VARIANCE ESTIMATION WITH IMPUTED MEANS

Joseph L. Schafer and Nathaniel Schenker *

Joseph L. Schafer, Department of Statistics, 1 Oxford Street, Cambridge, MA 02138

Key Words: Missing data, sample surveys.

1 INTRODUCTION

In many large datasets, especially those arising from sample surveys, item nonresponse is often handled by imputation—*i.e.*, missing data are filled in with plausible values, and the dataset is analyzed as if it were complete. Imputation is attractive because it allows standard complete-data methods of analysis to be used on the imputed dataset. The nominal inferences (confidence levels, P-values, etc.) that result from such an analysis may be seriously misleading, however, because uncertainty due to missing data has not been taken into account (see, for example, chapter 3 of Little and Rubin, 1987).

Multiple imputation (Rubin, 1987) provides a general framework for incorporating missing-data uncertainty into inference. Generating proper multiple imputations is often a difficult task, however, as general-purpose algorithms are not widely available (see, *e.g.*, Rubin and Schafer, 1990). Moreover, managing even a small number of imputations may be computationally burdensome in statistical computing environments where the handling of missing data is usually, at best, an afterthought. In large survey applications, the desire of practitioners to produce quality variance estimates may be outweighed by the practical difficulties of generating and managing a multiply-imputed dataset.

In this paper, we develop a "quick and dirty" analytic method that can be used to correct variance estimates to account for missing data in special cases. Our method (1) can be used when there is a single variable subject to nonresponse, and when the complete-data estimator is a smooth function of linear statistics; (2) is similar in spirit to multiple imputation, but it requires only single imputation of predictive means; (3) is based on asymptotic expansions of classical survey estimators and their variance estimates, and can be thought of as a first-order approximation to what would be obtained from an infinite number of multiple imputations; and (4) is computationally attractive when missing data can be modeled with a single-parameter error distribution, *e.g.*, Bernoulli or Poisson.

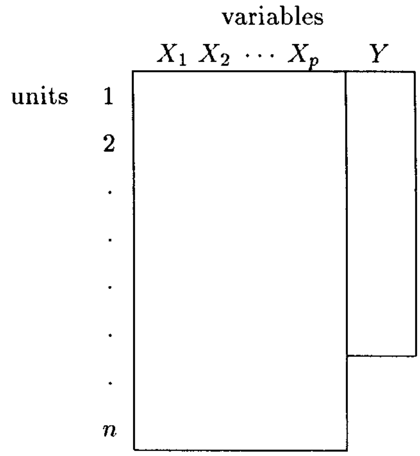

Figure 1: Rectangular dataset, one variable subject to nonresponse.

We detail our assumptions in Section 2 and give a precise definition of mean imputation based on a generic parametric model for the missing data. Our basic results are presented in Section 3. In Section 4, we demonstrate the utility of our approach with an application to the Census Bureau's Post-Enumeration Survey. This paper represents a clarification and extension of earlier work by Schenker (1989), who considered the imputation of probabilities for a single Bernoulli variable subject to nonresponse. The relationship of this method to the results of Schenker (1989) is discussed in Section 3.

2 SETUP AND ASSUMPTIONS

2.1 The Missing-Data Pattern

We will consider rectangular datasets with the simple structure shown in Figure 1. Let Y denote a variable subject to nonresponse, and X denote other variables $(X_1, X_2, \ldots, X_p)$ that are assumed to be completely observed. In this paper, we deal only with the case of a real-valued Y, although extensions of this method to multidimensional Y are possible. Partition Y as $Y = (Y_{obs}, Y_{mis})$, where Y_{obs} denotes the observed values and Y_{mis} denotes

*Joseph L. Schafer, Department of Statistics, Harvard University. Nathaniel Schenker, Department of Biostatistics, UCLA School of Public Health. This research was supported in part by U.S. Bureau of the Census Joint Statistical Agreements 89-08, 89-17, and 90-23, and by National Institute of Mental Health Grant MH-37188. This research was partially conducted while Joseph L. Schafer was employed by the Statistical Support Division, U.S. Bureau of the Census. The views expressed are attributable to the authors and do not necessarily reflect those of the Bureau.

the missing values of Y. Let n denote the number of observational units (rows) in the dataset.

