in leaving out predictors rather than including too many, and the advice has always been to include as many variables as possible when doing multiple imputation. Nevertheless, because problems can occur when the imputer's model leaves out important predictor variables, the database constructor must include a description of the imputation model with the multiply-imputed database, so that ultimate users know which relationships among variables have been implicitly set to zero. This is obviously good advice in principle, but it may be difficult to do in practice.

4.3.3 Current issues concerning multiple imputation

Rubin (1996) also discussed current issues concerning multiple imputation. The first issue focuses on its implementation: operational difficulties for the database constructor and the ultimate user, as well as the acceptability of answers obtained partially through the use of simulation. The second issue concerns the frequentist validity of repeated-imputation inferences when the multiple imputations are not proper, but appear "reasonable" in some sense. Specifically, Rubin raised four questions and tried to answer them:

(1) Is multiple imputation unprincipled or unacceptable because it uses simulation?

It is critical to remember that multiple imputation does not pretend to create information through simulated values but simply to represent the observed information this way to make it amenable to valid analysis using complete-data tools. The extra noise created when using a finite number of imputations is the price to be paid for this luxury.

With multiple imputation, the simulation is only being used to handle the missing information, with reliance for handling the rest of the information left to the complete-data method, be it analytic or simulation-based. Jackknife and Bootstrap use many more simulations. More explicitly, hundreds or thousands of simulations will be needed for bootstrap or jackknife methods, whereas as few as five multiple imputations (or even three in some cases) are adequate under each model for nonresponse. The asymptotic efficiency of the repeated-imputation finite-m estimate relative to the infinite m estimate is $[1+(g\ /m)]^{-1/2}$ in units of standard deviations, which is close to one with realistic fractions of missing information γ and modest m.

(2) Is multiple imputation too much work for the user?
(3) Does it take too much work to create proper or approximately proper multiple imputations?
(4) Can repeated imputations under an appropriate Bayesian model lead to invalid inferences?

His arguments to these three questions are not very convincing and therefore are not repeated here. There are no "right" answers to questions (2) and (3). Different people may have different opinions. Regarding question (4), Fay (1996) seems to give a "yes" answer; that is, it is possible that multiple imputation under a Bayesian model may lead to invalid inferences.

Chapter 5 Simulation Study

5.1 Simulation design

The simulation design factors are described as follows.

5.1.1 Distribution

Four sets of variables were generated for the simulation study. The distribution type and name of each of the variables generated are described below.

(1) Five variables from $N(m, 1)$ denoted as Norm1, Norm2, Norm3, Norm4, Norm5 with m=1, …, 5, respectively;

(2) Five variables from a double exponential distribution denoted as Dexp1, Dexp2, Dexp3, Dexp4, and Dexp5 with means of 1, 2, 3, 4, and 5, respectively, and variances equal to 2;

(3) Five variables from mixed normal distributions (i.e., 95 percent $N(m, 1)$ and 5 percent $N(m, 3^2)$) denoted as MixNorm1, MixNorm2, MixNorm3, MixNorm4, and MixNorm5 with m=1, …, 5, respectively.

(4) Five variables from mixed normal distributions (i.e., 95 percent $N(m, 1)$ and 5 percent $c^2(4)-4+m$) denoted as MixNChi1, MixNChi2, MixNChi3, MixNChi4, and MixNChi5 with m=1, …, 5, respectively.

The first three sets of variables were symmetric about their means, while the fourth set of variables was right skewed. The five variables in each set had means of 1, 2, 3, 4, and 5 respectively. Each set of five variables were correlated with the following correlation matrix:

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

The correlation coefficients between different sets of variables were small.

5.1.2 Missing mechanism

(1) *MCAR*: Missing values in variables Norm1, Dexp1, MixNorm1, and MixNChi1 were missing completely at random (MCAR);

(2) *Tail values more likely missing (unconfounded)*: Missing values in Norm2 were created with probability of $exp(-\lambda\ |Norm1\text{-}1|)$, where λ was determined so that on average 10 percent, 20 percent, 30 percent, and 40 percent missing values were generated for the four missing rate categories under study. This was an unconfounded missing mechanism. Since Norm1 and Norm2 were positively correlated with correlation coefficient 0.9, tail values were missing with higher probabilities. Missing values in Dexp2, MixNorm2, and MixNChi2 were similarly created using Dexp1, MixNorm1, and MixNChi1;

(3) *Large values more likely missing (unconfounded)*: Missing values in Norm3 were created with probability of $exp[-\lambda\ (Norm2\text{-}2)]$, where λ was determined so that on average 5 percent, 10 percent, 15 percent, and 20 percent missing values were generated for the four missing rate categories under study. This was an unconfounded missing mechanism. Since Norm2 and Norm3 were positively correlated with correlation coefficient 0.8, large values of Norm3 were missing with higher probabilities. Missing values in Dexp3, MixNorm3, and MixNChi3 were similarly created using Dexp2, MixNorm2, and MixNChi2;

(4) *Center values more likely missing (unconfounded)*: Missing values in Norm4 were created with probability of $1\text{-}exp[-\lambda\ |Norm3\text{-}3|]$, where λ was determined so that on average 10 percent, 20 percent, 30 percent, and 40 percent missing values were generated for the four missing rate categories under study. This was an unconfounded missing mechanism. Since Norm3 and Norm4 were positively correlated with correlation coefficient 0.7, center values of Norm4 were missing with higher probabilities. Missing values in Dexp4, MixNorm4, and MixNChi4 were similarly created using Dexp3, MixNorm3, and MixNChi3;

(5) *Tail values more likely missing (confounded)*: Missing values in Norm5 were created with probability of $1\text{-}exp[-\lambda\ |Norm5\text{-}5|]$, where λ was determined so that on average 10 percent, 20 percent, 30 percent, and 40 percent missing values were generated for the four missing rate categories under study. This was a confounded missing mechanism since the probabilities of missing Norm5 depended on itself. Missing values in Dexp5, MixNorm5, and MixNChi5 were similarly created.

We use the term “one-side missing mechanism” for mechanism (3) and the term “two-side missing mechanism” for the other four mechanisms for the convenience of description.

5.1.3 Missing rates

For missing mechanisms (1), (2), (4), and (5), the four types of missing rates were 10 percent, 20 percent, 30 percent, and 40 percent, while for missing mechanisms (3), the four types of missing rates were 5 percent, 10 percent, 15 percent, and 20 percent.

5.1.4 Imputation methods

(1) *Mean Imputation* (deterministic): Missing values were replaced with the sample mean.

(2) *Ratio Imputation* (deterministic): Missing values in y were replaced by

$$y_i = \frac{\bar{y}_{obs} - 1}{\bar{x}_{obs}} x_i + 1,$$

where $\bar{y}_{obs}$ and $\bar{x}_{obs}$ were the means of the observed values for the target variable and auxiliary variables respectively. Norm1, Norm2, Norm3, and Norm4 served as auxiliary variables for Norm2, Norm3, Norm4, and Norm5, respectively.

Since the means of the target variables were one more than the means of the auxiliary variables, we subtracted 1 from the numerators of the ratios and added 1 back to the final imputed values. This means that we used ratio imputation model $E(y-1) = bx$ instead of $E(y) = bx$ because the later model led to very bad results.

We did not use ratio imputation for Norm1 since we needed to create a complete auxiliary variable to start the ratio imputation process. Because missing values in Norm1 were missing completely at random, we started with this variable and imputed its missing values using the mean with disturbance method described in (5) below.

The other three sets of five variables were imputed in the same way as the normal variables.

(3) *Sequential nearest neighbor hot deck method* (deterministic): This is also called the traditional hot deck method. To impute any one of the five variables in each set, the data were first sorted by the other four variables of that set. The observed mean served as the starting stored value. Then the sequential imputation process started to check each record in the sorted data file. If a record had a response for the target variable, the stored value was updated by this new response value; if a record missed the target variable, the currently stored value would serve as the imputation value.

(4) *Random imputation method* (random): Randomly drew imputations from the observed values (with replacement).

(5) *Mean imputation with disturbance* (random): Random disturbances drawn from $N(0, s^2)$ were added to the mean imputation (1), where s^2 is the sample variance.

(6) *Ratio imputation with disturbance* (random): Random disturbances were drawn from $N(0, s^2)$ were added to the ratio imputation (2), where s^2 is the sample variance.

(7) *Approximate Bayesian Bootstrap (ABB) method* (random): First drew *r* values randomly with replacement from the observed values $Y_1, \ldots, Y_r$ to create Y^*_{obs}, and then drew *m* values randomly with replacement from Y^*_{obs} for imputation, where *r* and *m* were the number of observed values and that of missing values.

(8) *Bayesian Bootstrap (BB) method* (random): First, drew *r-1* uniform random numbers between 0 and 1, and let their ordered values be $a_1, \ldots, a_{r-1}$; also let a_0=0 and a_r=1, where *r* was the number of respondents. Then, drew each of the *m* missing values by drawing from $Y_1, \ldots, Y_r$ with probabilities $(a_1 - a_0)$, $(a_2 - a_1)$, …, $(1 - a_{r-1})$; that is, independently *m* times, drew a uniform random number *u*, and imputed Y_i if $a_{i-1} < u \le a_i$ (*i*=1, 2, …, *r*).

(9) *PROC IMPUTE* (random): First, used a stepwise regression approach to find the best regression equations and then used the predicted regression values to form the "optimal" imputation classes. Then, for each missing record, two observed values were drawn and weighted to form the imputation value. One of the two observed values were drawn according to the estimated distribution of the observed values from its own imputation class and the other from the nearest imputation class.

(10) *Data Augmentation* (random): This Bayesian iterative method assumed two distributions: the distribution of the data and the prior distribution of the parameters. The imputation process consisted of two steps: (i) *I*-step: with current parameter estimates, drew imputations for the missing values from the predicted distribution of the data; (ii) *P*-step: with both the observed data and the imputed values of the missing data, drew parameter estimates from their posterior distribution. To start this iterative process, we may use the EM algorithm to obtain initial parameter estimates for the first *I*-step. *Schafer's software* was used to implement this method in our simulation. This software assumes multivariate normal distribution for the data, and normal prior for the parameters of means and normal-inverted Wishart for the variance-covariance parameters.

(11) *Adjusted data augmentation method* (random): If the normality assumption for the continuous data in Schafer's software 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}. We can accomplish this as follows. First m^* and s^{*2} were drawn in the same way from their posterior distributions as in Schafer's software. Then the components of m-dimensional vector $X = (X_1, \ldots, X_m)$ were drawn with replacement from Y_{obs}. Under repeated draws from Y_{obs}, the standardized variable

$$Z_i = (X_i - \bar{y}_r) / \sqrt{(r-1)s_r^2 / r}$$

had expected value 0 and variance 1. Finally, the m components of Y_{mis} were set equal to $m^* + s^* Z_i$, i=1, 2, …., m.

For each combination formed by the above simulation factors, 200 replicate runs were performed. We assessed the imputation methods based on their average performance over the 200 replications. The sample size for each replicate data set was 100.

5.2 Simulation results

We compared the imputation methods in terms of bias of parameter estimates (mean, median, first and third quartiles), bias of variance estimates (single and multiple imputations), coverage probability, confidence interval width, and average imputation error. Analyses and conclusions according to each criterion based on the simulation results follow. The detailed simulation results are presented in tables 5.2.1.1–5.2.7.5.

5.2.1 Bias of population mean estimates

Tables 5.2.1.1–5.2.1.5 present the biases of population mean estimates for the 11 imputation methods under study. Table 5.2.1.1 combines the four missing rate categories with overall missing rates of around 25 percent for missing mechanisms (1), (2), (4), and (5), and about 10 percent for missing mechanism (3). The remaining four tables describe the biases for missing rate categories 10 percent, 20 percent, 30 percent, and 40 percent. The numbers of missing values for one-side missing mechanism (3) are about half of those for the other four two-side missing mechanisms.

For symmetric distributions (normal, double exponential, and mixed normal) and two-side missing mechanisms, the population mean estimates based on the incomplete data are theoretically unbiased. Therefore, the values in the first three rows in each block except block 3 of tables 5.2.1.1–5.2.1.5 are all pretty close to zero. For these cases, it does not make much sense to compare the imputation methods in terms of improvement of biases.

When large values are more likely to be missing, block 3 of table 5.2.1.1 shows that the negative biases caused by missing values, which are the same as those for the mean imputation method, are considerable for all four types of distributions although there are only about 10 percent missing values. As the distributions depart further from normal, the biases become more and more serious. The ratio imputation method, ratio imputation with disturbance method, and Schafer's software perfectly corrected the biases. PROC IMPUTE and the sequential nearest neighbor hot deck method improved the biases substantially, but PROC IMPUTE has a significant advantage over the hot deck method. Since the adjusted data augmentation method introduces more impact of the observed data and the observed data are biased for missing mechanism (3), this method results in only slight (negligible) improvement of the biases. All other imputation methods are helpless with the nonresponse biases because these methods do not use any auxiliary information from other variables.

We believe that one reason why the ratio imputation method performs so well is because we used the same variables to create and to impute the missing values for each target variable. The second reason is the high correlation coefficients (at least 0.6) between the target variables and the auxiliary variables used by the ratio imputation method. The ratio imputation method is more sensitive to the model specification because it directly uses the predicted values from the equations as imputation values. Actually, when we used ratio imputation model $E(y) = bx$ instead of $E(y-1) = bx$ in our first attempt, the results were worse than any other method. Later we subtracted 1 from y so that the means of y-1 and x were equal. But this is not a requirement of the ratio imputation method. It is more natural for many analysts to consider the model $E(y) = bx$ to impute y with auxiliary variable x rather than $E(y-1) = bx$. Therefore, we should be very cautious in the selection of ratio imputation models in real applications where the underlying missing mechanisms and the data distributions are generally unknown.

The fourth row of each block in tables 5.2.1.1–5.2.1.5 present the biases for the right skewed distribution, the mixer of 95 percent Normal and 5 percent Chi-square. These biases are not severe when the missing rates are low. As the missing rates increase, the biases become considerable. For the MCAR missing mechanism, all imputation methods are supposed to provide unbiased mean estimates. For missing mechanisms (2) and (3), since tail values are more likely missing and the right side has more tail values with the right skewed distributions, the mean estimates based on the incomplete data will underestimate the population mean. It is evident that the biases with the confounded mechanism (5) are much more serious than with the unconfounded mechanism (2). On the other hand, for missing mechanism (4), when center values are more likely missing, the estimates based on the incomplete data tend to overestimate the population mean. But the right skewness will not have as much effect with this missing mechanism as with missing mechanisms (2) and (5) since center values have much less effect on the mean estimates than tail values. That is why row 4 of block 4 in tables 5.2.1.2–5.2.1.4 does not show positive biases. However, the positive biases are substantial in row 4 of block 4 in table 5.2.1.5 when the missing rate increases to 40 percent.

We found earlier that ratio imputation with or without disturbance, Schafer's software, PROC IMPUTE, and hot deck are all very effective in improving the biases caused by missing mechanism (3). However, the improvement is much less impressive for the biases caused by the right skewness of the distributions, although these methods can still provide improvement in most cases when considerable biases exist with the incomplete data. Overall, they are still a little better than the other methods.

Table 5.2.1.1—Bias of population mean estimates (overall *)

Missing Mechanism	Distribution	Missing Rate	Mean Imp.	Ratio Imp.	Hot Deck	Random	Mean +e	Ratio +e	ABB	BB	Proc Impute	Schafer	Adj. DA
1. MCAR	Normal	24.7%	-0.005		0.012	-0.007	-0.009		-0.003	-0.008	-0.003	-0.006	-0.004
	Dexp	25.0%	-0.004		0.014	0.001	0.000		-0.009	-0.015	-0.003	-0.004	0.003
	MixNorm	24.9%	0.003		0.025	0.003	0.009		0.009	0.002	-0.004	-0.005	0.001
	MixNChi	24.9%	0.009		0.079	0.011	0.011		0.011	0.033	0.014	0.008	0.022
2. Unconfounded	Normal	17.7%	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	18.0%	-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	18.5%	0.003	-0.009	0.001	0.000	0.001	-0.003	0.004	0.006	0.002	-0.004	0.001
missing)	MixNChi	16.8%	-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	9.5%	-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	9.2%	-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	9.8%	-0.109	0.001	-0.024	-0.109	-0.110	0.004	-0.112	-0.104	0.011	0.001	-0.098
missing)	MixNChi	9.1%	-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	22.5%	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	19.6%	-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	21.0%	0.010	0.031	0.008	0.007	0.010	0.030	0.018	0.016	-0.004	-0.002	0.007
missing)	MixNChi	23.9%	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	25.1%	-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	26.9%	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	27.1%	-0.010	-0.005	-0.004	-0.009	-0.013	-0.007	-0.011	-0.006	-0.014	-0.019	-0.006
missing)	MixNChi	27.1%	-0.076	-0.022	-0.045	-0.071	-0.070	-0.015	-0.072	-0.078	-0.065	-0.032	-0.062

* "Overall" means that the four missing rate categories are combined. Biases in population means for each separate missing rate category are reported in tables 5.2.1.2 to 5.2.1.5.

Table 5.2.1.2—Bias of population mean estimates with about 10% missing values *

Missing Mechanism	Distribution	Mean Imp.	Ratio Imp.	Hot Deck	Random	Mean +e	Ratio +e	ABB	BB	Proc Impute	Schafer	Adj DA
1. MCAR	Normal	-0.019		-0.023	-0.021	-0.021		-0.024	-0.014	-0.019	-0.021	-0.020
	Dexp	-0.006		0.005	-0.006	-0.010		-0.001	-0.007	0.003	0.008	-0.003
	MixNorm	0.033		0.036	0.039	0.045		0.043	0.035	0.031	0.027	0.040
	MixNChi	-0.028		-0.031	-0.013	-0.024		-0.034	-0.017	-0.027	-0.012	0.001
2. Unconfounded	Normal	-0.001	-0.007	-0.005	-0.002	-0.002	-0.010	0.000	0.003	-0.004	-0.007	-0.003
(tail values	Dexp	0.005	0.000	0.016	-0.002	0.000	0.004	0.009	0.004	0.009	0.005	-0.001
more likely	MixNorm	0.016	0.023	0.022	0.022	0.017	0.021	0.021	0.018	0.026	0.026	0.027
missing)	MixNChi	-0.023	-0.046	0.007	-0.024	-0.013	-0.035	-0.024	-0.032	-0.030	-0.011	-0.015
3. Unconfounded	Normal	-0.042	0.014	0.002	-0.042	-0.044	0.016	-0.041	-0.044	0.007	0.008	-0.040
(large values	Dexp	-0.061	0.008	-0.011	-0.062	-0.059	0.009	-0.056	-0.062	-0.001	-0.002	-0.059
more likely	MixNorm	-0.035	0.028	0.011	-0.035	-0.039	0.031	-0.039	-0.036	0.020	0.020	-0.031
missing)	MixNChi	-0.091	0.015	-0.031	-0.084	-0.081	0.017	-0.091	-0.090	-0.022	0.004	-0.079
4. Unconfounded	Normal	-0.010	-0.006	-0.020	-0.003	-0.006	0.000	-0.009	-0.012	-0.011	-0.012	-0.003
(Center values	Dexp	0.014	0.020	0.009	0.006	0.007	0.027	0.015	0.006	0.012	0.013	0.005
more likely	MixNorm	0.024	0.027	0.022	0.027	0.023	0.027	0.032	0.034	0.020	0.018	0.028
missing)	MixNChi	-0.013	-0.005	-0.024	-0.020	-0.024	-0.018	-0.036	-0.010	-0.023	-0.030	-0.013
5. Confounded	Normal	0.005	0.003	0.011	0.006	0.007	0.010	0.007	0.004	0.006	0.000	0.006
(tail values	Dexp	0.026	0.024	0.034	0.033	0.029	0.028	0.025	0.026	0.025	0.017	0.030
more likely	MixNorm	0.004	0.010	0.009	0.004	0.001	0.012	0.010	0.011	0.006	0.010	0.006
missing)	MixNChi	-0.060	-0.008	-0.035	-0.050	-0.045	-0.006	-0.064	-0.050	-0.029	-0.010	-0.041

* There are about 5% missing values for missing mechanism 3.

