AN OVERVIEW OF MULTIPLE IMPUTATION

Donald B. Rubin, Harvard University
One Oxford Street, Cambridge, MA 02138

KEY WORDS: data-base construction, missing data, nonresponse, public-use files, survey methods.

ABSTRACT

Multiple imputation for nonresponse in public-use files replaces each missing value by two or more plausible values. The values can be chosen to represent both uncertainty about which values to impute assuming the reasons for nonresponse are known and uncertainty about the reasons for nonresponse. The theoretical underpinnings and several examples are given in Rubin (1987). This presentation illustrates the dramatic improvements possible when using multiple rather than single imputation and provides a brief overview of current technology and lacunae that, hopefully, will be addressed and filled by current research efforts. The two important applications of multiple imputation that this overview introduces, demonstrate the substantial improvements that can accrue from the straightforward use of multiple imputation in practice.

1. INTRODUCTION

Anyone involved in constructing data bases from survey data knows that essentially every survey suffers from nonresponse and the resultant missing values. Nevertheless, data-base constructors often feel obligated to create fixed-up data-bases for distribution -- ones without missing values. When missing values are left blank, complete-data statistics that would have been used in the absence of missing data can no longer be calculated, and data analysts can no longer use standard complete-data methods to draw inferences.

1.1 Imputation

It is not surprising, therefore, that a very common method of handling missing values is to fill them in -- impute them. That is, with imputation, each missing value is replaced by a real value. Many different procedures have been proposed for imputation, for instance, filling in the respondents' mean for that variable or a value predicted from the modeling of the missing variable given observed variables using respondent data; as a specific example, when the missing value is personal income, a linear regression model predicting log(income) from demographic characteristics such as age, sex, education and occupation might be regarded as reasonable. A common method of imputation in large surveys is "hot-deck" imputation; see Madow et al. (1983) for relevant definitions and references.

1.2 Advantages of Single Imputation

In addition to the obvious advantage of allowing complete-data methods of analysis, imputation performed by the data collector (e.g., the Census Bureau) also has the important advantage of allowing the use of information available to the data collector but not available to an external data analyst such as a university social scientist analyzing a public-use file. This information may involve detailed knowledge of interviewing procedures and reasons for nonresponse that are too cumbersome to place on public-use files, or may be facts, such as street addresses of dwelling units, that cannot be placed on public-use files because of confidentiality constraints. This kind of information, even though inaccessible to the user of a public-use file, can often improve the imputed values.

A third advantage of imputation by the data-base constructor is that the missing data problem is handled once, rather than many times by the users. This implies consistency of the data-bases across users, and a consequent consistency of answers from identical analyses. Too often the apparently same analysis (e.g., least squares regression) when applied to the apparently same data base will result in different conclusions because of differences in the way users and programs handle missing data. This situtation leads to unnecessary confusion and wasted resources. Imputation by the data-base constructor leads to greater consistency and thereby to reduced costs of this type.

1.3 Disadvantages of Single Imputation

Just as there are obvious advantages to imputing one value for each missing value, there are obvious disadvantages of this procedure arising from the fact that the one imputed value cannot itself represent any 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. Equally serious, single imputation cannot represent any additional uncertainty that arises when the reasons for nonresponse are not known.

The underrepresentation of uncertainty with single imputation can be a major problem. To illustrate this, assume that single imputations have been created "properly", meaning, as in Rubin (1987, Chapter 4) that imputations are randomly drawn from an appropriate distribution (worse results would be obtained for best-predicted-value methods). Table 1 presents frequency evaluations in large sample cases with 30% missing information -- a lot of missing data but not too extreme in many survey contexts. Note that the actual confidence coverage for a scalar parameter is quite a bit less than the nominal coverage, but even more dramatic, the rejection rate for a true null hypothesis about a 10-component parameter (e.g., a 10-component regression coefficient) is much larger than nominal. To ignore the sort of problems with single imputation demonstrated in Table 1 is to follow a path unrelated to scientific inference.