2.2 The Complete-Data Estimation Problem

Let Q denote a scalar quantity to be estimated. We desire both an efficient point estimate for Q and a standard error that includes missing-data uncertainty. Let $\hat{Q}$ denote the complete-data point estimate for Q, *i.e.*, the estimate that we would use if no data were missing. $\hat{Q}$ is a function of both the observed and missing data,

$$\hat{Q} = \hat{Q}(X, Y_{obs}, Y_{mis}). \tag{1}$$

Let U denote the complete-data variance estimate for $\hat{Q}$, *i.e.*, the estimate of $V(Q - \hat{Q})$ that we would use if no data were missing. U is also a function of both the observed and missing data,

$$U = U(X, Y_{obs}, Y_{mis}). \tag{2}$$

We will only consider estimators $\hat{Q}$ that are smooth functions of linear statistics. If w_i denotes a weight (*e.g.*, a survey weight) associated with row i of the dataset, we require that $\hat{Q}$ be a function of weighted (by w_i) sums of the columns of the dataset,

$$\begin{aligned} \hat{Q} &= g(T_{X_1}, \ldots, T_{X_p}, T_Y) \qquad (3)\\ &= g\left(\sum_{i=1}^{n} w_i X_{i1}, \ldots, \sum_{i=1}^{n} w_i X_{ip}, \sum_{i=1}^{n} w_i Y_i\right) \end{aligned}$$

where X_{ij} denotes the ith element of the jth column of the dataset, Y_i denotes the ith element (observed or missing) of the last column, and the function g is smooth. Typically, the estimand Q will be the same function g of the expectations of the linear statistics,

$$Q = g(ET_{X_1}, \ldots, ET_{X_p}, ET_Y), \tag{4}$$

where the expectations are taken over repeated sampling of X and Y; hence $\hat{Q}$ can be thought of as a method of moments estimate of Q. The form (3) includes many estimators typically used in survey practice, including means and proportions, subdomain means, ratios of means, etc., but does not include medians, variances, or correlations.

To stabilize the arguments of g in (3), we require that $\max_i w_i = O(n^{-1})$. In surveys, this can be achieved by scaling the usual survey weights (the inverse probabilities of selection) to sum to one. The weights are allowed to be functions of the observed data (X, Y_{obs}). For example, we allow $\hat{Q}$ to be a poststratified estimator with poststrata defined by categories of X. The weights may not, however, be functions of the missing data Y_{mis}.

We assume that U has the form

$$U = \left(\frac{\partial g(T)}{\partial T}\right)^T W \left(\frac{\partial g(T)}{\partial T}\right), \tag{5}$$

where

$$T = (T_{X_1}, \ldots, T_{X_p}, T_Y)^T, \tag{6}$$

and where W is the classical unbiased variance estimate for T in a stratified pps cluster sample (see, *e.g.*, Wolter, 1985). For example, the diagonal element of W corresponding to the variance of T_{X_k} has the form

$$\sum_s \frac{1}{n_s(n_s - 1)} \sum_c \left(n_s \sum_i w_i X_{ik} - \sum_{c,i} w_i X_{ik}\right)^2$$

where s indexes sampling strata, c indexes clusters within strata, and i indexes sample units within clusters. This is a very common variance estimate in survey practice; it includes, as special cases, variance estimates for simple random samples, stratified and cluster samples, unequal probability samples (such as pps designs), and many multistage designs as well. The derivatives in (5) account for potential nonlinearity of g, a method known in survey literature as Taylor linearization (*e.g.*, Wolter, 1985).

2.3 Regularity of the Complete-Data Problem

We must impose some regularity conditions on the complete-data problem. Assume an asymptotic sequence in which

$$U^{-1/2}(Q - \hat{Q}) \xrightarrow{\mathcal{L}} N(0, 1) \tag{7}$$

as $n \longrightarrow \infty$. When the units in our dataset constitute a simple random sample from an infinite population, (7) is easily verified by appealing to standard central limit theorem arguments. Results such as (7) have also been demonstrated for finite populations and more complex sample designs under a variety of asymptotic conditions (Wolter, 1985). Even if such a result has not been formally demonstrated for a particular estimator or sample design, survey practitioners have still found that appealing to asymptotic normality often provides a useful first-order approximation for statistical inference.