Table 5.2.1.3—Bias of population mean estimates with about 20% missing values *

Missing Mechanism	Distribution	Mean Imp.	Ratio Imp.	Hot Deck	Random	Mean +e	Ratio +e	ABB	BB	Proc Impute	Schafer	Adj DA
1. MCAR	Normal	-0.003		0.003	0.001	-0.005		0.002	-0.010	-0.007	-0.011	0.000
	Dexp	-0.013		-0.014	0.006	-0.001		-0.023	-0.024	-0.024	-0.012	0.000
	MixNorm	-0.003		0.004	-0.003	0.006		-0.005	0.004	-0.012	-0.012	-0.006
	MixNChi	-0.012		0.004	-0.028	0.016		-0.007	-0.008	0.022	0.016	-0.016
2. Unconfounded	Normal	-0.001	-0.008	0.003	0.000	0.001	-0.017	0.005	-0.002	-0.004	-0.001	-0.002
(tail values	Dexp	-0.007	-0.003	0.019	-0.008	-0.003	0.009	-0.013	-0.006	-0.004	-0.006	-0.010
more likely	MixNorm	0.009	-0.014	0.005	0.003	0.016	-0.006	0.008	0.006	-0.004	-0.010	-0.003
missing)	MixNChi	0.024	-0.009	0.086	0.026	0.020	-0.009	0.026	0.039	-0.001	0.020	0.018
3. Unconfounded	Normal	-0.081	0.005	-0.019	-0.080	-0.083	-0.002	-0.080	-0.087	0.003	0.001	-0.075
(large values	Dexp	-0.092	0.019	-0.020	-0.090	-0.102	0.020	-0.094	-0.088	0.015	0.016	-0.082
more likely	MixNorm	-0.105	-0.003	-0.033	-0.112	-0.105	0.000	-0.103	-0.100	-0.009	-0.007	-0.106
missing)	MixNChi	-0.159	-0.004	-0.070	-0.167	-0.151	-0.003	-0.153	-0.145	-0.056	-0.014	-0.149
4. Unconfounded	Normal	0.001	0.016	0.019	-0.003	-0.002	0.013	0.014	0.006	0.000	0.005	-0.004
(Center values	Dexp	-0.004	0.011	-0.006	-0.010	-0.005	0.015	-0.009	-0.004	-0.011	-0.009	-0.011
more likely	MixNorm	0.002	0.020	0.001	0.003	0.002	0.022	0.003	-0.001	0.001	0.000	0.004
missing)	MixNChi	-0.004	0.026	-0.018	0.015	0.000	0.041	0.014	0.029	-0.015	-0.021	0.017
5. Confounded	Normal	0.000	0.004	0.000	0.006	-0.009	0.011	0.001	0.007	-0.009	0.002	0.007
(tail values	Dexp	0.003	0.009	0.005	0.007	0.006	0.005	0.012	-0.016	0.003	0.002	0.006
more likely	MixNorm	0.003	-0.002	0.001	0.002	0.001	-0.015	0.012	0.011	-0.002	-0.011	0.007
missing)	MixNChi	-0.086	-0.031	-0.032	-0.065	-0.078	-0.005	-0.076	-0.096	-0.065	-0.034	-0.062

* There are about 10% missing values for missing mechanism 3.

Table 5.2.1.4—Bias of population mean estimates with about 30% missing values *

Missing Mechanism	Distribution	Mean Imp.	Ratio Imp.	Hot Deck	Random	Mean +e	Ratio +e	ABB	BB	Proc Impute	Schafer	Adj DA
1. MCAR	Normal	0.017		0.044	0.013	0.019		0.010	0.005	0.023	0.014	0.014
	Dexp	0.035		0.052	0.035	0.038		0.034	0.025	0.031	0.030	0.043
	MixNorm	-0.034		-0.002	-0.027	-0.036		-0.020	-0.056	-0.042	-0.039	-0.031
	MixNChi	0.005		0.090	0.003	0.013		0.005	0.039	0.006	-0.019	0.017
2. Unconfounded	Normal	0.032	0.019	0.029	0.039	0.041	0.012	0.022	0.035	0.019	0.021	0.037
(tail values	Dexp	-0.007	-0.020	-0.022	-0.009	-0.002	-0.020	0.004	-0.014	0.005	-0.001	-0.002
more likely	MixNorm	-0.019	-0.046	-0.043	-0.026	-0.025	-0.036	-0.024	-0.011	-0.040	-0.046	-0.034
missing)	MixNChi	-0.038	-0.048	-0.038	-0.037	-0.030	-0.058	-0.046	-0.025	-0.058	-0.017	-0.026
3. Unconfounded	Normal	-0.093	0.018	-0.001	-0.097	-0.094	0.020	-0.092	-0.088	0.028	0.019	-0.084
(large values	Dexp	-0.139	0.002	-0.054	-0.141	-0.137	0.000	-0.136	-0.134	0.031	0.001	-0.125
more likely	MixNorm	-0.160	-0.035	-0.056	-0.160	-0.167	-0.032	-0.165	-0.158	-0.017	-0.029	-0.146
missing)	MixNChi	-0.203	-0.028	-0.057	-0.203	-0.211	-0.032	-0.187	-0.202	-0.081	-0.040	-0.182
4. Unconfounded	Normal	0.041	0.063	0.034	0.051	0.044	0.066	0.040	0.028	0.019	0.030	0.043
(Center values	Dexp	-0.022	0.010	-0.004	-0.021	0.002	0.008	-0.006	-0.020	-0.045	-0.019	-0.022
more likely	MixNorm	-0.012	0.019	-0.016	-0.009	-0.019	0.022	0.005	-0.001	-0.023	-0.028	-0.015
missing)	MixNChi	-0.027	0.025	-0.023	-0.044	-0.027	0.033	-0.045	-0.031	-0.050	-0.045	-0.043
5. Confounded	Normal	-0.001	0.013	0.002	0.000	0.004	0.018	-0.009	0.004	-0.001	-0.006	-0.007
(tail values	Dexp	-0.014	-0.005	0.000	-0.009	-0.023	-0.019	-0.003	0.000	-0.029	-0.007	-0.008
more likely	MixNorm	-0.036	-0.022	-0.031	-0.042	-0.045	-0.017	-0.043	-0.027	-0.041	-0.045	-0.042
missing)	MixNChi	-0.095	-0.047	-0.064	-0.105	-0.096	-0.034	-0.098	-0.109	-0.091	-0.063	-0.100

* There are about 15% missing values for missing mechanism 3.

Table 5.2.1.5—Bias of population mean estimates with about 40% missing values *

Missing Mechanism	Distribution	Mean Imp.	Ratio Imp.	Hot Deck	Random	Mean +e	Ratio +e	ABB	BB	Proc Impute	Schafer	Adj DA
1. MCAR	Normal	-0.018		0.023	-0.021	-0.027		-0.002	-0.015	-0.010	-0.006	-0.009
	Dexp	-0.032		0.012	-0.030	-0.026		-0.045	-0.053	-0.022	-0.040	-0.029
	MixNorm	0.018		0.062	0.001	0.021		0.016	0.025	0.009	0.004	-0.001
	MixNChi	0.072		0.253	0.083	0.039		0.078	0.118	0.054	0.048	0.084
2. Unconfounded	Normal	-0.010	-0.013	-0.028	-0.007	-0.018	-0.013	-0.005	-0.006	-0.018	-0.017	-0.006
(tail values	Dexp	-0.003	-0.022	-0.045	-0.009	-0.013	-0.028	0.014	-0.013	-0.005	-0.011	-0.016
more likely	MixNorm	0.004	0.003	0.020	0.002	-0.003	0.009	0.011	0.011	0.025	0.013	0.014
missing)	MixNChi	-0.018	-0.032	0.008	-0.008	-0.023	-0.029	0.001	-0.027	-0.001	0.008	-0.016
3. Unconfounded	Normal	-0.159	-0.027	-0.064	-0.160	-0.156	-0.020	-0.159	-0.158	0.000	-0.023	-0.141
(large values	Dexp	-0.178	-0.016	-0.052	-0.170	-0.179	-0.022	-0.192	-0.166	0.033	-0.005	-0.145
more likely	MixNorm	-0.134	0.013	-0.018	-0.131	-0.130	0.018	-0.139	-0.122	0.049	0.020	-0.109
missing)	MixNChi	-0.182	0.012	-0.086	-0.185	-0.183	0.015	-0.175	-0.178	-0.022	0.022	-0.163
4. Unconfounded	Normal	0.018	0.054	0.003	0.017	0.004	0.050	0.004	0.026	-0.017	-0.007	0.014
(Center values	Dexp	-0.011	0.046	-0.053	-0.003	-0.004	0.058	-0.002	-0.010	-0.018	-0.005	-0.011
more likely	MixNorm	0.024	0.059	0.025	0.008	0.033	0.052	0.031	0.033	-0.013	0.003	0.009
missing)	MixNChi	0.109	0.144	0.154	0.147	0.149	0.159	0.148	0.083	0.038	0.081	0.125
5. Confounded	Normal	-0.017	-0.011	-0.011	-0.014	-0.023	-0.022	-0.032	-0.021	-0.028	-0.020	-0.020
(tail values	Dexp	0.007	0.035	0.009	0.010	0.005	0.035	-0.008	0.004	0.024	0.013	-0.006
more likely	MixNorm	-0.012	-0.006	0.005	0.000	-0.008	-0.007	-0.022	-0.018	-0.019	-0.028	0.006
missing)	MixNChi	-0.062	-0.004	-0.046	-0.062	-0.063	-0.017	-0.051	-0.055	-0.076	-0.022	-0.045

* There are about 20% mis sing values for missing mechanism 3.

5.2.2 Bias of variance estimates with single imputation

Tables 5.2.2.1–5.2.2.5 report the relative biases of variance estimates based on the incomplete data and the data imputed by the 11 methods. The relative biases are defined as:

$$Relative \quad Bias = \frac{(Estimated \quad Var) \quad - \quad (True \quad Var)}{True \quad Var}. \qquad (5.1)$$

In this formula, we are discussing the variance among the data $Var(y_i)$, not the variance of the mean estimates $Var(\bar{y})$, although the relative biases of the two variance estimates are equal for all the imputation methods. We will use the statement "the variance is 20 percent overestimated" if the relative bias is 0.20, and say "the variance is 20 percent underestimated" if the relative bias is -0.20.

For the MCAR missing mechanism, the variance estimates based on the incomplete data are supposed to be unbiased, which was confirmed by the simulation. It is to be expected that the mean imputation method seriously underestimates the variances since the data were centralized by using the mean as the imputed values for all missing cases. One way to correct this underestimation is to multiply the variance estimates by the factor $(n-1)/(r-1)$, where n is the sample size and r is the number of observed values. The other way is to add random variation to the mean as imputation values as done by the mean with disturbance imputation method. Actually, the variance estimates based on the incomplete data and those based on the mean with disturbance imputation method are always approximately equal across all missing mechanisms and all distributions.

For MCAR, all other methods seem fine except the sequential hot deck method which provides a few very large variance estimates for the mixed distribution of 95 percent normal and 5 percent Chi-square. For example, the sequential hot deck overestimated the variance by 70 percent and 24 percent respectively when there are 40 percent and 30 percent missing values. This is probably because some extremely large values were imputed too many times by the hot deck sequential imputation scheme. Therefore, the sequential hot deck imputation method is dangerous even for MCAR missing mechanism if extreme values or outliers exist in the observed data. For other distributions, the hot deck method works well.

For unconfounded missing mechanism (2) where tail values are more likely missing, the incomplete data shrink to the center and, therefore, the variance estimates based on the incomplete data are too small. This underestimation is much less serious than for the confounded missing mechanism (5) where tail values are also more likely missing but the missing probabilities depend on the target variable itself. For mechanism (2), Schafer's software performs better than the ratio imputation, which is better than PROC IMPUTE, which is better than the hot deck method. However, all four methods dramatically improved the negative biases of the variance estimates. The ratio imputation with disturbance method tends to overestimate the variances.

Slight improvement has been found with the adjusted data augmentation method. It is evident and expectable that the BB, ABB, random, and the mean with disturbance imputation methods all have almost the same variance estimates as the incomplete data, while the mean imputation method worsens the variance estimates.

For unconfounded missing mechanism (3) where large values are more likely missing, the incomplete data have shorter range than the complete data; therefore, the incomplete data will underestimate the true variance. Since the missing rates are always less than 20 percent, the underestimation of the variances is not severe. Except for one case, all negative biases are smaller than 11 percent of the true variances. In this cases all imputation methods except the mean imputation provide fine variance estimates. However, Schafer's software, ratio imputation, PROC IMPUTE, and the hot deck method still shows some advantage over the other methods.

For unconfounded missing mechanism (4) where center values are more likely missing, the incomplete data overestimate the variances and so do the random, mean imputation with disturbance, ratio imputation with disturbance, ABB, and BB methods, while the mean imputation still underestimates the variances. These methods cannot improve the positive biases at all. Overall, Schafer's software has the best performance, followed by the hot deck method, which is followed by PROC IMPUTE, which is followed by the ratio imputation. All four methods substantially improved the positive biases of variance estimates. The hot deck method has one bad case in which it overestimates the variance by 23 percent for the mixer of normal and Chi-square when the missing rate is 40 percent, but it is still a significant improvement over the incomplete data which overestimate the variance by 37 percent. Again, the adjusted data augmentation method can improve the biases slightly.

For confounded missing mechanism (5) where tail values are more likely missing and the missing probabilities depend on the target variable itself, the incomplete data underestimate the variances much more seriously than for unconfounded missing mechanism (2). Again, the random, mean imputation with disturbance, ratio imputation with disturbance, ABB, and BB methods do not help at all with the biases. Schafer's software, adjusted data augmentation and the hot deck method only slightly improve them. PROC IMPUTE only have improvement with the mixed distribution of normal and Chi-square which has much more serious underestimated variances than the other distributions. For this distribution, PROC IMPUTE is better than Schafer's software, adjusted data augmentation, and the hot deck method. For this confounded missing mechanism, the only methods which can substantially improve the biases in variance estimates are ratio imputation with or without disturbance. These two methods are the only ones in this study that directly use auxiliary variables to predict missing values. This probably implies that we may have to use some directly predictive approach such as regression imputation or ratio imputation to impute missing values if the missing mechanism is confounded; that is, if the missing probabilities depend on the target variable itself.

In summary, for the MCAR missing mechanism, all imputation methods can provide acceptable variance estimates except the mean imputation method, which needs to be adjusted with a factor of $(n-1)/(r-1)$. For unconfounded missing mechanisms, Schafer's software performs best, and ratio imputation, PROC IMPUTE, and the hot deck method can all improve the biases of variance estimates dramatically, but the ratio imputation with disturbance method tends to overestimate the variance. For the confounded missing mechanism, only the ratio imputation method with or without disturbance substantially improves the biases. The random, ABB, BB, and mean imputation with disturbance methods are almost equivalent to the incomplete data for all missing mechanisms, while the adjusted data augmentation method always helps a little, but never much.

Table 5.2.2.1—Relative bias of variance estimates with single imputation (overall *)

Missing Mechanism	Distribution	Incomp	Mean Imp.	Ratio Imp.	Hot Deck	Random	Mean +e	Ratio +e	ABB	BB	Proc Impute	Schafer	Adj. DA
1. MCAR	Normal	-0.002	-0.250		-0.039	-0.019	-0.010		-0.008	-0.009	-0.027	0.012	-0.010
	Dexp	0.029	-0.234		-0.020	0.019	0.024		0.006	0.010	0.001	0.014	0.024
	MixNorm	0.003	-0.247		-0.039	-0.004	-0.004		-0.006	-0.028	-0.027	0.004	0.006
	MixNChi	0.016	-0.242		0.195	-0.011	0.007		-0.008	0.064	-0.044	0.026	0.018
2. Unconfounded	Normal	-0.125	-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	-0.236	-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	-0.191	-0.341	0.064	-0.025	-0.205	-0.193	0.206	-0.205	-0.196	-0.006	-0.002	-0.162
missing)	MixNChi	-0.424	-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	-0.047	-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	-0.042	-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	-0.045	-0.138	-0.020	-0.024	-0.051	-0.051	0.068	-0.049	-0.041	0.041	0.004	-0.044
missing)	MixNChi	-0.107	-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	0.126	-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	0.110	-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	0.121	-0.123	-0.083	-0.002	0.121	0.115	0.162	0.122	0.123	-0.036	-0.002	0.095
missing)	MixNChi	0.146	-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	-0.272	-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	-0.350	-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	-0.352	-0.514	-0.178	-0.330	-0.353	-0.351	0.054	-0.361	-0.353	-0.375	-0.323	-0.338
missing)	MixNChi	-0.674	-0.750	-0.228	-0.629	-0.678	-0.676	-0.075	-0.676	-0.680	-0.488	-0.550	-0.644

* "Overall" means that the four missing rate categories are combined. Relative biases of variance estimates for each separate missing rate category are reported in tables 5.2.2.2 to 5.2.2.5.

Table 5.2.2.2—Relative bias of variance estimates with single imputation with about 10% missing values*

Missing Mechanism	Distribution	Incomp	Mean Imp.	Ratio Imp.	Hot Deck	Random	Mean +e	Ratio +e	ABB	BB	Proc Impute	Schafer	Adj. DA
1. MCAR	Normal	0.000	-0.091		0.001	-0.009	-0.008		0.003	-0.006	-0.003	0.012	-0.005
	Dexp	-0.012	-0.112		-0.025	-0.016	-0.016		-0.021	-0.018	-0.016	-0.009	-0.011
	MixNorm	0.011	-0.092		-0.008	-0.001	0.008		-0.001	0.012	-0.007	0.016	0.003
	MixNChi	0.015	-0.076		-0.016	0.010	0.018		0.009	0.116	-0.029	0.033	0.028
2. Unconfounded	Normal	-0.052	-0.140	0.030	0.016	-0.044	-0.051	0.120	-0.057	-0.053	0.017	0.026	-0.031
(tail values	Dexp	-0.157	-0.236	0.013	-0.032	-0.149	-0.157	0.071	-0.161	-0.152	-0.038	-0.021	-0.130
more likely	MixNorm	-0.135	-0.218	0.036	-0.007	-0.145	-0.135	0.134	-0.142	-0.141	-0.019	0.004	-0.129
missing)	MixNChi	-0.383	-0.439	0.049	-0.238	-0.393	-0.388	0.112	-0.379	-0.398	-0.074	0.046	-0.358
3. Unconfounded	Normal	-0.014	-0.061	0.019	0.010	-0.019	-0.016	0.075	-0.011	-0.018	0.014	0.016	-0.017
(large values	Dexp	-0.038	-0.080	-0.005	-0.021	-0.048	-0.039	0.028	-0.037	-0.036	-0.012	-0.010	-0.045
more likely	MixNorm	-0.037	-0.084	-0.006	-0.015	-0.034	-0.048	0.041	-0.039	-0.038	-0.007	-0.010	-0.031
missing)	MixNChi	-0.047	-0.087	0.073	0.041	-0.048	-0.047	0.122	-0.056	-0.053	-0.014	0.065	-0.048
4. Unconfounded	Normal	0.028	-0.068	-0.064	-0.019	0.023	0.026	0.021	0.023	0.020	-0.035	-0.023	0.021
(Center values	Dexp	0.053	-0.030	-0.028	0.016	0.049	0.058	0.055	0.050	0.064	-0.009	0.014	0.046
more likely	MixNorm	0.016	-0.072	-0.069	-0.038	0.017	0.008	0.014	0.022	0.012	-0.044	-0.028	0.015
missing)	MixNChi	0.008	-0.109	-0.107	-0.071	0.064	0.004	0.008	-0.037	-0.021	-0.085	-0.062	0.035
5. Confounded	Normal	-0.139	-0.223	-0.088	-0.132	-0.137	-0.145	0.001	-0.133	-0.141	-0.154	-0.122	-0.132
(tail values	Dexp	-0.180	-0.272	-0.096	-0.156	-0.178	-0.192	-0.009	-0.177	-0.186	-0.191	-0.152	-0.177
more likely	MixNorm	-0.182	-0.275	-0.129	-0.167	-0.193	-0.184	-0.033	-0.197	-0.177	-0.196	-0.163	-0.190
missing)	MixNChi	-0.441	-0.505	-0.147	-0.403	-0.449	-0.441	-0.070	-0.441	-0.458	-0.306	-0.335	-0.430

* There are about 5% missing values for missing mechanism 3.

Table 5.2.2.3—Relative bias of variance estimates with single imputation with about 20% missing values *

Missing Mechanism	Distribution	Incomp	Mean Imp.	Ratio Imp.	Hot Deck	Random	Mean +e	Ratio +e	ABB	BB	Proc Impute	Schafer	Adj. DA
1. MCAR	Normal	0.015	-0.198		-0.013	0.008	0.010		0.015	0.006	-0.015	0.014	0.015
	Dexp	0.034	-0.178		0.008	0.032	0.029		0.004	0.027	-0.001	0.039	0.038
	MixNorm	-0.004	-0.203		-0.034	-0.021	-0.028		-0.002	-0.022	-0.023	0.001	-0.006
	MixNChi	-0.099	-0.281		-0.145	-0.101	-0.095		-0.133	-0.153	-0.031	-0.043	-0.077
2. Unconfounded	Normal	-0.141	-0.278	0.020	-0.010	-0.118	-0.155	0.149	-0.148	-0.136	0.013	-0.019	-0.089
(tail values	Dexp	-0.209	-0.343	0.075	-0.030	-0.220	-0.215	0.228	-0.225	-0.232	-0.001	0.020	-0.187
more likely	MixNorm	-0.162	-0.307	0.082	0.000	-0.183	-0.158	0.211	-0.184	-0.165	-0.012	-0.002	-0.146
missing)	MixNChi	-0.395	-0.508	0.117	0.012	-0.418	-0.394	0.223	-0.380	-0.375	-0.124	0.078	-0.336
3. Unconfounded	Normal	-0.061	-0.136	-0.022	-0.046	-0.068	-0.067	0.041	-0.069	-0.060	-0.005	-0.017	-0.060
(large values	Dexp	-0.022	-0.097	0.015	-0.005	-0.011	-0.019	0.093	-0.021	-0.027	0.027	0.030	-0.007
more likely	MixNorm	-0.060	-0.139	-0.014	-0.030	-0.063	-0.061	0.056	-0.069	-0.054	0.007	-0.007	-0.056
missing)	MixNChi	-0.212	-0.279	-0.091	-0.159	-0.218	-0.210	-0.046	-0.226	-0.190	-0.165	-0.084	-0.199
4. Unconfounded	Normal	0.117	-0.097	-0.077	0.014	0.104	0.103	0.146	0.126	0.109	-0.019	0.006	0.096
(Center values	Dexp	0.100	-0.063	-0.053	0.037	0.093	0.096	0.123	0.099	0.114	-0.018	0.015	0.080
more likely	MixNorm	0.097	-0.089	-0.072	0.003	0.086	0.091	0.117	0.087	0.092	-0.033	-0.004	0.075
missing)	MixNChi	0.095	-0.090	-0.084	-0.008	0.094	0.107	0.074	0.094	0.163	-0.064	-0.006	0.077
5. Confounded	Normal	-0.224	-0.379	-0.118	-0.201	-0.224	-0.234	0.099	-0.229	-0.230	-0.252	-0.199	-0.211
(tail values	Dexp	-0.267	-0.419	-0.135	-0.249	-0.270	-0.273	0.068	-0.281	-0.263	-0.280	-0.230	-0.261
more likely	MixNorm	-0.286	-0.439	-0.142	-0.266	-0.274	-0.288	0.057	-0.297	-0.289	-0.303	-0.252	-0.252
missing)	MixNChi	-0.709	-0.789	-0.220	-0.650	-0.712	-0.710	-0.103	-0.708	-0.714	-0.513	-0.565	-0.679

* There are about 10% missing values for missing mechanism 3.

Table 5.2.2.4—Relative bias of variance estimates with single imputation with about 30% missing values *

Missing Mechanism	Distribution	Incomp	Mean Imp.	Ratio Imp.	Hot Deck	Random	Mean +e	Ratio +e	ABB	BB	Proc Impute	Schafer	Adj. DA
1. MCAR	Normal	-0.022	-0.312		-0.058	-0.053	-0.022		-0.034	-0.024	-0.041	0.010	-0.043
	Dexp	0.037	-0.275		-0.025	0.001	0.039		0.035	-0.011	0.009	0.008	0.010
	MixNorm	-0.002	-0.303		-0.051	-0.010	0.005		-0.004	-0.055	-0.024	-0.005	0.000
	MixNChi	0.077	-0.255		0.238	0.072	0.077		0.089	0.090	0.022	0.067	0.050
2. Unconfounded	Normal	-0.155	-0.335	0.047	-0.001	-0.162	-0.167	0.200	-0.152	-0.162	0.104	0.013	-0.125
(tail values	Dexp	-0.262	-0.427	0.083	-0.063	-0.287	-0.260	0.207	-0.273	-0.267	-0.002	0.021	-0.224
more likely	MixNorm	-0.239	-0.403	0.107	-0.066	-0.250	-0.233	0.254	-0.260	-0.241	-0.026	0.010	-0.195
missing)	MixNChi	-0.459	-0.557	0.000	-0.240	-0.477	-0.461	0.105	-0.448	-0.470	-0.040	-0.039	-0.390
3. Unconfounded	Normal	-0.040	-0.152	-0.018	-0.018	-0.037	-0.034	0.099	-0.036	-0.033	0.031	0.019	-0.027
(large values	Dexp	-0.044	-0.149	-0.024	-0.028	-0.039	-0.048	0.070	-0.044	-0.047	0.084	0.009	-0.028
more likely	MixNorm	-0.011	-0.130	-0.004	0.013	-0.031	-0.016	0.111	-0.007	-0.003	0.090	0.043	-0.024
missing)	MixNChi	-0.100	-0.197	-0.039	0.021	-0.104	-0.091	0.068	-0.069	-0.073	-0.069	-0.025	-0.101
4. Unconfounded	Normal	0.163	-0.147	-0.090	0.021	0.148	0.149	0.194	0.164	0.157	-0.041	0.024	0.120
(Center values	Dexp	0.142	-0.134	-0.103	0.015	0.153	0.163	0.153	0.135	0.138	-0.046	-0.012	0.132
more likely	MixNorm	0.192	-0.116	-0.065	0.053	0.198	0.188	0.247	0.197	0.198	-0.009	0.036	0.164
missing)	MixNChi	0.113	-0.217	-0.196	-0.102	0.101	0.108	0.117	0.151	0.102	-0.168	-0.087	0.038
5. Confounded	Normal	-0.314	-0.518	-0.174	-0.303	-0.332	-0.321	0.134	-0.318	-0.316	-0.363	-0.276	-0.313
(tail values	Dexp	-0.408	-0.604	-0.183	-0.372	-0.428	-0.425	0.086	-0.404	-0.410	-0.444	-0.375	-0.405
more likely	MixNorm	-0.430	-0.621	-0.206	-0.406	-0.427	-0.426	0.064	-0.449	-0.434	-0.472	-0.395	-0.416
missing)	MixNChi	-0.734	-0.817	-0.290	-0.691	-0.733	-0.738	-0.101	-0.742	-0.736	-0.514	-0.611	-0.694

* There are about 15% missing values for missing mechanism 3.

Table 5.2.2.5—Relative bias of variance estimates with single imputation with about 40% missing values *

Missing Mechanism	Distribution	Incomp	Mean Imp.	Ratio Imp.	Hot Deck	Random	Mean +e	Ratio +e	ABB	BB	Proc Impute	Schafer	Adj. DA
1. MCAR	Normal	-0.001	-0.398		-0.085	-0.023	-0.019		-0.018	-0.013	-0.049	0.014	-0.009
	Dexp	0.056	-0.372		-0.036	0.059	0.045		0.005	0.043	0.013	0.018	0.058
	MixNorm	0.006	-0.389		-0.064	0.015	-0.001		-0.015	-0.046	-0.053	0.005	0.025
	MixNChi	0.071	-0.356		0.702	-0.027	0.027		0.003	0.205	-0.137	0.047	0.071
2. Unconfounded	Normal	-0.151	-0.363	0.034	-0.009	-0.169	-0.154	0.220	-0.162	-0.133	0.188	-0.002	-0.142
(tail values	Dexp	-0.317	-0.482	0.055	-0.135	-0.322	-0.318	0.190	-0.318	-0.310	-0.005	-0.057	-0.254
more likely	MixNorm	-0.230	-0.435	0.032	-0.027	-0.241	-0.247	0.227	-0.235	-0.239	0.035	-0.019	-0.180
missing)	MixNChi	-0.457	-0.573	-0.135	-0.349	-0.396	-0.472	-0.052	-0.454	-0.461	-0.200	-0.104	-0.345
3. Unconfounded	Normal	-0.071	-0.200	-0.049	-0.062	-0.076	-0.073	0.105	-0.068	-0.072	0.075	0.000	-0.062
(large values	Dexp	-0.063	-0.196	-0.074	-0.042	-0.060	-0.056	0.039	-0.062	-0.071	0.069	-0.019	-0.047
more likely	MixNorm	-0.070	-0.200	-0.055	-0.064	-0.075	-0.078	0.065	-0.081	-0.068	0.073	-0.011	-0.063
missing)	MixNChi	-0.069	-0.196	-0.035	-0.111	-0.100	-0.079	0.083	-0.080	-0.075	-0.040	0.006	-0.082
4. Unconfounded	Normal	0.197	-0.233	-0.099	0.039	0.183	0.195	0.325	0.164	0.192	-0.049	0.009	0.131
(Center values	Dexp	0.143	-0.226	-0.154	0.002	0.140	0.122	0.202	0.155	0.129	-0.091	-0.043	0.094
more likely	MixNorm	0.181	-0.215	-0.125	-0.026	0.182	0.173	0.270	0.184	0.188	-0.058	-0.014	0.126
missing)	MixNChi	0.369	-0.160	-0.116	0.226	0.401	0.330	0.392	0.538	0.248	-0.079	0.069	0.317
5. Confounded	Normal	-0.412	-0.657	-0.205	-0.386	-0.436	-0.412	0.190	-0.395	-0.426	-0.466	-0.390	-0.414
(tail values	Dexp	-0.546	-0.745	-0.235	-0.506	-0.556	-0.549	0.077	-0.553	-0.555	-0.577	-0.513	-0.535
more likely	MixNorm	-0.509	-0.721	-0.235	-0.480	-0.517	-0.505	0.127	-0.503	-0.512	-0.529	-0.482	-0.492
missing)	MixNChi	-0.812	-0.888	-0.254	-0.774	-0.818	-0.816	-0.026	-0.812	-0.812	-0.621	-0.689	-0.771

* There are about 20% missing values for missing mechanism 3.

5.2.3 Bias of variance estimates of population mean with five sets of imputations

Five sets of imputations were created for the eight random imputation methods under study. Variance estimates based on the five sets of multiple imputations are obtained through Rubin's multiple imputation theory:

$$\hat{V} = \frac{1}{m}\sum_{i=1}^{m}\hat{V}_i + \left(\frac{m+1}{m}\right)\frac{1}{m-1}\sum_{i=1}^{m}\left(\hat{q}_i - \bar{q}\right), \tag{5.2}$$

where $\hat{q}_i$ and $\hat{V}_i$ are the parameter estimate and variance estimate, respectively, based on $i\text{-}th$ (i=1, …, m) set of imputations. The first term in (5.2) is called the within-imputation variability, and the second term is referred as the between-imputation variability.

Tables 5.2.3.1–5.2.3.5 present the relative biases of variance estimates of population mean estimates. The relative biases are defined as in (5.1). Multiple imputation variance estimates are generally larger than single imputation variance estimates since multiple imputation adds the between-imputation variation.

If the data are missing completely at random, all methods except PROC IMPUTE and Schafer's software substantially overestimate the variances. For the combined data with about 25 percent missing values, the random, mean with disturbance, ratio with disturbance, and adjusted data augmentation methods all overestimate the variance by 25 percent to 35 percent, while ABB and BB methods overestimate the variances by 35 percent to 55 percent. Even with a 10 percent missing rate, these methods overestimate the variances by more than 10 percent in most cases. It seems that the second term in (2.2) is too much to add to the variance estimates. The ABB and BB methods, which introduce more variation than the random method and are considered "proper" by Rubin (1987), seem to overestimate the variances most seriously. PROC IMPUTE provides the best variance estimates with this ideal missing mechanism although it is not "proper" according to Rubin's definition. Its multiple imputation variance estimates can be considered unbiased. Schafer's software is the second best and it slightly overestimates the variances.

For unconfounded missing mechanisms (2) and (3) where the incomplete data underestimate the variances, the multiple imputation variance estimates corrected more negative biases than the single imputation variance estimates, as expected. PROC IMPUTE and Schafer's software again have the best overall performance. All other methods except the ratio with disturbance method produce fine variance estimates. The ratio with disturbance method significantly overestimate the variances even for these two missing mechanisms when the incomplete data are more concentrated around the center than the population distribution.

For the unconfounded missing mechanism (4) when center values are more likely missing and the incomplete data are more diversified than the population distribution, the relative performances across the different imputation methods are similar to those for the ideal missing mechanism (1). PROC IMPUTE works best and provides approximately unbiased variance estimates, Schafer's software is the second best and slightly overestimates the variances. Other methods all overestimate the variances; the ABB and BB methods are the worst in terms of bias of variance estimates.

For confounded missing mechanism (5) when the incomplete data seriously underestimate the variance, the extra variation introduced by multiple imputation helps reduce the negative biases of single imputation variance estimates for all methods except the ratio with disturbance imputation method. The ratio with disturbance imputation method again overestimate the variances. Except for the mixed distribution of normal and Chi-square, PROC IMPUTE has the largest negative biases and the ABB and BB methods have the smallest biases, while all the other methods are close to the ABB and BB methods. For the mixed right-skewed distribution of normal and Chi-square, PROC IMPUTE has the smallest negative biases; however, all methods except the ratio with disturbance method still substantially underestimate the variances.

In summary, the ratio with disturbance imputation method always overestimates the variances for all types of missing mechanisms when between-imputation variation is introduced via multiple imputations. For this method, the idea of multiple imputation is obviously inappropriate. PROC IMPUTE seems to have the least between-imputation variation and it provides approximately unbiased variance estimates for the MCAR and all unconfounded missing mechanisms. The ABB and BB methods introduce the most between-imputation variation and most seriously overestimate the variances for the MCAR and missing mechanism (4) when the incomplete data are more diversified than the true distribution. For these two types of missing mechanisms, multiple imputation variance estimates of all methods except PROC IMPUTE tend to overestimate the true variances. For the other missing mechanisms when the incomplete data are less diversified than the true distribution, introducing between-imputation variation can help reduce the negative biases of variance estimates except for the ratio with disturbance method.

Table 5.2.3.1—Relative bias of variance estimates with five sets of imputations (overall *)

Missing Mechanism	Distribution	Random	Mean +e	Ratio +e	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

* “Overall” means that the four missing rate categories are combined. Relative biases of variance estimates with five sets of imputations for each separate missing rate category are reported in tables 5.2.3.2 to 5.2.3.5.

Table 5.2.3.2—Relative bias of variance estimates with five sets of imputations with about 10% missing values *

Missing Mechanism	Distribution	Random	Mean +e	Ratio +e	ABB	BB	Proc Impute	Schafer	Adj DA
1. MCAR	Normal	0.098	0.129		0.137	0.074	0.010	0.031	0.104
	Dexp	0.104	0.127		0.133	0.083	-0.009	0.014	0.109
	MixNorm	0.147	0.110		0.139	0.125	0.004	0.035	0.147
	MixNChi	0.097	0.113		0.116	0.190	-0.009	0.051	0.134
2. Unconfounded	Normal	0.063	0.035	0.227	0.053	0.026	0.026	0.033	0.085
(tail values	Dexp	-0.074	-0.069	0.163	-0.049	-0.084	-0.032	0.001	-0.045
more likely	MixNorm	-0.028	-0.040	0.209	-0.030	-0.054	0.001	0.025	0.008
missing)	MixNChi	-0.290	-0.328	0.172	-0.326	-0.324	-0.042	0.059	-0.207
3. Unconfounded	Normal	0.040	0.036	0.102	0.051	0.027	0.021	0.034	0.043
(large values	Dexp	-0.002	0.012	0.086	0.002	0.009	-0.010	0.007	0.002
more likely	MixNorm	0.013	0.006	0.095	0.014	0.001	0.002	0.008	0.019
missing)	MixNChi	0.043	0.003	0.171	-0.023	0.026	0.000	0.070	0.028
4. Unconfounded	Normal	0.141	0.137	0.158	0.156	0.118	-0.017	0.025	0.141
(Center values	Dexp	0.157	0.149	0.149	0.163	0.145	0.009	0.051	0.146
more likely	MixNorm	0.113	0.117	0.133	0.139	0.095	-0.030	0.025	0.112
missing)	MixNChi	0.233	0.163	0.144	0.175	0.111	-0.076	-0.017	0.190
5. Confounded	Normal	-0.033	-0.054	0.115	-0.015	-0.050	-0.121	-0.047	-0.018
(tail values	Dexp	-0.081	-0.073	0.124	-0.073	-0.093	-0.163	-0.077	-0.072
more likely	MixNorm	-0.075	-0.050	0.075	-0.038	-0.071	-0.171	-0.100	-0.069
missing)	MixNChi	-0.397	-0.360	0.016	-0.368	-0.396	-0.257	-0.277	-0.366

* There are about 5% missing values for missing mechanism 3.

Table 5.2.3.3—Relative bias of variance estimates with five sets of imputations with about 20% missing values *

Missing Mechanism	Distribution	Random	Mean +e	Ratio +e	ABB	BB	Proc Impute	Schafer	Adj DA
1. MCAR	Normal	0.218	0.266		0.340	0.278	0.033	0.056	0.234
	Dexp	0.316	0.285		0.317	0.265	0.016	0.093	0.320
	MixNorm	0.204	0.234		0.221	0.207	-0.005	0.060	0.201
	MixNChi	0.128	0.084		0.053	0.087	-0.002	-0.008	0.166
2. Unconfounded	Normal	0.008	0.026	0.318	0.023	0.023	0.019	0.014	0.066
(tail values	Dexp	-0.044	-0.033	0.367	-0.005	-0.059	0.015	0.045	0.046
more likely	MixNorm	-0.026	0.051	0.368	0.063	-0.015	0.013	0.024	0.038
missing)	MixNChi	-0.233	-0.278	0.345	-0.227	-0.214	-0.089	0.094	-0.068
3. Unconfounded	Normal	0.026	0.024	0.134	0.024	0.011	0.002	0.007	0.043
(large values	Dexp	0.083	0.062	0.179	0.095	0.053	0.043	0.054	0.094
more likely	MixNorm	0.018	0.031	0.164	0.013	0.009	0.023	0.028	0.035
missing)	MixNChi	-0.112	-0.143	0.054	-0.112	-0.114	-0.148	-0.057	-0.086
4. Unconfounded	Normal	0.374	0.323	0.399	0.431	0.343	0.013	0.114	0.339
(Center values	Dexp	0.283	0.300	0.326	0.346	0.264	0.012	0.112	0.260
more likely	MixNorm	0.297	0.328	0.333	0.360	0.336	0.015	0.090	0.278
missing)	MixNChi	0.290	0.311	0.281	0.422	0.352	-0.045	0.072	0.240
5. Confounded	Normal	-0.009	-0.025	0.284	0.026	-0.046	-0.202	-0.072	0.014
(tail values	Dexp	-0.083	-0.119	0.243	-0.036	-0.068	-0.230	-0.115	-0.068
more likely	MixNorm	-0.115	-0.087	0.285	-0.045	-0.144	-0.260	-0.104	-0.092
missing)	MixNChi	-0.602	-0.605	0.050	-0.569	-0.591	-0.492	-0.520	-0.517

* There are about 10% missing values for missing mechanism 3.

Table 5.2.3.4—Relative bias of variance estimates with five sets of imputations with about 30% missing values *

Missing Mechanism	Distribution	Random	Mean +e	Ratio +e	ABB	BB	Proc Impute	Schafer	Adj DA
1. MCAR	Normal	0.281	0.340		0.516	0.423	0.001	0.071	0.302
	Dexp	0.367	0.361		0.642	0.512	0.039	0.101	0.363
	MixNorm	0.358	0.394		0.537	0.329	-0.007	0.067	0.384
	MixNChi	0.298	0.569		0.482	0.785	0.065	0.131	0.402
2. Unconfounded	Normal	0.076	0.036	0.408	0.158	0.054	0.121	0.037	0.167
(tail values	Dexp	-0.094	-0.096	0.443	-0.010	-0.041	0.014	0.049	0.051
more likely	MixNorm	-0.026	-0.085	0.439	-0.009	-0.054	-0.012	0.037	0.108
missing)	MixNChi	-0.318	-0.326	0.216	-0.271	-0.340	-0.012	0.006	-0.131
3. Unconfounded	Normal	0.089	0.094	0.204	0.123	0.088	0.044	0.061	0.108
(large values	Dexp	0.067	0.070	0.189	0.092	0.053	0.114	0.048	0.082
more likely	MixNorm	0.138	0.135	0.245	0.150	0.141	0.106	0.091	0.154
missing)	MixNChi	0.005	-0.001	0.198	0.035	0.022	-0.032	0.023	0.043
4. Unconfounded	Normal	0.481	0.509	0.570	0.772	0.683	0.019	0.163	0.414
(Center values	Dexp	0.506	0.430	0.475	0.502	0.500	0.001	0.142	0.448
more likely	MixNorm	0.648	0.491	0.651	0.589	0.515	0.057	0.208	0.553
missing)	MixNChi	0.597	0.526	0.614	0.897	0.575	-0.143	0.033	0.430
5. Confounded	Normal	-0.089	-0.082	0.375	0.115	-0.016	-0.287	-0.119	-0.048
(tail values	Dexp	-0.172	-0.205	0.402	-0.113	-0.093	-0.374	-0.211	-0.146
more likely	MixNorm	-0.237	-0.255	0.419	-0.038	-0.139	-0.405	-0.222	-0.216
missing)	MixNChi	-0.627	-0.643	0.111	-0.587	-0.598	-0.459	-0.549	-0.541

* There are about 15% missing values for missing mechanism 3.

Table 5.2.3.5—Relative bias of variance estimates with five sets of imputations with about 40% missing values *

Missing Mechanism	Distribution	Random	Mean +e	Ratio +e	ABB	BB	Proc Impute	Schafer	Adj DA
1. MCAR	Normal	0.418	0.353		0.844	0.683	0.026	0.102	0.481
	Dexp	0.520	0.519		0.741	0.938	0.037	0.141	0.515
	MixNorm	0.421	0.475		0.703	0.731	-0.006	0.073	0.426
	MixNChi	0.691	0.515		0.920	1.167	-0.094	0.103	0.592
2. Unconfounded	Normal	0.094	0.088	0.503	0.171	0.155	0.208	0.034	0.170
(tail values	Dexp	-0.142	-0.146	0.401	0.008	-0.053	0.005	-0.032	-0.015
more likely	MixNorm	-0.023	0.009	0.418	0.074	0.040	0.039	0.048	0.093
missing)	MixNChi	-0.323	-0.297	0.088	-0.336	-0.308	-0.187	-0.071	-0.182
3. Unconfounded	Normal	0.119	0.104	0.215	0.133	0.063	0.086	0.037	0.148
(large values	Dexp	0.112	0.090	0.185	0.147	0.081	0.091	0.036	0.138
more likely	MixNorm	0.081	0.062	0.203	0.092	0.077	0.081	0.032	0.110
missing)	MixNChi	0.063	0.077	0.268	0.116	0.053	-0.018	0.053	0.087
4. Unconfounded	Normal	0.641	0.692	0.810	0.871	0.833	0.029	0.218	0.538
(Center values	Dexp	0.451	0.537	0.567	0.798	0.732	-0.043	0.148	0.370
more likely	MixNorm	0.675	0.650	0.633	0.811	0.907	0.008	0.156	0.548
missing)	MixNChi	1.155	0.906	0.889	1.515	1.247	-0.052	0.294	0.926
5. Confounded	Normal	-0.089	-0.093	0.596	0.060	0.076	-0.383	-0.135	-0.063
(tail values	Dexp	-0.346	-0.291	0.533	-0.187	-0.119	-0.518	-0.347	-0.306
more likely	MixNorm	-0.197	-0.292	0.477	0.036	-0.157	-0.475	-0.299	-0.126
missing)	MixNChi	-0.718	-0.728	0.241	-0.668	-0.657	-0.592	-0.618	-0.592

* There are about 20% missing values for missing mechanism 3.