2.4 The Missing-Data Model

In the situation of Figure 1, it is clear that X provides potentially useful information for predicting the missing values of Y. For example, if Y is continuous, we might fit a normal regression model to the cases for which Y is observed, and use the fitted model to predicted Y_{mis}. This approach implicitly assumes that the conditional distribution of Y given X when Y is missing is the same as it is when Y is observed; this is appropriate if the nonresponse mechanism is *ignorable*, in the sense defined by Rubin (1976). Virtually all of the procedures commonly used to handle missing data in surveys and elsewhere in statistical practice are based on an assumption of ignorability. The observed data, of course, give us no information to support or contradict this assumption; such support must come from a source external to the observed data. Other approaches are possible, but every missing-data procedure must be based on some assumption that cannot be verified from (X, Y_{obs}) alone.

We assume that one can correctly specify a probability model for the missing data Y_{mis} given (X, Y_{obs}). A typical specification for this model will include unknown but estimable parameters, which we call θ, as well as some further assumptions, which we call M, that are completely untestable from (X, Y_{obs}) (*e.g.*, the assumption of ignorability). All of our inferences must assume that M is correct. We shall suppress M in the notation from this point onward, with the understanding that M is being conditioned on implicitly throughout.

Let $\hat{\theta}$ denote an efficient estimate of θ based on the observed data (X, Y_{obs}) under the assumed missing-data model. Also, let Γ denote an estimate of $V(\theta - \hat{\theta})$, also based on (X, Y_{obs}). For example, $\hat{\theta}$ may be a maximum likelihood (ML) estimate, and Γ may be the inverse of the observed or expected information matrix evaluated at $\hat{\theta}$. We will assume that $\Gamma = O(n^{-1})$ and that

$$\Gamma^{-1/2}(\theta - \hat{\theta}) \xrightarrow{\mathcal{L}} N(0, I). \qquad (8)$$

This implicitly assumes that the missing-data model is sufficiently regular that standard ML asymptotic theory (see, *e.g.*, Cox and Hinkley 1974) applies, that the fraction of missing information is bounded away from one, and that the dimension of θ is fixed. We will assume further that the missing-data model imposes an uncorrelated (given θ) error structure on the missing-data values. More precisely, let mis denote the set of indices i such that Y_i is an element of Y_{mis}. Assume that

$$\begin{aligned} E(Y_i \mid X, Y_{obs}, \theta) &= \mu_i(\theta) \\ V(Y_i \mid X, Y_{obs}, \theta) &= \sigma_i^2(\theta) \\ Cov(Y_i, Y_{i'} \mid X, Y_{obs}, \theta) &= 0, \quad i \neq i' \end{aligned}$$

where μ_i and σ_i^2 are smooth functions of θ for all $i \in mis$.

These conditions on the missing-data model are not, in practice, overly restrictive. They are satisfied by normal linear regression and analysis of variance models, logistic regression, loglinear and other generalized linear models (GLIM's) as defined by McCullagh and Nelder (1989)—most of the commonly used statistical models that are appropriate for predicting a univariate Y from multivariate X in a rectangular dataset.

2.5 Definition of Mean Imputation

The missing-data model, once it has been specified, is of great value in imputing Y_{mis}. Let $\mu(\theta)$ denote the vector with elements $\mu_i(\theta)$, $i \in mis$; that is,

$$\mu(\theta) = E(Y_{mis} \mid X, Y_{obs}, \theta).$$

If we were going to fill in the missing data with one set of "best" values, we might choose $\mu(\hat{\theta})$, the ML estimate of the mean of Y_{mis}. We will refer to this technique as mean imputation, because it replaces the missing data values with their predicted means under the missing-data model.[1]

Mean imputation be can efficient for point estimation of Q; in fact, we demonstrate below that $\hat{Q}(X, Y_{obs}, \mu(\hat{\theta}))$ is a first-order approximation to the "best" estimate of Q. Mean imputation, however, can seriously distort inferences if the mean-imputed dataset is treated as a complete dataset in the computation of variance estimates. The complete-data variance estimate calculated from a mean-imputed dataset, $U(X, Y_{obs}, \mu(\hat{\theta}))$, is typically a downwardly biased estimate of the true variance, *i.e.*,

$$EU(X, Y_{obs}, \mu(\hat{\theta})) < V(Q - \hat{Q}(X, Y_{obs}, \mu(\hat{\theta}))).$$

Examples of how mean imputation results in biased variance estimates can be found in Section 3.4 of Little and Rubin (1987). For this reason, mean imputation is rarely used in survey practice. Survey practitioners typically use random imputation methods, such as the hot deck, which produce variance estimates that are less biased. Any method

[1] Little and Rubin (1987) call this approach "imputing conditional means" to distinguish it from the method of imputing the unconditional or marginal mean (simply the average of the values in Y_{obs}) for every element of Y_{mis}. In multivariate settings, imputing marginal means may introduce serious biases and can almost never be recommended.

that substitutes a single number for each missing value and then treats the dataset as if it were complete, however, will tend to distort inferences.