5.2.4 Coverage rates

The coverage rate is defined as the ratio of the number of simulation replications in which the confidence interval estimates cover the true value to the total number of simulation replications. Tables 5.2.4.1–5.2.4.5 report the coverage rates of the 95 percent confidence interval estimates covering the true means for the combined missing category and separate missing categories, respectively.

Schafer's software obviously has the best coverage rates. It has almost perfect rates across the five missing mechanisms for all missing rate categories. The adjusted data augmentation method also has almost perfect coverage rates for all missing rate categories and all missing mechanisms except mechanism (3). This method has fairly low coverage rates for this missing mechanism when missing rates are higher than 20 percent. The reason is that this method substantially underestimated the true mean for this missing mechanism. It seems that imputation methods based on Bayesian theory give better coverage rates under similar conditions, which concurs with Rubin's point of view.

Ratio and ratio with disturbance imputation methods have great coverage rates for missing mechanisms (2), (3), and (5) when tail values or large values are missing at higher probabilities. Although the two methods are not as good for missing mechanism (4) when the incomplete data are more diversified than the true distribution, they are still acceptable when missing rates are lower than 30 percent. With 40 percent missing values, the coverage rates of the two ratio imputation methods are moderately low (from 78 percent for mixed distribution of normal and Chi-square and 90 percent for the normal distribution). This is because the two methods significantly overestimate the mean for this missing mechanism, as shown in our bias analyses.

PROC IMPUTE has very good coverage rates except for missing mechanism (5). Some rates are low for mechanism (5) when missing rates are higher than 25 percent. 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 mean imputation, random imputation, mean with disturbance imputation, ABB, and BB methods. The coverage rates of these methods are too low, especially for missing mechanisms (3) and (5), when missing rates are higher than 20 percent.

Table 5.2.4.1—Coverage rates with single imputation (overall [*])

Missing Mechanism	Distribution	Incomp	Mean Imp.	Ratio Imp.	Hot Deck	Random	Mean +e	Ratio +e	ABB	BB	Proc Impute	Schafer	Adj. DA
1. MCAR	Normal	89.0%	84.5%		93.5%	87.5%	86.5%		85.5%	85.5%	92.0%	96.0%	93.5%
	Dexp	92.0%	85.0%		87.5%	88.5%	88.0%		84.5%	86.0%	93.0%	94.5%	94.5%
	MixNorm	89.5%	85.0%		91.5%	89.5%	85.0%		84.0%	87.0%	93.0%	95.0%	95.5%
	MixNChi	89.5%	84.0%		87.0%	88.5%	87.5%		86.5%	86.0%	92.5%	94.5%	95.5%
2. Unconfounded	Normal	94.0%	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	95.5%	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	89.5%	84.5%	94.5%	85.5%	87.5%	88.0%	96.0%	84.5%	87.0%	94.0%	95.0%	92.5%
missing)	MixNChi	90.5%	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	83.5%	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	83.5%	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	79.5%	76.0%	92.0%	91.0%	80.0%	77.0%	93.5%	80.0%	76.5%	93.0%	94.0%	82.5%
missing)	MixNChi	83.5%	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	92.5%	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	93.5%	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	92.0%	88.5%	92.0%	85.5%	88.5%	87.0%	93.5%	88.0%	89.0%	90.0%	96.5%	96.5%
missing)	MixNChi	90.0%	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	92.5%	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	90.0%	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	91.5%	84.5%	95.5%	85.0%	88.5%	88.0%	95.5%	85.0%	86.0%	84.0%	94.5%	96.0%
missing)	MixNChi	81.5%	74.5%	96.0%	81.0%	75.0%	81.0%	95.0%	74.0%	77.0%	85.0%	95.0%	90.5%

* "Overall" means that the four missing rate categories are combined. Coverage rates for each separate missing rate category are reported in tables 5.2.4.1 to 5.2.4.5.

Table 5.2.4.2—Coverage rates with single imputation with about 10% missing values *

Missing Mechanism	Distribution	Incomp	Mean Imp.	Ratio Imp.	Hot Deck	Random	Mean +e	Ratio +e	ABB	BB	Proc Impute	Schafer	Adj. DA
1. MCAR	Normal	92.0%	92.0%		94.0%	88.0%	92.0%		92.0%	90.0%	92.0%	94.0%	92.0%
	Dexp	96.0%	96.0%		96.0%	94.0%	94.0%		92.0%	94.0%	96.0%	96.0%	98.0%
	MixNorm	92.0%	90.0%		94.0%	92.0%	92.0%		88.0%	90.0%	94.0%	96.0%	98.0%
	MixNChi	96.0%	92.0%		92.0%	92.0%	94.0%		92.0%	96.0%	92.0%	90.0%	94.0%
2. Unconfounded	Normal	92.0%	90.0%	94.0%	92.0%	94.0%	94.0%	94.0%	98.0%	92.0%	92.0%	94.0%	96.0%
(tail values	Dexp	98.0%	98.0%	96.0%	100%	88.0%	94.0%	98.0%	96.0%	98.0%	96.0%	98.0%	92.0%
more likely	MixNorm	88.0%	84.0%	94.0%	86.0%	90.0%	86.0%	98.0%	80.0%	88.0%	90.0%	96.0%	90.0%
missing)	MixNChi	92.0%	90.0%	92.0%	90.0%	92.0%	94.0%	92.0%	90.0%	98.0%	92.0%	94.0%	96.0%
3. Unconfounded	Normal	94.0%	94.0%	96.0%	98.0%	94.0%	94.0%	96.0%	92.0%	92.0%	94.0%	96.0%	94.0%
(large values	Dexp	92.0%	90.0%	96.0%	94.0%	88.0%	88.0%	96.0%	92.0%	90.0%	94.0%	96.0%	92.0%
more likely	MixNorm	92.0%	92.0%	94.0%	96.0%	94.0%	92.0%	96.0%	92.0%	92.0%	96.0%	96.0%	96.0%
missing)	MixNChi	90.0%	86.0%	96.0%	96.0%	90.0%	88.0%	96.0%	88.0%	86.0%	94.0%	96.0%	94.0%
4. Unconfounded	Normal	96.0%	96.0%	98.0%	96.0%	96.0%	96.0%	96.0%	96.0%	94.0%	98.0%	98.0%	98.0%
(Center values	Dexp	100%	100%	100%	98.0%	100%	94.0%	100%	100%	100%	100%	100%	100%
more likely	MixNorm	94.0%	94.0%	96.0%	94.0%	92.0%	90.0%	94.0%	90.0%	96.0%	94.0%	94.0%	96.0%
missing)	MixNChi	92.0%	90.0%	94.0%	90.0%	90.0%	90.0%	90.0%	90.0%	92.0%	94.0%	94.0%	94.0%
5. Confounded	Normal	92.0%	88.0%	92.0%	92.0%	88.0%	92.0%	92.0%	92.0%	88.0%	90.0%	92.0%	88.0%
(tail values	Dexp	92.0%	90.0%	92.0%	92.0%	92.0%	86.0%	92.0%	92.0%	90.0%	92.0%	92.0%	96.0%
more likely	MixNorm	94.0%	94.0%	98.0%	96.0%	96.0%	94.0%	98.0%	92.0%	92.0%	98.0%	96.0%	96.0%
missing)	MixNChi	88.0%	86.0%	98.0%	92.0%	90.0%	88.0%	98.0%	88.0%	86.0%	92.0%	96.0%	92.0%

* There are about 5% missing values for missing mechanism 3.

Table 5.2.4.3—Coverage rates with single imputation with about 20% missing values *

Missing Mechanism	Distribution	Incomp	Mean Imp.	Ratio Imp.	Hot Deck	Random	Mean +e	Ratio +e	ABB	BB	Proc Impute	Schafer	Adj. DA
1. MCAR	Normal	88.0%	86.0%		98.0%	90.0%	86.0%		88.0%	86.0%	92.0%	94.0%	96.0%
	Dexp	90.0%	86.0%		88.0%	90.0%	88.0%		86.0%	88.0%	92.0%	90.0%	92.0%
	MixNorm	96.0%	88.0%		96.0%	94.0%	88.0%		88.0%	88.0%	96.0%	98.0%	96.0%
	MixNChi	98.0%	94.0%		98.0%	98.0%	92.0%		98.0%	94.0%	100%	100%	98.0%
2. Unconfounded	Normal	98.0%	96.0%	98.0%	94.0%	96.0%	94.0%	98.0%	94.0%	96.0%	98.0%	96.0%	100%
(tail values	Dexp	94.0%	94.0%	94.0%	92.0%	94.0%	94.0%	94.0%	90.0%	94.0%	92.0%	94.0%	96.0%
more likely	MixNorm	98.0%	96.0%	98.0%	94.0%	96.0%	98.0%	98.0%	96.0%	96.0%	98.0%	96.0%	96.0%
missing)	MixNChi	94.0%	90.0%	100%	84.0%	92.0%	92.0%	100.0%	94.0%	92.0%	96.0%	98.0%	96.0%
3. Unconfounded	Normal	94.0%	94.0%	98.0%	98.0%	92.0%	84.0%	98.0%	92.0%	90.0%	98.0%	98.0%	96.0%
(large values	Dexp	82.0%	82.0%	92.0%	86.0%	82.0%	80.0%	90.0%	80.0%	84.0%	88.0%	88.0%	84.0%
more likely	MixNorm	82.0%	78.0%	94.0%	96.0%	80.0%	76.0%	94.0%	84.0%	78.0%	96.0%	96.0%	82.0%
missing)	MixNChi	84.0%	84.0%	94.0%	94.0%	88.0%	88.0%	98.0%	88.0%	86.0%	94.0%	100.0%	92.0%
4. Unconfounded	Normal	92.0%	90.0%	92.0%	88.0%	88.0%	94.0%	94.0%	92.0%	92.0%	92.0%	96.0%	96.0%
(Center values	Dexp	90.0%	86.0%	88.0%	86.0%	84.0%	88.0%	90.0%	88.0%	86.0%	88.0%	94.0%	90.0%
more likely	MixNorm	92.0%	92.0%	96.0%	94.0%	88.0%	92.0%	98.0%	90.0%	92.0%	94.0%	98.0%	98.0%
missing)	MixNChi	88.0%	88.0%	94.0%	88.0%	90.0%	88.0%	92.0%	92.0%	90.0%	88.0%	96.0%	96.0%
5. Confounded	Normal	98.0%	92.0%	98.0%	92.0%	96.0%	90.0%	94.0%	90.0%	92.0%	96.0%	98.0%	100%
(tail values	Dexp	98.0%	94.0%	98.0%	90.0%	96.0%	96.0%	94.0%	98.0%	94.0%	92.0%	96.0%	100%
more likely	MixNorm	100%	96.0%	98.0%	88.0%	100%	94.0%	96.0%	92.0%	98.0%	90.0%	98.0%	100%
missing)	MixNChi	82.0%	76.0%	94.0%	78.0%	72.0%	82.0%	94.0%	72.0%	80.0%	86.0%	96.0%	88.0%

* There are about 10% missing values for missing mechanism 3.

Table 5.2.4.4—Coverage rates with single imputation with about 30% missing values *

Missing Mechanism	Distribution	Incomp	Mean Imp.	Ratio Imp.	Hot Deck	Random	Mean +e	Ratio +e	ABB	BB	Proc Impute	Schafer	Adj. DA
1. MCAR	Normal	84.0%	80.0%		88.0%	86.0%	80.0%		82.0%	86.0%	96.0%	98.0%	94.0%
	Dexp	88.0%	74.0%		84.0%	84.0%	84.0%		78.0%	82.0%	88.0%	96.0%	92.0%
	MixNorm	80.0%	76.0%		86.0%	82.0%	74.0%		76.0%	90.0%	88.0%	92.0%	90.0%
	MixNChi	86.0%	82.0%		82.0%	78.0%	80.0%		82.0%	84.0%	88.0%	94.0%	96.0%
2. Unconfounded	Normal	96.0%	90.0%	98.0%	96.0%	92.0%	86.0%	98.0%	98.0%	94.0%	94.0%	98.0%	96.0%
(tail values	Dexp	98.0%	98.0%	100%	88.0%	96.0%	96.0%	98.0%	94.0%	98.0%	98.0%	98.0%	100%
more likely	MixNorm	90.0%	82.0%	94.0%	90.0%	84.0%	88.0%	94.0%	86.0%	86.0%	90.0%	94.0%	90.0%
missing)	MixNChi	86.0%	84.0%	88.0%	94.0%	90.0%	84.0%	90.0%	86.0%	86.0%	84.0%	86.0%	100%
3. Unconfounded	Normal	82.0%	80.0%	90.0%	90.0%	82.0%	80.0%	92.0%	82.0%	80.0%	94.0%	96.0%	84.0%
(large values	Dexp	84.0%	82.0%	94.0%	92.0%	80.0%	82.0%	96.0%	84.0%	82.0%	94.0%	96.0%	84.0%
more likely	MixNorm	74.0%	70.0%	84.0%	80.0%	72.0%	68.0%	88.0%	72.0%	68.0%	88.0%	90.0%	74.0%
missing)	MixNChi	82.0%	82.0%	88.0%	84.0%	78.0%	82.0%	88.0%	82.0%	82.0%	92.0%	92.0%	84.0%
4. Unconfounded	Normal	92.0%	82.0%	90.0%	90.0%	88.0%	84.0%	88.0%	86.0%	88.0%	96.0%	100.0%	96.0%
(Center values	Dexp	90.0%	86.0%	86.0%	84.0%	88.0%	92.0%	92.0%	86.0%	92.0%	88.0%	88.0%	94.0%
more likely	MixNorm	90.0%	82.0%	90.0%	74.0%	84.0%	80.0%	94.0%	86.0%	84.0%	94.0%	94.0%	94.0%
missing)	MixNChi	92.0%	88.0%	92.0%	92.0%	90.0%	86.0%	86.0%	84.0%	92.0%	94.0%	90.0%	94.0%
5. Confounded	Normal	90.0%	88.0%	92.0%	82.0%	90.0%	84.0%	90.0%	90.0%	84.0%	88.0%	96.0%	96.0%
(tail values	Dexp	86.0%	80.0%	96.0%	92.0%	86.0%	84.0%	92.0%	82.0%	88.0%	86.0%	96.0%	96.0%
more likely	MixNorm	86.0%	76.0%	96.0%	76.0%	80.0%	82.0%	96.0%	76.0%	74.0%	70.0%	88.0%	92.0%
missing)	MixNChi	78.0%	74.0%	94.0%	80.0%	72.0%	80.0%	92.0%	70.0%	68.0%	86.0%	94.0%	92.0%

* There are about 15% missing values for missing mechanism 3.

Table 5.2.4.5—Coverage rates with single imputation with about 40% missing values *

Missing Mechanism	Distribution	Incomp	Mean Imp.	Ratio Imp.	Hot Deck	Random	Mean +e	Ratio +e	ABB	BB	Proc Impute	Schafer	Adj. DA
1. MCAR	Normal	92.0%	80.0%		94.0%	86.0%	88.0%		80.0%	80.0%	88.0%	98.0%	92.0%
	Dexp	94.0%	84.0%		82.0%	86.0%	86.0%		82.0%	80.0%	96.0%	96.0%	96.0%
	MixNorm	90.0%	86.0%		90.0%	90.0%	86.0%		84.0%	80.0%	94.0%	94.0%	98.0%
	MixNChi	78.0%	68.0%		76.0%	86.0%	84.0%		74.0%	70.0%	90.0%	94.0%	94.0%
2. Unconfounded	Normal	90.0%	82.0%	96.0%	86.0%	86.0%	82.0%	90.0%	94.0%	90.0%	90.0%	98.0%	94.0%
(tail values	Dexp	92.0%	86.0%	96.0%	74.0%	90.0%	90.0%	96.0%	86.0%	88.0%	98.0%	98.0%	100%
more likely	MixNorm	82.0%	76.0%	92.0%	72.0%	80.0%	80.0%	94.0%	76.0%	78.0%	98.0%	94.0%	94.0%
missing)	MixNChi	90.0%	86.0%	96.0%	86.0%	86.0%	86.0%	92.0%	84.0%	88.0%	90.0%	96.0%	96.0%
3. Unconfounded	Normal	64.0%	54.0%	94.0%	88.0%	58.0%	60.0%	92.0%	58.0%	64.0%	94.0%	96.0%	74.0%
(large values	Dexp	76.0%	74.0%	96.0%	96.0%	72.0%	74.0%	90.0%	66.0%	74.0%	92.0%	96.0%	82.0%
more likely	MixNorm	70.0%	64.0%	96.0%	92.0%	74.0%	72.0%	96.0%	72.0%	68.0%	92.0%	94.0%	78.0%
missing)	MixNChi	78.0%	76.0%	94.0%	92.0%	76.0%	78.0%	94.0%	76.0%	72.0%	94.0%	98.0%	86.0%
4. Unconfounded	Normal	90.0%	84.0%	86.0%	80.0%	90.0%	82.0%	88.0%	82.0%	88.0%	90.0%	94.0%	96.0%
(Center values	Dexp	94.0%	82.0%	90.0%	76.0%	88.0%	86.0%	92.0%	84.0%	84.0%	84.0%	92.0%	96.0%
more likely	MixNorm	92.0%	86.0%	86.0%	80.0%	90.0%	86.0%	88.0%	86.0%	84.0%	78.0%	100.0%	98.0%
missing)	MixNChi	88.0%	78.0%	78.0%	82.0%	86.0%	82.0%	80.0%	84.0%	90.0%	92.0%	96.0%	98.0%
5. Confounded	Normal	90.0%	80.0%	98.0%	90.0%	92.0%	84.0%	94.0%	88.0%	80.0%	90.0%	96.0%	100%
(tail values	Dexp	84.0%	72.0%	98.0%	90.0%	82.0%	72.0%	98.0%	78.0%	80.0%	84.0%	98.0%	100%
more likely	MixNorm	86.0%	72.0%	90.0%	80.0%	78.0%	82.0%	92.0%	80.0%	80.0%	78.0%	96.0%	96.0%
missing)	MixNChi	78.0%	62.0%	98.0%	74.0%	66.0%	74.0%	96.0%	66.0%	74.0%	76.0%	94.0%	90.0%

* There are about 20% missing values for missing mechanism 3.

5.2.5 Confidence interval width

A 95 percent confidence interval width was obtained via the distribution of the 200 mean estimates based on the 200 simulation replications. The lower confidence limit was equal to the average of the fifth and sixth smallest estimates, and the upper confidence limit was equal to the average of the fifth and sixth largest estimates. Shorter confidence interval alone does not necessarily imply a better method. A method which provides shorter confidence intervals with higher coverage rates is generally preferred because the method is more likely to provide more concentrated point estimates around the true values.

Table 5.2.5.1 presents the confidence interval widths for the estimates based on the complete data and the data imputed by the 11 imputation methods. For missing mechanisms (2), (3), and (5), tail values or large values are more likely missing and the incomplete data are less diversified than the true distribution, and so are the imputed data. Therefore, the estimates based on the imputed data tend to have less variation than the complete data, and consequently the confidence intervals tend to be too short. This tendency can especially be seen in missing mechanism (5). The readers may need to compare the methods in terms of confidence interval widths along with the biases of variance estimates discussed in section 5.2.2 and coverage rates described in section 5.2.4.

On the other hand, for missing mechanism (4), the incomplete data are more diversified than the complete data, and therefore the estimates based on the imputed data tend to have more variation. Consequently, the confidence intervals based on the imputed data tend to be too wide.

Overall, Schafer's software and the adjusted data augmentation method have the shortest confidence intervals across the five missing mechanism. We also found in the preceding section that these 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.

Table 5.2.5.1—Confidence interval width with single imputation (overall *)

Missing Mechanism	Distribution	Comp.	Mean Imp.	Ratio Imp.	Hot Deck	Random	Mean +e	Ratio +e	ABB	BB	Proc Impute	Schafer	Adj. DA
1. MCAR	Normal	0.375	0.453		0.417	0.496	0.518		0.491	0.488	0.466	0.390	0.393
	Dexp	0.564	0.629		0.689	0.610	0.713		0.681	0.685	0.598	0.557	0.497
	MixNorm	0.429	0.494		0.532	0.598	0.618		0.634	0.585	0.478	0.428	0.481
	MixNChi	0.797	1.015		1.504	1.179	1.094		1.134	1.289	0.847	0.841	0.959
2. Unconfounded	Normal	0.369	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	0.513	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	0.425	0.481	0.459	0.618	0.515	0.507	0.515	0.544	0.538	0.465	0.447	0.444
missing)	MixNChi	0.884	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	0.362	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	0.543	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	0.483	0.527	0.532	0.493	0.525	0.550	0.543	0.567	0.510	0.519	0.488	0.465
missing)	MixNChi	0.846	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	0.376	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	0.521	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	0.472	0.554	0.496	0.617	0.612	0.600	0.523	0.601	0.622	0.564	0.436	0.474
missing)	MixNChi	0.893	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	0.411	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	0.559	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	0.468	0.432	0.388	0.512	0.436	0.473	0.450	0.529	0.492	0.429	0.376	0.353
missing)	MixNChi	0.765	0.627	0.677	0.678	0.685	0.658	0.769	0.698	0.637	0.773	0.622	0.578

* "Overall" means that the four missing rate categories are combined.

5.2.6 Bias of quartile estimates

We obtained estimates of median and the first and third quartiles for all imputed data to investigate how imputation affects the data distribution. Tables 5.2.6.1–5.2.6.3 give the biases of the first quartile, the third quartile, and the median estimates, respectively, for the combined missing rate categories.

The mean imputation method is obviously the worst in terms of quartile estimates across all five missing mechanisms. The data are centralized so that the first quartiles are substantially overestimated, while the third quartiles are substantially underestimated. The median estimates are pretty much similar to those of the incomplete data. The only exceptions are the first quartile estimates for missing mechanism (3) in which the positive biases are very small. This is because both missing values created via missing mechanism (3) and the means imputed for the missing values are larger than the first quartiles so that the first quartile estimates based on the imputed data are very close to those based on the complete data. We will not include this method for discussion in this section.

For the MCAR missing mechanism, all methods except the mean and the mean with disturbance imputation methods give fine estimates for all the quartiles. The mean with disturbance imputation method gives fine estimates for the normal and the contaminated normal distributions, but it has significantly larger negative biases of the first quartile estimates and significantly larger positive biases of the third quartile estimates for the double exponential distribution and the mixed distribution of normal and Chi-square. This implies that the disturbance drawn from $N(0, s_{obs}^2)$ diversified the true data, where s_{obs}^2 is calculated from the observed data from the double exponential distribution or the mixed distribution of normal and Chi-square.

For unconfounded missing mechanism (2), since the incomplete data are less diversified than the true distributions, the first quartiles are overestimated while the third quartiles are underestimated. Five methods—Schafer's software, PROC IMPUTE, hot deck, ratio and ratio with disturbance imputation—all substantially reduce the biases of the first and third quartile estimates compared to the incomplete data. The adjusted data augmentation method has slight improvement for the third quartile estimates, but no improvement for the biases of the first quartile estimates. The random, mean with disturbance, ABB, and BB imputation methods do not improve the first and second quartile estimates compared to the incomplete data. For this missing mechanism, all methods provide fine median estimates because values are missing symmetrically at both tails.

For unconfounded missing mechanism (3), since the incomplete data are less diversified than the true distributions, the first quartiles are overestimated while the third quartiles are underestimated by the incomplete data. Similar results to those for mechanism (2) have been found for the first and third quartile estimates. The biases of these quartile estimates based on the data imputed by Schafer's software, ratio imputation, ratio with disturbance imputation, PROC IMPUTE, and hot deck are at least twice smaller than those based on the incomplete data. Among these five methods, hot deck is obviously worse than Schafer's software, PROC IMPUTE, and the ratio imputation method. All other methods except the mean imputation method have some improvement over the incomplete data but it is not substantial. For this missing mechanism, the medians are underestimated by the incomplete data. Schafer's software and PROC IMPUTE reduce the negative biases by 4 to 50 times, while hot deck, ratio imputation, ratio with disturbance imputation reduce the negative biases by 2 to 10 times. All other methods reduce the biases of the incomplete data median estimates slightly.

For unconfounded missing mechanism (4), since the incomplete data are more diversified than the true distribution, the first quartiles are underestimated while the third quartiles are overestimated by the incomplete data. The hot deck method has the best overall performance in terms of biases of quartile estimates, followed by PROC IMPUTE and Schafer's software. Among these three methods, Schafer's software is best for normal distribution, but much worse than hot deck and PROC IMPUTE for the mixed distribution of normal and Chi-square. The other methods do not improve the biases over the incomplete data. Although the ratio imputation method shrinks the diversified incomplete data, the imputed data are shrunk too much so that they have less variation than the true distribution. The magnitudes of the biases of the first quartile estimates are larger than those of the incomplete data, but it is the other way around for the third quartile estimates. On the other hand, the random imputation, ABB, BB, and adjusted data augmentation methods have slightly better first quartile estimates but slightly worse third quartile estimates in terms of bias. All methods except ratio imputation and ratio with disturbance imputation provide as good median estimates as the incomplete data. Ratio imputation and ratio with disturbance imputation worsen the median estimates compared to the incomplete data.

For confounded missing mechanism (5), since the incomplete data are less diversified than the true distributions, the first quartiles are overestimated while the third quartiles are underestimated by the incomplete data. The ratio with disturbance imputation method obviously has the best performance and reduces the biases of the incomplete quartile estimates by two to six times. Ratio imputation and Schafer's software also improve the quartile estimates over the incomplete data. The other methods slightly worsen the first quartile estimates while slightly improving the third quartile estimates. All methods give fine median estimates with this missing mechanism.

Table 5.2.6.1—Biases of the first quartile estimates (overall *)

Missing Mechanism	Distribution	Incomp	Mean Imp.	Ratio Imp.	Hot Deck	Random	Mean +e	Ratio +e	ABB	BB	Proc Impute	Schafer	Adj. DA
1. MCAR	Normal	-0.020	0.251		0.038	-0.001	-0.006		0.007	-0.004	-0.016	-0.013	0.001
	Dexp	-0.022	0.289		0.028	-0.004	-0.062		-0.004	-0.010	0.004	-0.045	-0.007
	MixNorm	-0.015	0.271		0.033	-0.003	-0.012		0.007	0.004	0.004	-0.015	-0.011
	MixNChi	-0.019	0.290		0.044	-0.003	-0.084		0.002	0.008	0.049	-0.058	-0.027
2. Unconfounded	Normal	0.054	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	0.079	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	0.058	0.247	-0.015	-0.003	0.071	0.061	-0.004	0.072	0.072	-0.004	-0.001	0.059
missing)	MixNChi	0.064	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	-0.082	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	-0.096	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	-0.096	0.008	0.008	-0.009	-0.088	-0.085	-0.022	-0.083	-0.079	0.002	0.001	-0.082
missing)	MixNChi	-0.103	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	-0.061	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	-0.113	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	-0.075	0.193	0.111	0.006	-0.064	-0.065	-0.062	-0.061	-0.056	0.023	-0.024	-0.056
missing)	MixNChi	-0.146	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	0.111	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	0.175	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	0.129	0.388	0.096	0.137	0.140	0.127	0.033	0.146	0.150	0.157	0.103	0.135
missing)	MixNChi	0.173	0.467	0.124	0.189	0.206	0.162	0.021	0.192	0.197	0.135	0.143	0.172

* “Overall” means that the four missing rate categories are combined. The results for each separate missing rate category are reported in tables 5.2.6.2 to 5.2.6.3.

Table 5.2.6.2—Biases of the third quartile estimates (overall *)

Missing Mechanism	Distribution	Incomp	Mean Imp.	Ratio Imp.	Hot Deck	Random	Mean +e	Ratio +e	ABB	BB	Proc Impute	Schafer	Adj. DA
1. MCAR	Normal	-0.023	-0.273		-0.008	-0.010	-0.019		-0.004	-0.011	0.012	-0.005	-0.007
	Dexp	-0.016	-0.287		-0.018	0.014	0.063		0.003	-0.001	-0.002	0.043	0.023
	MixNorm	-0.017	-0.274		-0.013	-0.006	0.021		-0.002	-0.011	-0.018	0.004	-0.001
	MixNChi	-0.010	-0.254		-0.001	0.013	0.109		0.014	0.019	0.008	0.061	0.058
2. Unconfounded	Normal	-0.066	-0.207	0.015	-0.005	-0.050	-0.055	0.000	-0.054	-0.047	0.018	0.002	-0.043
(tail values	Dexp	-0.109	-0.265	0.011	0.001	-0.090	-0.072	0.006	-0.089	-0.093	-0.006	-0.005	-0.071
more likely	MixNorm	-0.070	-0.222	0.017	0.012	-0.061	-0.050	0.020	-0.065	-0.057	0.007	-0.004	-0.045
missing)	MixNChi	-0.081	-0.223	-0.004	0.023	-0.068	-0.018	0.007	-0.062	-0.068	-0.033	-0.005	-0.028
3. Unconfounded	Normal	-0.130	-0.193	0.003	-0.027	-0.119	-0.115	-0.003	-0.119	-0.118	0.017	0.004	-0.105
(large values	Dexp	-0.174	-0.262	-0.007	-0.031	-0.157	-0.145	-0.010	-0.158	-0.156	0.024	0.001	-0.140
more likely	MixNorm	-0.146	-0.219	-0.008	-0.022	-0.137	-0.134	0.002	-0.138	-0.125	0.017	-0.001	-0.120
missing)	MixNChi	-0.145	-0.216	-0.010	-0.036	-0.135	-0.101	-0.018	-0.126	-0.128	-0.018	-0.014	-0.108
4. Unconfounded	Normal	0.067	-0.180	-0.057	0.009	0.081	0.071	0.102	0.086	0.081	-0.044	0.008	0.072
(Center values	Dexp	0.062	-0.196	-0.091	-0.004	0.070	0.102	0.136	0.090	0.080	-0.060	0.018	0.057
more likely	MixNorm	0.068	-0.173	-0.060	0.009	0.084	0.095	0.135	0.105	0.103	-0.035	0.019	0.077
missing)	MixNChi	0.077	-0.221	-0.028	0.014	0.106	0.250	0.266	0.105	0.102	-0.039	0.135	0.108
5. Confounded	Normal	-0.151	-0.340	-0.091	-0.122	-0.141	-0.129	-0.037	-0.139	-0.131	-0.158	-0.116	-0.139
(tail values	Dexp	-0.197	-0.448	-0.107	-0.150	-0.179	-0.166	-0.030	-0.174	-0.178	-0.185	-0.139	-0.171
more likely	MixNorm	-0.163	-0.407	-0.102	-0.132	-0.145	-0.141	-0.045	-0.150	-0.150	-0.169	-0.134	-0.135
missing)	MixNChi	-0.188	-0.428	-0.119	-0.120	-0.172	-0.132	-0.034	-0.170	-0.177	-0.140	-0.109	-0.123

* “Overall” means that the four missing rate categories are combined.

Table 5.2.6.3—Biases of median estimates (overall [*])

Missing Mechanism	Distribution	Incomp	Mean Imp.	Ratio Imp.	Hot Deck	Random	Mean +e	Ratio +e	ABB	BB	Proc Impute	Schafer	Adj. DA
1. MCAR	Normal	-0.013	-0.003		0.008	-0.011	-0.006		0.001	-0.005	-0.004	-0.004	-0.009
	Dexp	-0.007	-0.007		0.006	0.005	0.001		0.002	0.000	-0.003	-0.001	0.006
	MixNorm	-0.009	0.005		0.011	-0.003	0.008		0.002	0.001	-0.002	-0.008	-0.005
	MixNChi	0.000	0.029		0.011	0.011	0.010		0.019	0.014	0.044	-0.002	0.017
2. Unconfounded	Normal	-0.001	0.005	-0.001	0.011	0.005	0.008	0.000	0.008	0.011	-0.001	0.000	0.006
(tail values	Dexp	-0.007	-0.006	0.001	0.008	-0.001	-0.001	-0.007	0.002	-0.010	0.000	-0.003	-0.002
more likely	MixNorm	-0.004	-0.005	-0.001	0.008	-0.001	-0.004	0.002	0.003	0.004	0.002	-0.005	0.000
missing)	MixNChi	0.003	0.025	-0.007	0.028	0.008	0.013	-0.002	0.016	0.011	-0.010	0.000	0.012
3. Unconfounded	Normal	-0.099	-0.097	0.021	-0.015	-0.094	-0.094	-0.024	-0.090	-0.096	0.006	-0.003	-0.085
(large values	Dexp	-0.115	-0.118	0.026	-0.025	-0.103	-0.115	-0.026	-0.111	-0.102	0.006	-0.008	-0.093
more likely	MixNorm	-0.116	-0.106	0.026	-0.011	-0.107	-0.112	-0.022	-0.115	-0.104	0.007	-0.002	-0.093
missing)	MixNChi	-0.124	-0.118	0.014	-0.030	-0.116	-0.114	-0.048	-0.110	-0.113	-0.021	-0.033	-0.100
4. Unconfounded	Normal	0.011	0.008	0.056	0.004	0.028	0.013	0.035	0.019	0.020	-0.014	0.001	0.023
(Center values	Dexp	-0.022	-0.006	0.061	-0.026	-0.019	-0.005	0.025	-0.004	-0.006	-0.019	-0.002	-0.021
more likely	MixNorm	0.005	0.010	0.056	-0.007	0.018	0.014	0.032	0.022	0.026	-0.008	0.001	0.015
missing)	MixNChi	-0.006	0.082	0.118	0.010	0.007	0.028	0.047	0.006	0.009	-0.025	0.003	0.005
5. Confounded	Normal	-0.011	0.007	0.015	-0.003	0.006	-0.002	0.004	-0.005	0.004	-0.003	0.001	0.001
(tail values	Dexp	-0.002	-0.001	0.024	0.011	0.008	0.005	0.013	0.006	0.002	0.001	0.010	0.003
more likely	MixNorm	-0.012	-0.008	0.003	-0.001	-0.002	-0.009	-0.006	-0.004	-0.004	-0.007	-0.020	0.001
missing)	MixNChi	0.008	0.004	0.018	0.042	0.014	0.016	0.014	0.017	0.011	0.003	0.019	0.020

* “Overall” means that the four missing rate categories are combined.

5.2.7 Average imputation error

Average imputation error is defined as

$$\sqrt{\frac{1}{m}\sum_{i=1}^{m}(y_i^* - y_i)^2}\ ,$$

where m is the number of missing values, y_i is the true value which is intentionally set to missing, and y_i^* is the imputed value for the $i\text{-}th$ missing case. That an imputation method has smaller average imputation errors only implies that the method provides imputations on average closer to the real values. This does not necessarily means that it gives more accurate estimates for all types of statistics, although this is true in many situations.

Tables 5.2.7.1–5.2.7.5 present average imputation errors for the combined missing rate categories and each separate missing rate category, respectively. The figures in the tables have been standardized by dividing the true standard deviation from the original imputation errors.

Across all missing mechanisms, the random imputation, mean with disturbance imputation, ABB, BB, and adjusted data augmentation methods all have the similar imputation errors that are significantly larger than the imputation errors for the other methods for almost all distributions, all missing rates, and all missing categories.

The ratio imputation method always has the smallest or close to smallest average imputation errors. Schafer's software and PROC IMPUTE are competitive candidates. These three methods have substantially smaller average imputation errors than the others. The hot deck, ratio with disturbance imputation, and mean imputation methods sit in the middle in terms of average imputation error. They are significantly worse than the three best methods, but they are better than the worst five methods. Mean imputation has very small imputation errors for missing mechanism (4) because center values are more likely missing with this missing mechanism and the mean imputation method imputes the mean values for them.

It is also noticed that most methods give fairly consistent average imputation errors, while PROC IMPUTE and hot deck have much larger average imputation errors for the mixed distribution of normal and Chi-square than they do with the other three distributions for all missing mechanisms except mechanism (4). This probably indicates that these two methods are not very good at recovering tail or large missing values.

The relative performance of the imputation methods in terms of average imputation error is very consistent across the missing rate categories.

Table 5.2.7.1—Average imputation error (overall *)

Missing Mechanism	Distribution	Missing rates	Mean Imp.	Ratio Imp.	Hot Deck	Random	Mean +e	Ratio +e	ABB	BB	Proc Impute	Schafer	Adj. DA
1. MCAR	Normal	24.7%	1.008		0.971	1.402	1.406		1.406	1.432	0.671	0.607	1.405
	Dexp	25.0%	0.970		1.068	1.392	1.411		1.375	1.393	0.584	0.611	1.394
	MixNorm	24.9%	0.997		1.114	1.378	1.386		1.397	1.371	0.591	0.626	1.388
	MixNChi	24.9%	0.989		1.721	1.362	1.417		1.386	1.477	0.836	0.628	1.396
2. Unconfounded	Normal	17.7%	1.258	0.496	1.330	1.581	1.556	1.056	1.556	1.569	0.671	0.516	1.622
(tail values	Dexp	18.0%	1.438	0.567	1.594	1.671	1.678	1.006	1.653	1.670	0.576	0.521	1.746
more likely	MixNorm	18.5%	1.359	0.543	1.558	1.597	1.615	1.033	1.605	1.612	0.576	0.519	1.654
missing)	MixNChi	16.8%	1.759	0.723	2.141	1.905	1.896	1.027	1.911	1.920	1.378	0.481	1.992
3. Unconfounded	Normal	9.5%	1.244	0.609	0.995	1.574	1.566	1.183	1.580	1.601	0.825	0.742	1.534
(large values	Dexp	9.2%	1.225	0.638	1.150	1.571	1.579	1.129	1.582	1.560	0.919	0.769	1.534
more likely	MixNorm	9.8%	1.227	0.616	1.081	1.554	1.558	1.119	1.583	1.568	0.924	0.739	1.512
missing)	MixNChi	9.1%	1.467	0.618	1.638	1.702	1.739	1.105	1.735	1.775	1.295	0.736	1.684
4. Unconfounded	Normal	22.5%	0.781	0.737	1.042	1.307	1.317	1.300	1.332	1.346	0.844	0.951	1.258
(Center values	Dexp	19.6%	0.762	0.739	1.064	1.307	1.303	1.277	1.310	1.330	0.841	0.941	1.254
more likely	MixNorm	21.0%	0.766	0.722	1.011	1.328	1.331	1.300	1.329	1.331	0.833	0.939	1.272
missing)	MixNChi	23.9%	0.781	0.774	1.103	1.375	1.343	1.321	1.420	1.321	0.789	0.969	1.285
5. Confounded	Normal	25.1%	1.332	0.978	1.382	1.546	1.557	1.405	1.569	1.557	1.167	1.237	1.560
(tail values	Dexp	26.9%	1.379	1.101	1.431	1.570	1.556	1.414	1.580	1.564	1.239	1.291	1.581
more likely	MixNorm	27.1%	1.378	1.094	1.419	1.573	1.575	1.420	1.561	1.579	1.223	1.287	1.587
missing)	MixNChi	27.1%	1.663	1.416	1.632	1.746	1.740	1.602	1.745	1.736	1.569	1.423	1.776

* "Overall" means that the four missing rate categories are combined. Average imputation errors for each separate missing rate category are reported in tables 5.2.7.2 to 5.2.7.5.

Table 5.2.7.2—Average imputation error with about 10% missing values *

Missing Mechanism	Distribution	Mean Imp.	Ratio Imp.	Hot Deck	Random	Mean +e	Ratio +e	ABB	BB	Proc Impute	Schafer	Adj DA
1. MCAR	Normal	1.061		0.822	1.450	1.375		1.440	1.452	0.641	0.632	1.458
	Dexp	0.993		0.782	1.358	1.401		1.368	1.366	0.583	0.634	1.377
	MixNorm	1.032		0.962	1.323	1.417		1.387	1.476	0.635	0.615	1.330
	MixNChi	1.164		1.077	1.522	1.585		1.540	1.848	1.162	0.545	1.567
2. Unconfounded	Normal	1.285	0.472	1.014	1.683	1.641	1.111	1.557	1.585	0.479	0.523	1.728
(tail values	Dexp	1.558	0.475	1.615	1.865	1.802	0.973	1.774	1.781	0.649	0.514	1.924
more likely	MixNorm	1.510	0.511	1.656	1.726	1.764	1.066	1.728	1.702	0.667	0.546	1.755
missing)	MixNChi	2.288	0.706	2.420	2.385	2.394	1.049	2.391	2.386	1.485	0.474	2.458
3. Unconfounded	Normal	1.272	0.588	0.948	1.560	1.616	1.167	1.619	1.635	0.737	0.708	1.546
(large values	Dexp	1.306	0.634	1.040	1.643	1.609	1.070	1.570	1.725	0.883	0.757	1.628
more likely	MixNorm	1.291	0.685	1.252	1.613	1.589	1.148	1.692	1.636	0.818	0.763	1.597
missing)	MixNChi	1.909	0.672	2.144	2.052	2.150	1.221	2.091	2.094	1.578	0.760	2.001
4. Unconfounded	Normal	0.771	0.775	1.036	1.267	1.323	1.273	1.246	1.276	0.835	0.950	1.246
(Center values	Dexp	0.647	0.650	1.008	1.167	1.178	1.202	1.199	1.246	0.779	0.899	1.147
more likely	MixNorm	0.768	0.758	0.981	1.326	1.308	1.227	1.276	1.231	0.846	0.953	1.318
missing)	MixNChi	1.138	1.143	1.295	1.666	1.505	1.475	1.411	1.449	1.120	1.233	1.581
5. Confounded	Normal	1.412	0.981	1.411	1.755	1.646	1.342	1.662	1.724	1.205	1.271	1.769
(tail values	Dexp	1.470	1.158	1.523	1.703	1.698	1.500	1.742	1.730	1.246	1.379	1.696
more likely	MixNorm	1.511	1.234	1.560	1.751	1.747	1.542	1.679	1.741	1.332	1.432	1.752
missing)	MixNChi	2.212	1.886	2.154	2.348	2.377	2.079	2.349	2.268	2.163	1.930	2.369

* There are about 5% missing values for missing mechanism 3.