Multiple imputation (Rubin, 1987) addresses the shortcomings of single imputation, while still retaining the convenience of imputation as a missing-data procedure. In multiple imputation, the missing data Y_{mis} is replaced by m random draws from their predictive distribution. Using Bayesian notation, we can write this distribution as

$$P(Y_{mis}\,|\,Y_{obs}) = \int P(Y_{mis}\,|\,Y_{obs},\theta)\,P(\theta\,|\,Y_{obs})\,d\theta, \quad (9)$$

which makes explicit the fact that the multiple imputations incorporate uncertainty about the parameter θ as well as uncertainty about the missing data Y_{mis} given θ. Generating proper multiple imputations from a distribution such as (9) may be a complicated task, however, for which no algorithm is readily available. It would be desirable, if possible, to have a "quick and dirty" method of measuring missing-data uncertainty without resorting to full-blown multiple imputation. We provide such a method, as we now describe.

3 VARIANCE ESTIMATION FROM A MEAN-IMPUTED DATASET

3.1 A Bayesian Interpretation

The usual frequentist interpretation of (7) regards Q as fixed and $\hat{Q}$ and U as random. A Bayesian interpretion, however, regards $\hat{Q}$ and U as fixed (given complete data) and Q as random. Exploiting the latter interpretation, we will now regard $\hat{Q}$ and U as the approximate complete-data posterior mean and variance of Q, respectively,

$$\begin{aligned} \hat{Q} &= E(Q \mid X, Y_{obs}, Y_{mis}), \\ U &= V(Q \mid X, Y_{obs}, Y_{mis}). \end{aligned} \quad (10)$$

Under sufficient regularity, posterior means and variances behave as in (7) (*e.g.*, Cox and Hinkley 1974), and in large samples the difference between classical estimates and Bayesian estimates will be small. We will also exploit the Bayesian interpretation of (8) and regard $\hat{\theta}$ and Γ as posterior moments of θ given the observed data,

$$\begin{aligned} \hat{\theta} &= E(\theta \mid X, Y_{obs}), \\ \Gamma &= V(\theta \mid X, Y_{obs}). \end{aligned} \quad (11)$$

With complete data, our state of knowledge about Q is summarized by $\hat{Q}$ and U. With incomplete data, however, inference should be based on the posterior moments given only the data actually observed, $E(Q \mid X, Y_{obs})$ and $V(Q \mid X, Y_{obs})$. Note that

$$E(Q \mid X, Y_{obs}) = E(\hat{Q} \mid X, Y_{obs})$$

and

$$V(Q \mid X, Y_{obs}) = V(\hat{Q}\,|\,X, Y_{obs}) + E(U\,|\,X, Y_{obs})$$

where the moments on the right-hand side of these equations are being evaluated over the posterior distribution $P(Y_{mis} \mid X, Y_{obs})$, the same distribution as (9) from which multiple imputations would be drawn. To obtain approximate posterior moments of Q, then, we need only to approximate the mean and variance of $\hat{Q}$ and the mean of U over the posterior distribution of Y_{mis}. We now state, without proof, these approximations.

3.2 Approximations to the Moments of $\hat{Q}$ and U

It can be shown that, under the assumptions outlined in Section 2,

$$E(\hat{Q} \mid X, Y_{obs}) = \hat{Q}(X, Y_{obs}, \mu(\hat{\theta})) + R_1 \quad (12)$$

where R_1 is the mean of a random variable that is $O_p(n^{-1})$. It can also be shown that

$$\begin{aligned} V(\hat{Q} \mid X, Y_{obs}) &= \left(\frac{\partial g(\hat{T})}{\partial T_Y}\right)^2 \sum_{i\in mis} w_i^2\,\sigma_i^2(\hat{\theta}) \\ &+ \left(\frac{\partial g(\hat{T})}{\partial T_Y}\right)^2 D_\mu(\hat{\theta})^T \Gamma D_\mu(\hat{\theta}) + R_2 \end{aligned} \quad (13)$$