Table 5.2.7.3—Average imputation error with about 20% missing values *

Missing Mechanism	Distribution	Mean Imp.	Ratio Imp.	Hot Deck	Random	Mean +e	Ratio +e	ABB	BB	Proc Impute	Schafer	Adj DA
1. MCAR	Normal	0.983		0.855	1.430	1.389		1.377	1.416	0.609	0.583	1.434
	Dexp	1.013		0.984	1.447	1.437		1.376	1.424	0.601	0.592	1.450
	MixNorm	1.021		0.966	1.378	1.352		1.428	1.383	0.599	0.628	1.399
	MixNChi	1.154		1.362	1.496	1.511		1.433	1.389	0.774	0.684	1.525
2. Unconfounded	Normal	1.307	0.520	1.286	1.662	1.560	1.088	1.586	1.612	0.618	0.514	1.709
(tail values	Dexp	1.497	0.534	1.607	1.711	1.736	1.040	1.727	1.673	0.582	0.508	1.775
more likely	MixNorm	1.340	0.543	1.581	1.556	1.648	1.061	1.536	1.617	0.571	0.530	1.611
missing)	MixNChi	1.649	0.751	2.306	1.782	1.838	1.083	1.823	1.885	1.105	0.490	1.902
3. Unconfounded	Normal	1.274	0.622	1.052	1.580	1.565	1.186	1.560	1.687	0.730	0.745	1.562
(large values	Dexp	1.315	0.694	1.249	1.712	1.720	1.192	1.624	1.616	0.840	0.805	1.673
more likely	MixNorm	1.289	0.580	1.124	1.647	1.615	1.105	1.548	1.601	0.991	0.650	1.627
missing)	MixNChi	1.635	0.632	1.773	1.847	1.852	1.002	1.813	1.905	1.540	0.732	1.844
4. Unconfounded	Normal	0.718	0.703	1.033	1.264	1.256	1.273	1.328	1.277	0.828	0.929	1.235
(Center values	Dexp	0.772	0.762	1.091	1.300	1.292	1.336	1.284	1.379	0.851	0.922	1.250
more likely	MixNorm	0.735	0.705	0.984	1.241	1.268	1.318	1.269	1.297	0.808	0.920	1.208
missing)	MixNChi	0.800	0.805	1.063	1.328	1.372	1.254	1.301	1.491	0.796	0.992	1.273
5. Confounded	Normal	1.364	0.994	1.359	1.616	1.608	1.428	1.608	1.583	1.158	1.269	1.633
(tail values	Dexp	1.472	1.145	1.468	1.708	1.673	1.524	1.661	1.693	1.266	1.337	1.718
more likely	MixNorm	1.436	1.064	1.422	1.645	1.641	1.413	1.659	1.683	1.243	1.297	1.671
missing)	MixNChi	1.777	1.539	1.727	1.841	1.835	1.677	1.861	1.860	1.617	1.513	1.868

* There are about 10% missing values for missing mechanism 3.

Table 5.2.7.4—Average imputation error with about 30% missing values *

Missing Mechanism	Distribution	Mean Imp.	Ratio Imp.	Hot Deck	Random	Mean +e	Ratio +e	ABB	BB	Proc Impute	Schafer	Adj DA
1. MCAR	Normal	1.014		0.991	1.381	1.413		1.431	1.433	0.669	0.608	1.386
	Dexp	0.972		1.059	1.347	1.440		1.416	1.370	0.601	0.638	1.350
	MixNorm	1.003		1.118	1.369	1.407		1.410	1.364	0.550	0.615	1.372
	MixNChi	0.919		1.677	1.358	1.405		1.395	1.390	0.791	0.667	1.334
2. Unconfounded	Normal	1.262	0.482	1.453	1.552	1.529	1.011	1.589	1.567	0.679	0.526	1.602
(tail values	Dexp	1.423	0.624	1.652	1.626	1.661	1.045	1.640	1.662	0.573	0.530	1.708
more likely	MixNorm	1.385	0.587	1.528	1.634	1.619	1.040	1.627	1.639	0.565	0.516	1.707
missing)	MixNChi	1.745	0.646	2.109	1.843	1.873	0.957	1.908	1.881	1.558	0.484	1.956
3. Unconfounded	Normal	1.226	0.622	0.971	1.609	1.592	1.161	1.588	1.569	0.800	0.767	1.565
(large values	Dexp	1.191	0.598	1.203	1.531	1.550	1.108	1.548	1.538	0.909	0.747	1.492
more likely	MixNorm	1.219	0.628	1.112	1.511	1.611	1.132	1.631	1.602	0.911	0.790	1.459
missing)	MixNChi	1.369	0.593	1.616	1.662	1.712	1.130	1.727	1.781	1.152	0.685	1.626
4. Unconfounded	Normal	0.766	0.732	1.018	1.284	1.323	1.303	1.347	1.342	0.835	0.947	1.231
(Center values	Dexp	0.761	0.729	1.027	1.328	1.363	1.270	1.290	1.327	0.825	0.955	1.281
more likely	MixNorm	0.747	0.709	1.068	1.341	1.351	1.304	1.311	1.363	0.834	0.951	1.276
missing)	MixNChi	0.642	0.629	0.846	1.201	1.244	1.227	1.285	1.242	0.657	0.841	1.096
5. Confounded	Normal	1.328	0.970	1.380	1.518	1.567	1.418	1.564	1.544	1.161	1.220	1.537
(tail values	Dexp	1.398	1.111	1.475	1.579	1.570	1.399	1.621	1.581	1.292	1.325	1.597
more likely	MixNorm	1.366	1.110	1.407	1.576	1.577	1.414	1.539	1.564	1.225	1.301	1.585
missing)	MixNChi	1.582	1.356	1.580	1.682	1.666	1.547	1.669	1.663	1.476	1.369	1.718

* There are about 15% missing values for missing mechanism 3.

Table 5.2.7.5—Average imputation error with about 40% missing values *

Missing Mechanism	Distribution	Mean Imp.	Ratio Imp.	Hot Deck	Random	Mean +e	Ratio +e	ABB	BB	Proc Impute	Schafer	Adj DA
1. MCAR	Normal	1.005		1.043	1.390	1.417		1.394	1.435	0.709	0.614	1.393
	Dexp	0.940		1.172	1.405	1.377		1.346	1.400	0.562	0.595	1.401
	MixNorm	0.971		1.213	1.400	1.378		1.373	1.341	0.605	0.637	1.409
	MixNChi	0.909		2.002	1.255	1.337		1.317	1.488	0.811	0.586	1.336
2. Unconfounded	Normal	1.209	0.499	1.353	1.510	1.543	1.051	1.505	1.535	0.754	0.505	1.538
(tail values	Dexp	1.358	0.567	1.519	1.602	1.600	0.956	1.560	1.630	0.544	0.525	1.686
more likely	MixNorm	1.287	0.516	1.531	1.541	1.529	0.994	1.585	1.553	0.551	0.505	1.599
missing)	MixNChi	1.602	0.769	1.875	1.835	1.719	1.027	1.754	1.756	1.380	0.474	1.878
3. Unconfounded	Normal	1.232	0.597	0.998	1.546	1.526	1.205	1.571	1.565	0.920	0.730	1.488
(large values	Dexp	1.173	0.638	1.083	1.499	1.511	1.127	1.588	1.491	0.977	0.770	1.455
more likely	MixNorm	1.174	0.602	0.959	1.512	1.466	1.107	1.521	1.495	0.928	0.734	1.454
missing)	MixNChi	1.254	0.612	1.356	1.504	1.526	1.109	1.556	1.561	1.122	0.769	1.504
4. Unconfounded	Normal	0.826	0.749	1.066	1.356	1.341	1.319	1.345	1.401	0.862	0.966	1.292
(Center values	Dexp	0.783	0.757	1.092	1.327	1.291	1.273	1.362	1.330	0.863	0.948	1.261
more likely	MixNorm	0.796	0.732	0.985	1.361	1.352	1.307	1.386	1.347	0.841	0.935	1.288
missing)	MixNChi	0.739	0.725	1.225	1.423	1.354	1.369	1.562	1.260	0.760	0.963	1.325
5. Confounded	Normal	1.300	0.975	1.388	1.474	1.502	1.400	1.530	1.510	1.167	1.225	1.486
(tail values	Dexp	1.292	1.055	1.351	1.456	1.445	1.345	1.460	1.436	1.180	1.217	1.466
more likely	MixNorm	1.319	1.056	1.386	1.479	1.491	1.394	1.492	1.489	1.182	1.228	1.495
missing)	MixNChi	1.447	1.203	1.418	1.512	1.499	1.426	1.502	1.516	1.395	1.215	1.543

* There are about 20% missing values for missing mechanism 3.

References

American Institutes for Research. (1980). Guidebook for imputation of missing data. Statistical Analysis Group in Education for the National Center for Education Statistics, Palo Alto, CA.

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

Bello, A. L. (1994). A bootstrap method for using imputation techniques for data with missing values. *Biometrical Journal*, 36(4), 453–464.

Bendel, R.B. (1978). Population correlation matrices for sampling experiments. *Communication in Statistics*, B7(2), 163–182.

Binder, D. A. (1996). Comment to articles by Rubin, Fay and Rao. *Journal of the American Statistical Association*, 91, 510–512.

Bishop, Y. M. M., Fienberg, S. E. and Holland, P. W. (1975). *Discrete Multivariate Analysis: Theory and Practice*. Cambridge, MA: MIT Press.

Champney, T. F. and Bell, R. (1982). Imputation of income: a procedural comparison. *American Statistical Association Proceedings of the Section on Survey Research Methods*, 431–436.

Cohen, M. P. (1996). A new approach to imputation. *American Statistical Association Proceedings of the Section on Survey Research Methods*, 293–298.

Dempster, A. P., Laird, N. M. and Rubin, D. R. (1977). Maximum likelihood estimation from incomplete data via the EM algorithm (with discussion). *J. R. Statist. Soc*., B39, 1–38.

Dear, R. E. (1959). *A principal-component missing data method for multiple regression methods*. Report SP-86. System Development Corporation. Santa Monica, CA.

Eltinge, J. L. (1996). Comment to articles by Rubin, Fay and Rao. *Journal of the American Statistical Association*, 91, 513–515.

Fay, R. E. (1996). Alternative paradigms for the analysis of imputed survey data. *Journal of the American Statistical Association*, 91, 490–498.

Ford, B. L. (1980). *Incomplete data in sample surveys: the theory of current practices*. National Academy of Sciences, Panel on Incomplete Data.

Gelman, A., Rubin, D. B., Carlin, J. and Stern, H. (1995) *Bayesian Data Analysis*. London: Chapman & Hall.

Geman, S. and Geman, D. (1984). Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 6, 721–741.

Gimotty, P. A. and Brown, M. B. (1990). Imputation procedures for categorical data: their effects on the goodness-of-fit chi-square statistic. *Communication in Statistics: Simulation*, 19(2), 681–703.

Gimotty, P. A. (1987). The asymptotic distribution of the goodness-of-fit chi-square statistic computed from imputed data. *Communication in Statistics A*, 16(1), 45–60.

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. *Journal of the American Statistical Association*, 77, 251–261.

Huseby, J. R., Schwertman, N. C. and Allen, D. M. (1980). Computation of the mean vector and dispersion matrix for incomplete multivariate data. *Communication in Statistics*, B3, 301–309.

Judkins, D. R. (1996). Comment to articles by Rubin, Fay and Rao. *Journal of the American Statistical Association*, 91, 507–510.

Kalton, G. and Kasprzyk, D. (1982). Imputing for missing survey responses. *American Statistical Association Proceedings of the Section on Survey Research Methods*, 22–31.

Kalton, G. and Kish, L. (1984). Some efficient random imputation methods. *Commun. Statist.-Theor. Meth.*, 13(16), 1919–1939.

Kish, L. (1965). *Survey Sampling*. Wiley, New York.

Kong, A., Liu, J. S. and Wong, W. H. (1994). Sequential imputations and Bayesian missing data problems. *Journal of the American Statistical Association*, 89, 278–288.

Krzanowski, W. J. (1988). Missing value imputation in multivariate data using the singular value decomposition of a matrix. *Biometrical Letters*, 25, 31–39.

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

Little, R. J. A. and Rubin, D. B. (1987). *Statistical Analysis with Missing Data.* New York: Wiley.

Meng, X. L. and Rubin, D. B. (1991). Using EM to obtain asymptotic variance-covariance matrices: the SEM algorithm. *Journal of the American Statistical Association*, 86, 899–909.

Olkin, I. And Tate, R. F. (1961). Multivariate correlation models with mixed discrete and continuous variables. *Annals of Mathematical Statistics*, 32, 448–465.

Rancourt, E., Lee, H. and S@rndal, C. E. (1994). Bias corrections for survey estimates from data with ratio imputed values for confounded nonresponse. *Survey Methodology*, 20, 37–147.

Rao, J. N. K. (1996). On variance estimation with imputed survey data. *Journal of the American Statistical Association*, 91, 499–506.

Rao, J. N. K. and Shao, J. (1992). Jackknife variance estimation with survey data under hot-deck imputation. *Biometrika*, 79, 811–822.

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.

Rosenbaum, P. R. and Rubin, D. B. (1984). Reducing bias in observational studies using subclassification in the propensity score. *Journal of the American Statistical Association*, 79, 516–524.

Rubin, D. B. (1978). Multiple imputations in sample surveys—A phenomenological Bayesian approach to nonresponse. *American Statistical Association Proceedings of the Section on Survey Research Methods*, 20–28.

Rubin, D. B. (1981). The Bayesian bootstrap. *Annals of Statistics*, 9, 130-134.

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. *Journal of the American Statistical Association*, 91, 473–489.

Rubin, D. B. and Schenker, N. (1986). Multiple imputation for interval estimation from simple random samples with ignorable nonresponse. *Journal of the American Statistical Association*, 81, 366–374.

Rubin, D. B. and Schenker, N. (1991). Multiple imputation in health-care databases: an overview and some application. *Statistics in Medicine*, 10, 585–598.

S@rndal, C.-E. (1992). Methods for estimating the precision of survey estimates when imputation has been used. *Survey Methodology*, 18, 241–252.

Schafer, J. L. (1997). *Analysis of Incomplete Multivariate Data*. New York: Chapman and Hall.

Schenker, N. (1985). *Multiple imputation for interval estimation from surveys with ignorable nonresponse*. Unpublished Ph.D. dissertation, University of Chicago.

Singh, A. C., Mantel, H. J., Kinack, M. D. and Rowe, G. (1993). Statistical matching: use of auxiliary information as an alternative to the conditional independent assumption. *Survey Methodology*, 19(1), 59–79.

Tanner, M. A. and Wong, W. H. (1987). The calculation of posterior distributions by data augmentation (with discussion). *Journal of the American Statistical Association*, 82, 528–550.

Listing of NCES Working Papers to Date

Working papers can be downloaded as pdf files from the NCES Electronic Catalog (http://nces.ed.gov/pubsearch/). You can also contact Sheilah Jupiter at (202) 502–7444 (sheilah_jupiter@ed.gov) if you are interested in any of the following papers.

Listing of NCES Working Papers by Program Area

No.	Title	NCES contact
Baccalaureate and Beyond (B&B)		
98–15	Development of a Prototype System for Accessing Linked NCES Data	Steven Kaufman
2001–15	Baccalaureate and Beyond Longitudinal Study: 2000/01 Follow-Up Field Test Methodology Report	Andrew G. Malizio
Beginning Postsecondary Students (BPS) Longitudinal Study		
98–11	Beginning Postsecondary Students Longitudinal Study First Follow-up (BPS:96–98) Field Test Report	Aurora D'Amico
98–15	Development of a Prototype System for Accessing Linked NCES Data	Steven Kaufman
1999–15	Projected Postsecondary Outcomes of 1992 High School Graduates	Aurora D'Amico
2001–04	Beginning Postsecondary Students Longitudinal Study: 1996–2001 (BPS:1996/2001) Field Test Methodology Report	Paula Knepper
Common Core of Data (CCD)		
95–12	Rural Education Data User's Guide	Samuel Peng
96–19	Assessment and Analysis of School-Level Expenditures	William J. Fowler, Jr.
97–15	Customer Service Survey: Common Core of Data Coordinators	Lee Hoffman
97–43	Measuring Inflation in Public School Costs	William J. Fowler, Jr.
98–15	Development of a Prototype System for Accessing Linked NCES Data	Steven Kaufman
1999–03	Evaluation of the 1996–97 Nonfiscal Common Core of Data Surveys Data Collection, Processing, and Editing Cycle	Beth Young
2000–12	Coverage Evaluation of the 1994–95 Common Core of Data: Public Elementary/Secondary School Universe Survey	Beth Young
2000–13	Non-professional Staff in the Schools and Staffing Survey (SASS) and Common Core of Data (CCD)	Kerry Gruber
2001–09	An Assessment of the Accuracy of CCD Data: A Comparison of 1988, 1989, and 1990 CCD Data with 1990–91 SASS Data	John Sietsema
2001–14	Evaluation of the Common Core of Data (CCD) Finance Data Imputations	Frank Johnson
Data Development		
2000–16a	Lifelong Learning NCES Task Force: Final Report Volume I	Lisa Hudson
2000–16b	Lifelong Learning NCES Task Force: Final Report Volume II	Lisa Hudson
Decennial Census School District Project		
95–12	Rural Education Data User's Guide	Samuel Peng
96–04	Census Mapping Project/School District Data Book	Tai Phan
98–07	Decennial Census School District Project Planning Report	Tai Phan
2001–12	Customer Feedback on the 1990 Census Mapping Project	Dan Kasprzyk
Early Childhood Longitudinal Study (ECLS)		
96–08	How Accurate are Teacher Judgments of Students' Academic Performance?	Jerry West
96–18	Assessment of Social Competence, Adaptive Behaviors, and Approaches to Learning with Young Children	Jerry West
97–24	Formulating a Design for the ECLS: A Review of Longitudinal Studies	Jerry West
97–36	Measuring the Quality of Program Environments in Head Start and Other Early Childhood Programs: A Review and Recommendations for Future Research	Jerry West
1999–01	A Birth Cohort Study: Conceptual and Design Considerations and Rationale	Jerry West

No.	Title	NCES contact
2000–04	Selected Papers on Education Surveys: Papers Presented at the 1998 and 1999 ASA and 1999 AAPOR Meetings	Dan Kasprzyk
2001–02	Measuring Father Involvement in Young Children's Lives: Recommendations for a Fatherhood Module for the ECLS-B	Jerry West
2001–03	Measures of Socio-Emotional Development in Middle Childhood	Elvira Hausken
2001–06	Papers from the Early Childhood Longitudinal Studies Program: Presented at the 2001 AERA and SRCD Meetings	Jerry West
Education Finance Statistics Center (EDFIN)		
94–05	Cost-of-Education Differentials Across the States	William J. Fowler, Jr.
96–19	Assessment and Analysis of School-Level Expenditures	William J. Fowler, Jr.
97–43	Measuring Inflation in Public School Costs	William J. Fowler, Jr.
98–04	Geographic Variations in Public Schools' Costs	William J. Fowler, Jr.
1999–16	Measuring Resources in Education: From Accounting to the Resource Cost Model Approach	William J. Fowler, Jr.
High School and Beyond (HS&B)		
95–12	Rural Education Data User's Guide	Samuel Peng
1999–05	Procedures Guide for Transcript Studies	Dawn Nelson
1999–06	1998 Revision of the Secondary School Taxonomy	Dawn Nelson
HS Transcript Studies		
1999–05	Procedures Guide for Transcript Studies	Dawn Nelson
1999–06	1998 Revision of the Secondary School Taxonomy	Dawn Nelson
International Adult Literacy Survey (IALS)		
97–33	Adult Literacy: An International Perspective	Marilyn Binkley
Integrated Postsecondary Education Data System (IPEDS)		
97–27	Pilot Test of IPEDS Finance Survey	Peter Stowe
98–15	Development of a Prototype System for Accessing Linked NCES Data	Steven Kaufman
2000–14	IPEDS Finance Data Comparisons Under the 1997 Financial Accounting Standards for Private, Not-for-Profit Institutes: A Concept Paper	Peter Stowe
National Assessment of Adult Literacy (NAAL)		
98–17	Developing the National Assessment of Adult Literacy: Recommendations from Stakeholders	Sheida White
1999–09a	1992 National Adult Literacy Survey: An Overview	Alex Sedlacek
1999–09b	1992 National Adult Literacy Survey: Sample Design	Alex Sedlacek
1999–09c	1992 National Adult Literacy Survey: Weighting and Population Estimates	Alex Sedlacek
1999–09d	1992 National Adult Literacy Survey: Development of the Survey Instruments	Alex Sedlacek
1999–09e	1992 National Adult Literacy Survey: Scaling and Proficiency Estimates	Alex Sedlacek
1999–09f	1992 National Adult Literacy Survey: Interpreting the Adult Literacy Scales and Literacy Levels	Alex Sedlacek
1999–09g	1992 National Adult Literacy Survey: Literacy Levels and the Response Probability Convention	Alex Sedlacek
2000–05	Secondary Statistical Modeling With the National Assessment of Adult Literacy: Implications for the Design of the Background Questionnaire	Sheida White
2000–06	Using Telephone and Mail Surveys as a Supplement or Alternative to Door-to-Door Surveys in the Assessment of Adult Literacy	Sheida White
2000–07	"How Much Literacy is Enough?" Issues in Defining and Reporting Performance Standards for the National Assessment of Adult Literacy	Sheida White
2000–08	Evaluation of the 1992 NALS Background Survey Questionnaire: An Analysis of Uses with Recommendations for Revisions	Sheida White
2000–09	Demographic Changes and Literacy Development in a Decade	Sheida White
2001–08	Assessing the Lexile Framework: Results of a Panel Meeting	Sheida White

No.	Title	NCES contact
National Assessment of Educational Progress (NAEP)		
95–12	Rural Education Data User's Guide	Samuel Peng
97–29	Can State Assessment Data be Used to Reduce State NAEP Sample Sizes?	Steven Gorman
97–30	ACT's NAEP Redesign Project: Assessment Design is the Key to Useful and Stable Assessment Results	Steven Gorman
97–31	NAEP Reconfigured: An Integrated Redesign of the National Assessment of Educational Progress	Steven Gorman
97–32	Innovative Solutions to Intractable Large Scale Assessment (Problem 2: Background Questionnaires)	Steven Gorman
97–37	Optimal Rating Procedures and Methodology for NAEP Open-ended Items	Steven Gorman
97–44	Development of a SASS 1993–94 School-Level Student Achievement Subfile: Using State Assessments and State NAEP, Feasibility Study	Michael Ross
98–15	Development of a Prototype System for Accessing Linked NCES Data	Steven Kaufman
1999–05	Procedures Guide for Transcript Studies	Dawn Nelson
1999–06	1998 Revision of the Secondary School Taxonomy	Dawn Nelson
2001–07	A Comparison of the National Assessment of Educational Progress (NAEP), the Third International Mathematics and Science Study Repeat (TIMSS-R), and the Programme for International Student Assessment (PISA)	Arnold Goldstein
2001–08	Assessing the Lexile Framework: Results of a Panel Meeting	Sheida White
2001–11	Impact of Selected Background Variables on Students' NAEP Math Performance	Arnold Goldstein
2001–13	The Effects of Accommodations on the Assessment of LEP Students in NAEP	Arnold Goldstein
National Education Longitudinal Study of 1988 (NELS:88)		
95–04	National Education Longitudinal Study of 1988: Second Follow-up Questionnaire Content Areas and Research Issues	Jeffrey Owings
95–05	National Education Longitudinal Study of 1988: Conducting Trend Analyses of NLS-72, HS&B, and NELS:88 Seniors	Jeffrey Owings
95–06	National Education Longitudinal Study of 1988: Conducting Cross-Cohort Comparisons Using HS&B, NAEP, and NELS:88 Academic Transcript Data	Jeffrey Owings
95–07	National Education Longitudinal Study of 1988: Conducting Trend Analyses HS&B and NELS:88 Sophomore Cohort Dropouts	Jeffrey Owings
95–12	Rural Education Data User's Guide	Samuel Peng
95–14	Empirical Evaluation of Social, Psychological, & Educational Construct Variables Used in NCES Surveys	Samuel Peng
96–03	National Education Longitudinal Study of 1988 (NELS:88) Research Framework and Issues	Jeffrey Owings
98–06	National Education Longitudinal Study of 1988 (NELS:88) Base Year through Second Follow-Up: Final Methodology Report	Ralph Lee
98–09	High School Curriculum Structure: Effects on Coursetaking and Achievement in Mathematics for High School Graduates—An Examination of Data from the National Education Longitudinal Study of 1988	Jeffrey Owings
98–15	Development of a Prototype System for Accessing Linked NCES Data	Steven Kaufman
1999–05	Procedures Guide for Transcript Studies	Dawn Nelson
1999–06	1998 Revision of the Secondary School Taxonomy	Dawn Nelson
1999–15	Projected Postsecondary Outcomes of 1992 High School Graduates	Aurora D'Amico
2001–16	Imputation of Test Scores in the National Education Longitudinal Study of 1988	Ralph Lee
National Household Education Survey (NHES)		
95–12	Rural Education Data User's Guide	Samuel Peng
96–13	Estimation of Response Bias in the NHES:95 Adult Education Survey	Steven Kaufman
96–14	The 1995 National Household Education Survey: Reinterview Results for the Adult Education Component	Steven Kaufman
96–20	1991 National Household Education Survey (NHES:91) Questionnaires: Screener, Early Childhood Education, and Adult Education	Kathryn Chandler
96–21	1993 National Household Education Survey (NHES:93) Questionnaires: Screener, School Readiness, and School Safety and Discipline	Kathryn Chandler
96–22	1995 National Household Education Survey (NHES:95) Questionnaires: Screener, Early Childhood Program Participation, and Adult Education	Kathryn Chandler

No.	Title	NCES contact
96–29	Undercoverage Bias in Estimates of Characteristics of Adults and 0- to 2-Year-Olds in the 1995 National Household Education Survey (NHES:95)	Kathryn Chandler
96–30	Comparison of Estimates from the 1995 National Household Education Survey (NHES:95)	Kathryn Chandler
97–02	Telephone Coverage Bias and Recorded Interviews in the 1993 National Household Education Survey (NHES:93)	Kathryn Chandler
97–03	1991 and 1995 National Household Education Survey Questionnaires: NHES:91 Screener, NHES:91 Adult Education, NHES:95 Basic Screener, and NHES:95 Adult Education	Kathryn Chandler
97–04	Design, Data Collection, Monitoring, Interview Administration Time, and Data Editing in the 1993 National Household Education Survey (NHES:93)	Kathryn Chandler
97–05	Unit and Item Response, Weighting, and Imputation Procedures in the 1993 National Household Education Survey (NHES:93)	Kathryn Chandler
97–06	Unit and Item Response, Weighting, and Imputation Procedures in the 1995 National Household Education Survey (NHES:95)	Kathryn Chandler
97–08	Design, Data Collection, Interview Timing, and Data Editing in the 1995 National Household Education Survey	Kathryn Chandler
97–19	National Household Education Survey of 1995: Adult Education Course Coding Manual	Peter Stowe
97–20	National Household Education Survey of 1995: Adult Education Course Code Merge Files User's Guide	Peter Stowe
97–25	1996 National Household Education Survey (NHES:96) Questionnaires: Screener/Household and Library, Parent and Family Involvement in Education and Civic Involvement, Youth Civic Involvement, and Adult Civic Involvement	Kathryn Chandler
97–28	Comparison of Estimates in the 1996 National Household Education Survey	Kathryn Chandler
97–34	Comparison of Estimates from the 1993 National Household Education Survey	Kathryn Chandler
97–35	Design, Data Collection, Interview Administration Time, and Data Editing in the 1996 National Household Education Survey	Kathryn Chandler
97–38	Reinterview Results for the Parent and Youth Components of the 1996 National Household Education Survey	Kathryn Chandler
97–39	Undercoverage Bias in Estimates of Characteristics of Households and Adults in the 1996 National Household Education Survey	Kathryn Chandler
97–40	Unit and Item Response Rates, Weighting, and Imputation Procedures in the 1996 National Household Education Survey	Kathryn Chandler
98–03	Adult Education in the 1990s: A Report on the 1991 National Household Education Survey	Peter Stowe
98–10	Adult Education Participation Decisions and Barriers: Review of Conceptual Frameworks and Empirical Studies	Peter Stowe

National Longitudinal Study of the High School Class of 1972 (NLS-72)

95–12	Rural Education Data User's Guide	Samuel Peng

National Postsecondary Student Aid Study (NPSAS)

96–17	National Postsecondary Student Aid Study: 1996 Field Test Methodology Report	Andrew G. Malizio
2000–17	National Postsecondary Student Aid Study:2000 Field Test Methodology Report	Andrew G. Malizio

National Study of Postsecondary Faculty (NSOPF)

97–26	Strategies for Improving Accuracy of Postsecondary Faculty Lists	Linda Zimbler
98–15	Development of a Prototype System for Accessing Linked NCES Data	Steven Kaufman
2000–01	1999 National Study of Postsecondary Faculty (NSOPF:99) Field Test Report	Linda Zimbler

Postsecondary Education Descriptive Analysis Reports (PEDAR)

2000–11	Financial Aid Profile of Graduate Students in Science and Engineering	Aurora D'Amico

Private School Universe Survey (PSS)

95–16	Intersurvey Consistency in NCES Private School Surveys	Steven Kaufman
95–17	Estimates of Expenditures for Private K–12 Schools	Stephen Broughman
96–16	Strategies for Collecting Finance Data from Private Schools	Stephen Broughman
96–26	Improving the Coverage of Private Elementary-Secondary Schools	Steven Kaufman
96–27	Intersurvey Consistency in NCES Private School Surveys for 1993–94	Steven Kaufman
97–07	The Determinants of Per-Pupil Expenditures in Private Elementary and Secondary Schools: An Exploratory Analysis	Stephen Broughman

No.	Title	NCES contact
97–22	Collection of Private School Finance Data: Development of a Questionnaire	Stephen Broughman
98–15	Development of a Prototype System for Accessing Linked NCES Data	Steven Kaufman
2000–04	Selected Papers on Education Surveys: Papers Presented at the 1998 and 1999 ASA and 1999 AAPOR Meetings	Dan Kasprzyk
2000–15	Feasibility Report: School-Level Finance Pretest, Private School Questionnaire	Stephen Broughman
Recent College Graduates (RCG)		
98–15	Development of a Prototype System for Accessing Linked NCES Data	Steven Kaufman
Schools and Staffing Survey (SASS)		
94–01	Schools and Staffing Survey (SASS) Papers Presented at Meetings of the American Statistical Association	Dan Kasprzyk
94–02	Generalized Variance Estimate for Schools and Staffing Survey (SASS)	Dan Kasprzyk
94–03	1991 Schools and Staffing Survey (SASS) Reinterview Response Variance Report	Dan Kasprzyk
94–04	The Accuracy of Teachers' Self-reports on their Postsecondary Education: Teacher Transcript Study, Schools and Staffing Survey	Dan Kasprzyk
94–06	Six Papers on Teachers from the 1990–91 Schools and Staffing Survey and Other Related Surveys	Dan Kasprzyk
95–01	Schools and Staffing Survey: 1994 Papers Presented at the 1994 Meeting of the American Statistical Association	Dan Kasprzyk
95–02	QED Estimates of the 1990–91 Schools and Staffing Survey: Deriving and Comparing QED School Estimates with CCD Estimates	Dan Kasprzyk
95–03	Schools and Staffing Survey: 1990–91 SASS Cross-Questionnaire Analysis	Dan Kasprzyk
95–08	CCD Adjustment to the 1990–91 SASS: A Comparison of Estimates	Dan Kasprzyk
95–09	The Results of the 1993 Teacher List Validation Study (TLVS)	Dan Kasprzyk
95–10	The Results of the 1991–92 Teacher Follow-up Survey (TFS) Reinterview and Extensive Reconciliation	Dan Kasprzyk
95–11	Measuring Instruction, Curriculum Content, and Instructional Resources: The Status of Recent Work	Sharon Bobbitt & John Ralph
95–12	Rural Education Data User's Guide	Samuel Peng
95–14	Empirical Evaluation of Social, Psychological, & Educational Construct Variables Used in NCES Surveys	Samuel Peng
95–15	Classroom Instructional Processes: A Review of Existing Measurement Approaches and Their Applicability for the Teacher Follow-up Survey	Sharon Bobbitt
95–16	Intersurvey Consistency in NCES Private School Surveys	Steven Kaufman
95–18	An Agenda for Research on Teachers and Schools: Revisiting NCES' Schools and Staffing Survey	Dan Kasprzyk
96–01	Methodological Issues in the Study of Teachers' Careers: Critical Features of a Truly Longitudinal Study	Dan Kasprzyk
96–02	Schools and Staffing Survey (SASS): 1995 Selected papers presented at the 1995 Meeting of the American Statistical Association	Dan Kasprzyk
96–05	Cognitive Research on the Teacher Listing Form for the Schools and Staffing Survey	Dan Kasprzyk
96–06	The Schools and Staffing Survey (SASS) for 1998–99: Design Recommendations to Inform Broad Education Policy	Dan Kasprzyk
96–07	Should SASS Measure Instructional Processes and Teacher Effectiveness?	Dan Kasprzyk
96–09	Making Data Relevant for Policy Discussions: Redesigning the School Administrator Questionnaire for the 1998–99 SASS	Dan Kasprzyk
96–10	1998–99 Schools and Staffing Survey: Issues Related to Survey Depth	Dan Kasprzyk
96–11	Towards an Organizational Database on America's Schools: A Proposal for the Future of SASS, with comments on School Reform, Governance, and Finance	Dan Kasprzyk
96–12	Predictors of Retention, Transfer, and Attrition of Special and General Education Teachers: Data from the 1989 Teacher Followup Survey	Dan Kasprzyk
96–15	Nested Structures: District-Level Data in the Schools and Staffing Survey	Dan Kasprzyk
96–23	Linking Student Data to SASS: Why, When, How	Dan Kasprzyk
96–24	National Assessments of Teacher Quality	Dan Kasprzyk
96–25	Measures of Inservice Professional Development: Suggested Items for the 1998–1999 Schools and Staffing Survey	Dan Kasprzyk
96–28	Student Learning, Teaching Quality, and Professional Development: Theoretical Linkages, Current Measurement, and Recommendations for Future Data Collection	Mary Rollefson
97–01	Selected Papers on Education Surveys: Papers Presented at the 1996 Meeting of the American Statistical Association	Dan Kasprzyk

No.	Title	NCES contact
97–07	The Determinants of Per-Pupil Expenditures in Private Elementary and Secondary Schools: An Exploratory Analysis	Stephen Broughman
97–09	Status of Data on Crime and Violence in Schools: Final Report	Lee Hoffman
97–10	Report of Cognitive Research on the Public and Private School Teacher Questionnaires for the Schools and Staffing Survey 1993–94 School Year	Dan Kasprzyk
97–11	International Comparisons of Inservice Professional Development	Dan Kasprzyk
97–12	Measuring School Reform: Recommendations for Future SASS Data Collection	Mary Rollefson
97–14	Optimal Choice of Periodicities for the Schools and Staffing Survey: Modeling and Analysis	Steven Kaufman
97–18	Improving the Mail Return Rates of SASS Surveys: A Review of the Literature	Steven Kaufman
97–22	Collection of Private School Finance Data: Development of a Questionnaire	Stephen Broughman
97–23	Further Cognitive Research on the Schools and Staffing Survey (SASS) Teacher Listing Form	Dan Kasprzyk
97–41	Selected Papers on the Schools and Staffing Survey: Papers Presented at the 1997 Meeting of the American Statistical Association	Steve Kaufman
97–42	Improving the Measurement of Staffing Resources at the School Level: The Development of Recommendations for NCES for the Schools and Staffing Survey (SASS)	Mary Rollefson
97–44	Development of a SASS 1993–94 School-Level Student Achievement Subfile: Using State Assessments and State NAEP, Feasibility Study	Michael Ross
98–01	Collection of Public School Expenditure Data: Development of a Questionnaire	Stephen Broughman
98–02	Response Variance in the 1993–94 Schools and Staffing Survey: A Reinterview Report	Steven Kaufman
98–04	Geographic Variations in Public Schools' Costs	William J. Fowler, Jr.
98–05	SASS Documentation: 1993–94 SASS Student Sampling Problems; Solutions for Determining the Numerators for the SASS Private School (3B) Second-Stage Factors	Steven Kaufman
98–08	The Redesign of the Schools and Staffing Survey for 1999–2000: A Position Paper	Dan Kasprzyk
98–12	A Bootstrap Variance Estimator for Systematic PPS Sampling	Steven Kaufman
98–13	Response Variance in the 1994–95 Teacher Follow-up Survey	Steven Kaufman
98–14	Variance Estimation of Imputed Survey Data	Steven Kaufman
98–15	Development of a Prototype System for Accessing Linked NCES Data	Steven Kaufman
98–16	A Feasibility Study of Longitudinal Design for Schools and Staffing Survey	Stephen Broughman
1999–02	Tracking Secondary Use of the Schools and Staffing Survey Data: Preliminary Results	Dan Kasprzyk
1999–04	Measuring Teacher Qualifications	Dan Kasprzyk
1999–07	Collection of Resource and Expenditure Data on the Schools and Staffing Survey	Stephen Broughman
1999–08	Measuring Classroom Instructional Processes: Using Survey and Case Study Fieldtest Results to Improve Item Construction	Dan Kasprzyk
1999–10	What Users Say About Schools and Staffing Survey Publications	Dan Kasprzyk
1999–12	1993–94 Schools and Staffing Survey: Data File User's Manual, Volume III: Public-Use Codebook	Kerry Gruber
1999–13	1993–94 Schools and Staffing Survey: Data File User's Manual, Volume IV: Bureau of Indian Affairs (BIA) Restricted-Use Codebook	Kerry Gruber
1999–14	1994–95 Teacher Followup Survey: Data File User's Manual, Restricted-Use Codebook	Kerry Gruber
1999–17	Secondary Use of the Schools and Staffing Survey Data	Susan Wiley
2000–04	Selected Papers on Education Surveys: Papers Presented at the 1998 and 1999 ASA and 1999 AAPOR Meetings	Dan Kasprzyk
2000–10	A Research Agenda for the 1999–2000 Schools and Staffing Survey	Dan Kasprzyk
2000–13	Non-professional Staff in the Schools and Staffing Survey (SASS) and Common Core of Data (CCD)	Kerry Gruber
2000–18	Feasibility Report: School-Level Finance Pretest, Public School District Questionnaire	Stephen Broughman
Third International Mathematics and Science Study (TIMSS)		
2001–01	Cross-National Variation in Educational Preparation for Adulthood: From Early Adolescence to Young Adulthood	Elvira Hausken
2001–05	Using TIMSS to Analyze Correlates of Performance Variation in Mathematics	Patrick Gonzales
2001–07	A Comparison of the National Assessment of Educational Progress (NAEP), the Third International Mathematics and Science Study Repeat (TIMSS-R), and the Programme for International Student Assessment (PISA)	Arnold Goldstein

Listing of NCES Working Papers by Subject

No.	Title	NCES contact
Achievement (student) - mathematics		
2001–05	Using TIMSS to Analyze Correlates of Performance Variation in Mathematics	Patrick Gonzales
Adult education		
96–14	The 1995 National Household Education Survey: Reinterview Results for the Adult Education Component	Steven Kaufman
96–20	1991 National Household Education Survey (NHES:91) Questionnaires: Screener, Early Childhood Education, and Adult Education	Kathryn Chandler
96–22	1995 National Household Education Survey (NHES:95) Questionnaires: Screener, Early Childhood Program Participation, and Adult Education	Kathryn Chandler
98–03	Adult Education in the 1990s: A Report on the 1991 National Household Education Survey	Peter Stowe
98–10	Adult Education Participation Decisions and Barriers: Review of Conceptual Frameworks and Empirical Studies	Peter Stowe
1999–11	Data Sources on Lifelong Learning Available from the National Center for Education Statistics	Lisa Hudson
2000–16a	Lifelong Learning NCES Task Force: Final Report Volume I	Lisa Hudson
2000–16b	Lifelong Learning NCES Task Force: Final Report Volume II	Lisa Hudson
Adult literacy—see Literacy of adults		
American Indian – education		
1999–13	1993–94 Schools and Staffing Survey: Data File User's Manual, Volume IV: Bureau of Indian Affairs (BIA) Restricted-Use Codebook	Kerry Gruber
Assessment/achievement		
95–12	Rural Education Data User's Guide	Samuel Peng
95–13	Assessing Students with Disabilities and Limited English Proficiency	James Houser
97–29	Can State Assessment Data be Used to Reduce State NAEP Sample Sizes?	Larry Ogle
97–30	ACT's NAEP Redesign Project: Assessment Design is the Key to Useful and Stable Assessment Results	Larry Ogle
97–31	NAEP Reconfigured: An Integrated Redesign of the National Assessment of Educational Progress	Larry Ogle
97–32	Innovative Solutions to Intractable Large Scale Assessment (Problem 2: Background Questions)	Larry Ogle
97–37	Optimal Rating Procedures and Methodology for NAEP Open-ended Items	Larry Ogle
97–44	Development of a SASS 1993–94 School-Level Student Achievement Subfile: Using State Assessments and State NAEP, Feasibility Study	Michael Ross
98–09	High School Curriculum Structure: Effects on Coursetaking and Achievement in Mathematics for High School Graduates—An Examination of Data from the National Education Longitudinal Study of 1988	Jeffrey Owings
2001–07	A Comparison of the National Assessment of Educational Progress (NAEP), the Third International Mathematics and Science Study Repeat (TIMSS-R), and the Programme for International Student Assessment (PISA)	Arnold Goldstein
2001–11	Impact of Selected Background Variables on Students' NAEP Math Performance	Arnold Goldstein
2001–13	The Effects of Accommodations on the Assessment of LEP Students in NAEP	Arnold Goldstein
Beginning students in postsecondary education		
98–11	Beginning Postsecondary Students Longitudinal Study First Follow-up (BPS:96–98) Field Test Report	Aurora D'Amico
2001–04	Beginning Postsecondary Students Longitudinal Study: 1996–2001 (BPS:1996/2001) Field Test Methodology Report	Paula Knepper

No.	Title	NCES contact
Civic participation		
97–25	1996 National Household Education Survey (NHES:96) Questionnaires: Screener/Household and Library, Parent and Family Involvement in Education and Civic Involvement, Youth Civic Involvement, and Adult Civic Involvement	Kathryn Chandler
Climate of schools		
95–14	Empirical Evaluation of Social, Psychological, & Educational Construct Variables Used in NCES Surveys	Samuel Peng
Cost of education indices		
94–05	Cost-of-Education Differentials Across the States	William J. Fowler, Jr.
Course-taking		
95–12	Rural Education Data User's Guide	Samuel Peng
98–09	High School Curriculum Structure: Effects on Coursetaking and Achievement in Mathematics for High School Graduates—An Examination of Data from the National Education Longitudinal Study of 1988	Jeffrey Owings
1999–05	Procedures Guide for Transcript Studies	Dawn Nelson
1999–06	1998 Revision of the Secondary School Taxonomy	Dawn Nelson
Crime		
97–09	Status of Data on Crime and Violence in Schools: Final Report	Lee Hoffman
Curriculum		
95–11	Measuring Instruction, Curriculum Content, and Instructional Resources: The Status of Recent Work	Sharon Bobbitt & John Ralph
98–09	High School Curriculum Structure: Effects on Coursetaking and Achievement in Mathematics for High School Graduates—An Examination of Data from the National Education Longitudinal Study of 1988	Jeffrey Owings
Customer service		
1999–10	What Users Say About Schools and Staffing Survey Publications	Dan Kasprzyk
2000–02	Coordinating NCES Surveys: Options, Issues, Challenges, and Next Steps	Valena Plisko
2000–04	Selected Papers on Education Surveys: Papers Presented at the 1998 and 1999 ASA and 1999 AAPOR Meetings	Dan Kasprzyk
2001–12	Customer Feedback on the 1990 Census Mapping Project	Dan Kasprzyk
Data quality		
97–13	Improving Data Quality in NCES: Database-to-Report Process	Susan Ahmed
2001–11	Impact of Selected Background Variables on Students' NAEP Math Performance	Arnold Goldstein
2001–13	The Effects of Accommodations on the Assessment of LEP Students in NAEP	Arnold Goldstein
Data warehouse		
2000–04	Selected Papers on Education Surveys: Papers Presented at the 1998 and 1999 ASA and 1999 AAPOR Meetings	Dan Kasprzyk
Design effects		
2000–03	Strengths and Limitations of Using SUDAAN, Stata, and WesVarPC for Computing Variances from NCES Data Sets	Ralph Lee
Dropout rates, high school		
95–07	National Education Longitudinal Study of 1988: Conducting Trend Analyses HS&B and NELS:88 Sophomore Cohort Dropouts	Jeffrey Owings
Early childhood education		
96–20	1991 National Household Education Survey (NHES:91) Questionnaires: Screener, Early Childhood Education, and Adult Education	Kathryn Chandler