where $\hat{T}$ is shorthand for the complete-data statistic T calculated with $\mu(\hat{\theta})$ substituted for Y_{mis}, where

$$D_\mu(\theta) = \sum_{i\in mis} w_i \left(\frac{\partial \mu_i(\theta)}{\partial \theta}\right), \quad (14)$$

and where R_2 is the mean of a random variable that is $O_p(n^{-3/2})$. Finally, it can also be shown that

$$\begin{aligned} E(U \mid X, Y_{obs}) &= U(X, Y_{obs}, \mu(\hat{\theta})) \\ &+ \left(\frac{\partial g(\hat{T})}{\partial T_Y}\right)^2 \sum_{i\in mis} w_i^2\,\sigma_i^2(\hat{\theta}) + R_3 \end{aligned} \quad (15)$$

where R_3 is the mean of a random variable that is $O_p(n^{-2})$. Proofs of these results follow from first-order Taylor expansions of the functions g and $\mu_i, i \in mis$.

3.3 Point Estimation with Imputed Means

It follows from (12) that the complete-data point estimate with means imputed for the missing values is a first-order approximation to the posterior mean,

$$E(Q \mid X, Y_{obs}) \approx \hat{Q}(X, Y_{obs}, \mu(\hat{\theta})). \qquad (16)$$

In large samples, then, it is desirable to use $\hat{Q}(X, Y_{obs}, \mu(\hat{\theta}))$, as it is an efficient estimate of Q. It is important to note, however, that this result assumes that the complete-data point estimate is a smooth function of linear statistics. It does not hold for an arbitrary estimator $\hat{Q}$; for example, it does not hold for a sample variance. In fact, the result (15) points out that a mean-imputed sample variance is biased downward.

An earlier version of this work (Schenker 1989) considered the special case of a Bernoulli Y. In simple random samples, the variance estimate typically used for a Bernoulli Y is $\bar{Y}(1-\bar{Y})/n$, where $\bar{Y}$ is the sample mean. In this special case the sample variance is a smooth function of a linear statistic, and the result (12) does apply. In more general cases, however, a sample variance does not have this special form, and a correction such as the one in (15) is needed.

3.4 Corrections to the Mean-Imputed Variance Estimate

It follows from (13) and (15) that a first-order approximation to the posterior variance is

$$V(Q \mid X, Y_{obs}) \approx U(X, Y_{obs}, \mu(\hat{\theta})) + C_1 + C_2, \quad (17)$$

where

$$C_1 = 2\left(\frac{\partial g(\hat{T})}{\partial T_Y}\right)^2 \sum_{i \in mis} w_i^2 \, \sigma_i^2(\hat{\theta}) \qquad (18)$$

and

$$C_2 = \left(\frac{\partial g(\hat{T})}{\partial T_Y}\right)^2 D_\mu(\hat{\theta})^T \Gamma D_\mu(\hat{\theta}). \qquad (19)$$

In (17), $U(X, Y_{obs}, \mu(\hat{\theta}))$ is the "naive" estimate that treats the mean-imputed dataset as complete data. The first correction term, C_1, is a component of variance that accounts for uncertainty in Y_{mis} given the imputed means. The second correction term, C_2, is an additional component of variance that accounts for uncertainty in the imputed means, *i.e.*, uncertainty due to the estimation of the parameters in the missing-data model.

The term C_1 is usually very simple to compute. For example, if the estimand Q is the population mean of Y and the missing values in Y are modeled by ordinary linear regression, then C_1 has the form

$$C_1 = 2 \sum_{i \in mis} w_i^2 \hat{\sigma}^2, \qquad (20)$$

where $\hat{\sigma}^2$ is the estimated residual variance of the regression. If Y is a binary variable, and the missing values of Y are modeled as Bernoulli with means $\hat{\pi}_i$, $i \in mis$ (for example, by logistic regression), then C_1 has the form

$$C_1 = 2 \sum_{i \in mis} w_i^2 \hat{\pi}_i (1 - \hat{\pi}_i) \qquad (21)$$

for estimating the mean of Y.