No.	Title	NCES contact
96–22	1995 National Household Education Survey (NHES:95) Questionnaires: Screener, Early Childhood Program Participation, and Adult Education	Kathryn Chandler
97–24	Formulating a Design for the ECLS: A Review of Longitudinal Studies	Jerry West
97–36	Measuring the Quality of Program Environments in Head Start and Other Early Childhood Programs: A Review and Recommendations for Future Research	Jerry West
1999–01	A Birth Cohort Study: Conceptual and Design Considerations and Rationale	Jerry West
2001–02	Measuring Father Involvement in Young Children's Lives: Recommendations for a Fatherhood Module for the ECLS-B	Jerry West
2001–03	Measures of Socio-Emotional Development in Middle School	Elvira Hausken
2001–06	Papers from the Early Childhood Longitudinal Studies Program: Presented at the 2001 AERA and SRCD Meetings	Jerry West
Educational attainment		
98–11	Beginning Postsecondary Students Longitudinal Study First Follow-up (BPS:96–98) Field Test Report	Aurora D'Amico
2001–15	Baccalaureate and Beyond Longitudinal Study: 2000/01 Follow-Up Field Test Methodology Report	Andrew G. Malizio
Educational research		
2000–02	Coordinating NCES Surveys: Options, Issues, Challenges, and Next Steps	Valena Plisko
Eighth-graders		
2001–05	Using TIMSS to Analyze Correlates of Performance Variation in Mathematics	Patrick Gonzales
Employment		
96–03	National Education Longitudinal Study of 1988 (NELS:88) Research Framework and Issues	Jeffrey Owings
98–11	Beginning Postsecondary Students Longitudinal Study First Follow-up (BPS:96–98) Field Test Report	Aurora D'Amico
2000–16a	Lifelong Learning NCES Task Force: Final Report Volume I	Lisa Hudson
2000–16b	Lifelong Learning NCES Task Force: Final Report Volume II	Lisa Hudson
2001–01	Cross-National Variation in Educational Preparation for Adulthood: From Early Adolescence to Young Adulthood	Elvira Hausken
Employment – after college		
2001–15	Baccalaureate and Beyond Longitudinal Study: 2000/01 Follow-Up Field Test Methodology Report	Andrew G. Malizio
Engineering		
2000–11	Financial Aid Profile of Graduate Students in Science and Engineering	Aurora D'Amico
Enrollment – after college		
2001–15	Baccalaureate and Beyond Longitudinal Study: 2000/01 Follow-Up Field Test Methodology Report	Andrew G. Malizio
Faculty – higher education		
97–26	Strategies for Improving Accuracy of Postsecondary Faculty Lists	Linda Zimbler
2000–01	1999 National Study of Postsecondary Faculty (NSOPF:99) Field Test Report	Linda Zimbler
Fathers – role in education		
2001–02	Measuring Father Involvement in Young Children's Lives: Recommendations for a Fatherhood Module for the ECLS-B	Jerry West
Finance – elementary and secondary schools		
94–05	Cost-of-Education Differentials Across the States	William J. Fowler, Jr.
96–19	Assessment and Analysis of School-Level Expenditures	William J. Fowler, Jr.
98–01	Collection of Public School Expenditure Data: Development of a Questionnaire	Stephen Broughman
1999–07	Collection of Resource and Expenditure Data on the Schools and Staffing Survey	Stephen Broughman

No.	Title	NCES contact
1999–16	Measuring Resources in Education: From Accounting to the Resource Cost Model Approach	William J. Fowler, Jr.
2000–18	Feasibility Report: School-Level Finance Pretest, Public School District Questionnaire	Stephen Broughman
2001–14	Evaluation of the Common Core of Data (CCD) Finance Data Imputations	Frank Johnson
Finance – postsecondary		
97–27	Pilot Test of IPEDS Finance Survey	Peter Stowe
2000–14	IPEDS Finance Data Comparisons Under the 1997 Financial Accounting Standards for Private, Not-for-Profit Institutes: A Concept Paper	Peter Stowe
Finance – private schools		
95–17	Estimates of Expenditures for Private K–12 Schools	Stephen Broughman
96–16	Strategies for Collecting Finance Data from Private Schools	Stephen Broughman
97–07	The Determinants of Per-Pupil Expenditures in Private Elementary and Secondary Schools: An Exploratory Analysis	Stephen Broughman
97–22	Collection of Private School Finance Data: Development of a Questionnaire	Stephen Broughman
1999–07	Collection of Resource and Expenditure Data on the Schools and Staffing Survey	Stephen Broughman
2000–15	Feasibility Report: School-Level Finance Pretest, Private School Questionnaire	Stephen Broughman
Geography		
98–04	Geographic Variations in Public Schools' Costs	William J. Fowler, Jr.
Graduate students		
2000–11	Financial Aid Profile of Graduate Students in Science and Engineering	Aurora D'Amico
Graduates of postsecondary education		
2001–15	Baccalaureate and Beyond Longitudinal Study: 2000/01 Follow-Up Field Test Methodology Report	Andrew G. Malizio
Imputation		
2000–04	Selected Papers on Education Surveys: Papers Presented at the 1998 and 1999 ASA and 1999 AAPOR Meeting	Dan Kasprzyk
2001–10	Comparison of Proc Impute and Schafer's Multiple Imputation Software	Sam Peng
2001–14	Evaluation of the Common Core of Data (CCD) Finance Data Imputations	Frank Johnson
2001–16	Imputation of Test Scores in the National Education Longitudinal Study of 1988	Ralph Lee
2001–17	A Study of Imputation Algorithms	Ralph Lee
Inflation		
97–43	Measuring Inflation in Public School Costs	William J. Fowler, Jr.
Institution data		
2000–01	1999 National Study of Postsecondary Faculty (NSOPF:99) Field Test Report	Linda Zimbler
Instructional resources and practices		
95–11	Measuring Instruction, Curriculum Content, and Instructional Resources: The Status of Recent Work	Sharon Bobbitt & John Ralph
1999–08	Measuring Classroom Instructional Processes: Using Survey and Case Study Field Test Results to Improve Item Construction	Dan Kasprzyk
International comparisons		
97–11	International Comparisons of Inservice Professional Development	Dan Kasprzyk
97–16	International Education Expenditure Comparability Study: Final Report, Volume I	Shelley Burns
97–17	International Education Expenditure Comparability Study: Final Report, Volume II, Quantitative Analysis of Expenditure Comparability	Shelley Burns
2001–01	Cross-National Variation in Educational Preparation for Adulthood: From Early Adolescence to Young Adulthood	Elvira Hausken
2001–07	A Comparison of the National Assessment of Educational Progress (NAEP), the Third International Mathematics and Science Study Repeat (TIMSS-R), and the Programme for International Student Assessment (PISA)	Arnold Goldstein

No.	Title	NCES contact
International comparisons – math and science achievement		
2001–05	Using TIMSS to Analyze Correlates of Performance Variation in Mathematics	Patrick Gonzales
Libraries		
94–07	Data Comparability and Public Policy: New Interest in Public Library Data Papers Presented at Meetings of the American Statistical Association	Carrol Kindel
97–25	1996 National Household Education Survey (NHES:96) Questionnaires: Screener/Household and Library, Parent and Family Involvement in Education and Civic Involvement, Youth Civic Involvement, and Adult Civic Involvement	Kathryn Chandler
Limited English Proficiency		
95–13	Assessing Students with Disabilities and Limited English Proficiency	James Houser
2001–11	Impact of Selected Background Variables on Students' NAEP Math Performance	Arnold Goldstein
2001–13	The Effects of Accommodations on the Assessment of LEP Students in NAEP	Arnold Goldstein
Literacy of adults		
98–17	Developing the National Assessment of Adult Literacy: Recommendations from Stakeholders	Sheida White
1999–09a	1992 National Adult Literacy Survey: An Overview	Alex Sedlacek
1999–09b	1992 National Adult Literacy Survey: Sample Design	Alex Sedlacek
1999–09c	1992 National Adult Literacy Survey: Weighting and Population Estimates	Alex Sedlacek
1999–09d	1992 National Adult Literacy Survey: Development of the Survey Instruments	Alex Sedlacek
1999–09e	1992 National Adult Literacy Survey: Scaling and Proficiency Estimates	Alex Sedlacek
1999–09f	1992 National Adult Literacy Survey: Interpreting the Adult Literacy Scales and Literacy Levels	Alex Sedlacek
1999–09g	1992 National Adult Literacy Survey: Literacy Levels and the Response Probability Convention	Alex Sedlacek
1999–11	Data Sources on Lifelong Learning Available from the National Center for Education Statistics	Lisa Hudson
2000–05	Secondary Statistical Modeling With the National Assessment of Adult Literacy: Implications for the Design of the Background Questionnaire	Sheida White
2000–06	Using Telephone and Mail Surveys as a Supplement or Alternative to Door-to-Door Surveys in the Assessment of Adult Literacy	Sheida White
2000–07	"How Much Literacy is Enough?" Issues in Defining and Reporting Performance Standards for the National Assessment of Adult Literacy	Sheida White
2000–08	Evaluation of the 1992 NALS Background Survey Questionnaire: An Analysis of Uses with Recommendations for Revisions	Sheida White
2000–09	Demographic Changes and Literacy Development in a Decade	Sheida White
2001–08	Assessing the Lexile Framework: Results of a Panel Meeting	Sheida White
Literacy of adults – international		
97–33	Adult Literacy: An International Perspective	Marilyn Binkley
Mathematics		
98–09	High School Curriculum Structure: Effects on Coursetaking and Achievement in Mathematics for High School Graduates—An Examination of Data from the National Education Longitudinal Study of 1988	Jeffrey Owings
1999–08	Measuring Classroom Instructional Processes: Using Survey and Case Study Field Test Results to Improve Item Construction	Dan Kasprzyk
2001–05	Using TIMSS to Analyze Correlates of Performance Variation in Mathematics	Patrick Gonzales
2001–07	A Comparison of the National Assessment of Educational Progress (NAEP), the Third International Mathematics and Science Study Repeat (TIMSS-R), and the Programme for International Student Assessment (PISA)	Arnold Goldstein
2001–11	Impact of Selected Background Variables on Students' NAEP Math Performance	Arnold Goldstein
Parental involvement in education		
96–03	National Education Longitudinal Study of 1988 (NELS:88) Research Framework and Issues	Jeffrey Owings

No.	Title	NCES contact
97–25	1996 National Household Education Survey (NHES:96) Questionnaires: Screener/Household and Library, Parent and Family Involvement in Education and Civic Involvement, Youth Civic Involvement, and Adult Civic Involvement	Kathryn Chandler
1999–01	A Birth Cohort Study: Conceptual and Design Considerations and Rationale	Jerry West
2001–06	Papers from the Early Childhood Longitudinal Studies Program: Presented at the 2001 AERA and SRCD Meetings	Jerry West
Participation rates		
98–10	Adult Education Participation Decisions and Barriers: Review of Conceptual Frameworks and Empirical Studies	Peter Stowe
Postsecondary education		
1999–11	Data Sources on Lifelong Learning Available from the National Center for Education Statistics	Lisa Hudson
2000–16a	Lifelong Learning NCES Task Force: Final Report Volume I	Lisa Hudson
2000–16b	Lifelong Learning NCES Task Force: Final Report Volume II	Lisa Hudson
Postsecondary education – persistence and attainment		
98–11	Beginning Postsecondary Students Longitudinal Study First Follow-up (BPS:96–98) Field Test Report	Aurora D'Amico
1999–15	Projected Postsecondary Outcomes of 1992 High School Graduates	Aurora D'Amico
Postsecondary education – staff		
97–26	Strategies for Improving Accuracy of Postsecondary Faculty Lists	Linda Zimbler
2000–01	1999 National Study of Postsecondary Faculty (NSOPF:99) Field Test Report	Linda Zimbler
Principals		
2000–10	A Research Agenda for the 1999–2000 Schools and Staffing Survey	Dan Kasprzyk
Private schools		
96–16	Strategies for Collecting Finance Data from Private Schools	Stephen Broughman
97–07	The Determinants of Per-Pupil Expenditures in Private Elementary and Secondary Schools: An Exploratory Analysis	Stephen Broughman
97–22	Collection of Private School Finance Data: Development of a Questionnaire	Stephen Broughman
2000–13	Non-professional Staff in the Schools and Staffing Survey (SASS) and Common Core of Data (CCD)	Kerry Gruber
2000–15	Feasibility Report: School-Level Finance Pretest, Private School Questionnaire	Stephen Broughman
Projections of education statistics		
1999–15	Projected Postsecondary Outcomes of 1992 High School Graduates	Aurora D'Amico
Public school finance		
1999–16	Measuring Resources in Education: From Accounting to the Resource Cost Model Approach	William J. Fowler, Jr.
2000–18	Feasibility Report: School-Level Finance Pretest, Public School District Questionnaire	Stephen Broughman
Public schools		
97–43	Measuring Inflation in Public School Costs	William J. Fowler, Jr.
98–01	Collection of Public School Expenditure Data: Development of a Questionnaire	Stephen Broughman
98–04	Geographic Variations in Public Schools' Costs	William J. Fowler, Jr.
1999–02	Tracking Secondary Use of the Schools and Staffing Survey Data: Preliminary Results	Dan Kasprzyk
2000–12	Coverage Evaluation of the 1994–95 Public Elementary/Secondary School Universe Survey	Beth Young
2000–13	Non-professional Staff in the Schools and Staffing Survey (SASS) and Common Core of Data (CCD)	Kerry Gruber

No.	Title	NCES contact
Public schools – secondary		
98–09	High School Curriculum Structure: Effects on Coursetaking and Achievement in Mathematics for High School Graduates—An Examination of Data from the National Education Longitudinal Study of 1988	Jeffrey Owings
Reform, educational		
96–03	National Education Longitudinal Study of 1988 (NELS:88) Research Framework and Issues	Jeffrey Owings
Response rates		
98–02	Response Variance in the 1993–94 Schools and Staffing Survey: A Reinterview Report	Steven Kaufman
School districts		
2000–10	A Research Agenda for the 1999–2000 Schools and Staffing Survey	Dan Kasprzyk
School districts, public		
98–07	Decennial Census School District Project Planning Report	Tai Phan
1999–03	Evaluation of the 1996–97 Nonfiscal Common Core of Data Surveys Data Collection, Processing, and Editing Cycle	Beth Young
School districts, public – demographics of		
96–04	Census Mapping Project/School District Data Book	Tai Phan
Schools		
97–42	Improving the Measurement of Staffing Resources at the School Level: The Development of Recommendations for NCES for the Schools and Staffing Survey (SASS)	Mary Rollefson
98–08	The Redesign of the Schools and Staffing Survey for 1999–2000: A Position Paper	Dan Kasprzyk
1999–03	Evaluation of the 1996–97 Nonfiscal Common Core of Data Surveys Data Collection, Processing, and Editing Cycle	Beth Young
2000–10	A Research Agenda for the 1999–2000 Schools and Staffing Survey	Dan Kasprzyk
Schools – safety and discipline		
97–09	Status of Data on Crime and Violence in Schools: Final Report	Lee Hoffman
Science		
2000–11	Financial Aid Profile of Graduate Students in Science and Engineering	Aurora D'Amico
2001–07	A Comparison of the National Assessment of Educational Progress (NAEP), the Third International Mathematics and Science Study Repeat (TIMSS-R), and the Programme for International Student Assessment (PISA)	Arnold Goldstein
Software evaluation		
2000–03	Strengths and Limitations of Using SUDAAN, Stata, and WesVarPC for Computing Variances from NCES Data Sets	Ralph Lee
Staff		
97–42	Improving the Measurement of Staffing Resources at the School Level: The Development of Recommendations for NCES for the Schools and Staffing Survey (SASS)	Mary Rollefson
98–08	The Redesign of the Schools and Staffing Survey for 1999–2000: A Position Paper	Dan Kasprzyk
Staff – higher education institutions		
97–26	Strategies for Improving Accuracy of Postsecondary Faculty Lists	Linda Zimbler
Staff – nonprofessional		
2000–13	Non-professional Staff in the Schools and Staffing Survey (SASS) and Common Core of Data (CCD)	Kerry Gruber

No.	Title	NCES contact
State		
1999–03	Evaluation of the 1996–97 Nonfiscal Common Core of Data Surveys Data Collection, Processing, and Editing Cycle	Beth Young
Statistical methodology		
97–21	Statistics for Policymakers or Everything You Wanted to Know About Statistics But Thought You Could Never Understand	Susan Ahmed
Statistical standards and methodology		
2001–05	Using TIMSS to Analyze Correlates of Performance Variation in Mathematics	Patrick Gonzales
Students with disabilities		
95–13	Assessing Students with Disabilities and Limited English Proficiency	James Houser
2001–13	The Effects of Accommodations on the Assessment of LEP Students in NAEP	Arnold Goldstein
Survey methodology		
96–17	National Postsecondary Student Aid Study: 1996 Field Test Methodology Report	Andrew G. Malizio
97–15	Customer Service Survey: Common Core of Data Coordinators	Lee Hoffman
97–35	Design, Data Collection, Interview Administration Time, and Data Editing in the 1996 National Household Education Survey	Kathryn Chandler
98–06	National Education Longitudinal Study of 1988 (NELS:88) Base Year through Second Follow-Up: Final Methodology Report	Ralph Lee
98–11	Beginning Postsecondary Students Longitudinal Study First Follow-up (BPS:96–98) Field Test Report	Aurora D'Amico
98–16	A Feasibility Study of Longitudinal Design for Schools and Staffing Survey	Stephen Broughman
1999–07	Collection of Resource and Expenditure Data on the Schools and Staffing Survey	Stephen Broughman
1999–17	Secondary Use of the Schools and Staffing Survey Data	Susan Wiley
2000–01	1999 National Study of Postsecondary Faculty (NSOPF:99) Field Test Report	Linda Zimbler
2000–02	Coordinating NCES Surveys: Options, Issues, Challenges, and Next Steps	Valena Plisko
2000–04	Selected Papers on Education Surveys: Papers Presented at the 1998 and 1999 ASA and 1999 AAPOR Meetings	Dan Kasprzyk
2000–12	Coverage Evaluation of the 1994–95 Public Elementary/Secondary School Universe Survey	Beth Young
2000–17	National Postsecondary Student Aid Study:2000 Field Test Methodology Report	Andrew G. Malizio
2001–04	Beginning Postsecondary Students Longitudinal Study: 1996–2001 (BPS:1996/2001) Field Test Methodology Report	Paula Knepper
2001–07	A Comparison of the National Assessment of Educational Progress (NAEP), the Third International Mathematics and Science Study Repeat (TIMSS-R), and the Programme for International Student Assessment (PISA)	Arnold Goldstein
2001–09	An Assessment of the Accuracy of CCD Data: A Comparison of 1988, 1989, and 1990 CCD Data with 1990–91 SASS Data	John Sietsema
2001–11	Impact of Selected Background Variables on Students' NAEP Math Performance	Arnold Goldstein
2001–13	The Effects of Accommodations on the Assessment of LEP Students in NAEP	Arnold Goldstein
Teachers		
98–13	Response Variance in the 1994–95 Teacher Follow-up Survey	Steven Kaufman
1999–14	1994–95 Teacher Followup Survey: Data File User's Manual, Restricted-Use Codebook	Kerry Gruber
2000–10	A Research Agenda for the 1999–2000 Schools and Staffing Survey	Dan Kasprzyk
Teachers – instructional practices of		
98–08	The Redesign of the Schools and Staffing Survey for 1999–2000: A Position Paper	Dan Kasprzyk
Teachers – opinions regarding safety		
98–08	The Redesign of the Schools and Staffing Survey for 1999–2000: A Position Paper	Dan Kasprzyk
Teachers – performance evaluations		
1999–04	Measuring Teacher Qualifications	Dan Kasprzyk

No.	Title	NCES contact
Teachers – qualifications of		
1999–04	Measuring Teacher Qualifications	Dan Kasprzyk
Teachers – salaries of		
94–05	Cost-of-Education Differentials Across the States	William J. Fowler, Jr.
Training		
2000–16a	Lifelong Learning NCES Task Force: Final Report Volume I	Lisa Hudson
2000–16b	Lifelong Learning NCES Task Force: Final Report Volume II	Lisa Hudson
Variance estimation		
2000–03	Strengths and Limitations of Using SUDAAN, Stata, and WesVarPC for Computing Variances from NCES Data Sets	Ralph Lee
2000–04	Selected Papers on Education Surveys: Papers Presented at the 1998 and 1999 ASA and 1999 AAPOR Meetings	Dan Kasprzyk
Violence		
97–09	Status of Data on Crime and Violence in Schools: Final Report	Lee Hoffman
Vocational education		
95–12	Rural Education Data User's Guide	Samuel Peng
1999–05	Procedures Guide for Transcript Studies	Dawn Nelson
1999–06	1998 Revision of the Secondary School Taxonomy	Dawn Nelson