When the elements of Y_{mis} are modeled with an error distribution that has a single parameter (*e.g.*, Bernoulli or Poisson), then the variances $\sigma_i^2(\theta)$ can be expressed as functions of the means $\mu_i(\theta)$, and the term C_1 can be computed from the mean-imputed dataset alone; no additional information is needed. When the error distribution has additional parameters, estimates of these parameters need to be retained to calculate C_1. For example, in the case of ordinary linear regression, the estimated residual variance $\hat{\sigma}^2$ is needed.

The second correction term, C_2, is usually more difficult to compute because it involves the variance of the parameters θ of the missing-data model. When the fraction of missing information is moderate, however, C_2 accounts for only a small proportion of the total variance in (17) and can usually be ignored (Rubin and Schenker, 1986).

4 APPLICATION TO THE PES

The U.S. Census Bureau's Post-Enumeration Survey (PES) attempts to measure errors of coverage (undercounting and overcounting) in the 1990 census. The PES estimates population size by dual-system estimation, a technique analagous to capture-recapture estimation wildlife studies. A full description of the 1990 PES methodology is not yet available, but an overview of a similar survey is given by Diffendal (1988).

Two overlapping samples of the population, denoted by the letters P and E, provide estimates of the gross undercount and gross overcount, respectively. In the P-sample, the outcome of interest, Y^P, is a binary variable indicating whether a person was counted ($Y^P = 1$) or missed ($Y^P = 0$) in the census. In the E-sample, the outcome of interest, Y^E, is a binary variable indicating whether a census person was erroneously included ($Y^E = 1$) or correctly included ($Y^E = 0$) in the census. The dual-system estimate of the population size, *DSE*, can be written as

$$DSE = (CEN - I)(1 - \bar{Y}^E)(\bar{Y}^P)^{-1}, \qquad (22)$$

where

$CEN =$ total census count,

$I =$ number of non-data-defined or imputed persons in the census count,

$\bar{Y}^P =$ P-sample weighted average of Y^P, and

$\bar{Y}^E =$ E-sample weighted average of Y^E.

From the standpoint of the PES, the quantities *CEN* and I are regarded as fixed. Hence, the estimate *DSE* has the form (3) required by the methods of this paper.

Both of the binary outcome variables Y^P and Y^E are subject to nonresponse. The missing data are modeled by means of hierarchical logistic regression models fit to the P and E-samples, and estimated probabilities of $Y^P = 1$ and $Y^E = 1$ from these models are imputed to the dataset. Details of the logistic regression models and fitting procedure are given by Belin *et.al.*, (1991).

Because the missing data are modeled as Bernoulli random variables, the first variance correction term, C_1, has a form similar to (21), with the addition of derivatives to account for the nonlinearity of *DSE*. The second correction term, C_2, is more problematic because a variance estimate Γ for the parameters of the missing-data model is not readily available. Because the fraction of missing data in the PES is not large, it would probably not be unreasonable to ignore C_2 in this case. As a safeguard, however, we obtain a rough estimate of the component C_2 by means of bootstrap resampling. Details of our calculations of C_1 and C_2 for the PES will be provided in a future article.

5 REFERENCES

Belin, T. R., Diffendal, G. J., Rubin, D. B., Schafer, J. L., and Zaslavsky, A. M., "Strategy for Handling Cases with Unresolved After-Followup Enumeration Status in Production of Adjusted Census Counts from 1990 Census and Post-Enumeration Survey," in preparation.

Cox, D. R. and Hinkley, D. V. (1974), *Theoretical Statistics*, London: Chapman and Hall.

Diffendal, G. (1988), "The 1986 Test of Adjustment-Related Operations in Central Los Angeles County," *Survey Methodology*, 14, 71–86.

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

McCullagh , P. and Nelder, J. A. (1983), *Generalized Linear Models*, London: Chapman and Hall.

Rubin, D. B. (1976), "Inference and Missing Data," *Biometrika*, 63, 581–592.

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

Rubin, D. B., and Schafer, J. L. (1990), "Efficiently Creating Multiple Imputations for Incomplete Multivariate Normal Data," *Proceedings of the American Statistical Association Statistical Computing Section*, to appear.

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.

Schenker, N. (1989), "The Use of Imputed Probabilities for Missing Binary Data," *Proceedings of the Fifth Annual Research Conference*, U.S. Bureau of the Census, 133–139.

Wolter, K. M. (1985), *Introduction to Variance Estimation*, New York: Springer-Verlag.