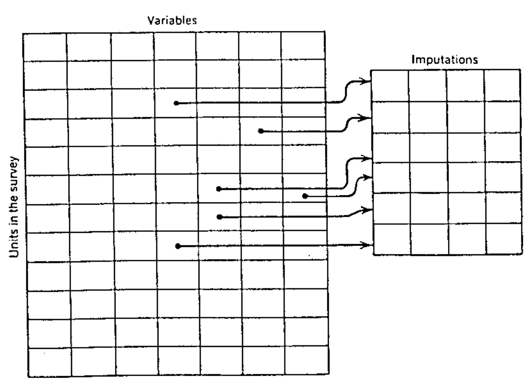
Table 1

Large-Sample Frequency Performance of Single Imputation with 30% Missing Information.

Confidence Coverage For Scalar Parameter	Nominal	90%	95%	99%
	Actual	77%	85%	94%
Significance Level for 10-Component Parameter	Nominal	1%	5%	10%
	Actual	25%	45%	57%

1.4 Multiple Imputation to the Rescue

Multiple imputation, first proposed in Rubin (1977,1978), retains the three major advantages of single imputation and rectifies its major disadvantages. As its name suggests, multiple imputation replaces each missing value by a vector composed of $M \geq 2$ possible values. The M values are ordered in the sense that the first components of the vectors for the missing values are used to create one completed data set, the second components of the vectors are used to create the second completed data set and so on; each completed data set is analyzed using standard complete-data methods. Figure 1 depicts a multiply-imputed data set.



Each row vector of imputations is of length M, where

Model for first imputation = ...
Model for second imputation = ...
.
.
Model for Mth imputation = ...

Figure 1. Data set with M imputations for each missing datum.

The first major advantage of single imputation is retained with multiple imputation, since standard complete-data methods are used to analyze each completed data set.

The second major advantage of imputation, that is, the ability to utilize data collectors' knowledge in handling the missing values, is not only retained but actually enhanced. In addition to allowing data collectors to use their knowledge to make point estimates for imputed values, multiple imputation allows data collectors to reflect their uncertainty as to which values to impute. This uncertainty is of two types: sampling variability assuming the reasons for nonresponse are known, and variability due to uncertainty about the reasons for nonresponse. Under each posited model for nonresponse, two or more imputations are created to reflect sampling variability under the model; imputations under more than one model for nonresponse reflect uncertainty about the reasons for nonresponse. The multiple imputations within one model are called repetitions, and repeated analyses based on them can be combined to form valid inferences under that model; the inferences under different models can be contrasted to reveal sensitivity of answers to posited reasons for nonresponse. Thus, multiple imputation rectifies both disadvantages of single imputation.

The third advantage of single imputation -- consistency of answers across users -- is also retained, since the same set of multiple imputations are being passed on to all users, and all users applying one analysis method will obtain the same answer.

The special advantages that multiple imputation can bring to statistical inference can be illustrated by reference to the cases of Table 1: with M=3 proper repetitions instead of M=1, and using simple procedures described in Section 3, nominal levels (to the nearest percent) are attained for all cases of Table 1!

Rubin (1987) is a comprehensive treatment of multiple imputation. Other references on multiple imputation include Rubin (1979, 1980, 1983, 1986a,b), Herzog and Rubin (1983), Li (1985), Schenker (1985), Rubin and Schenker (1986, 1987), Heitjan and Rubin (1986), Weld (1987), Raghunathan (1987), Little and Rubin (1987, Chapt. 11), and Treiman, Bielby and Cheng (1988).

2. GENERAL PROCEUDRES FOR CREATING A MULTIPLY-IMPUTED DATA SET

Multiple imputations ideally should be drawn according to the following general scheme. For each model being considered, the M imputations of the missing values, Y_{mis}, are M repetitions from the posterior predictive distribution of Y_{mis}, each repetition being an independent drawing of the parameters and missing values under appropriate Bayesian models for the data and the posited response mechanism. In practice, three important issues arise: explicit vs. implicit models, ignorable vs. nonignorable models, and proper imputation methods.

2.1 Explicit vs. Implicit Models

Explicit models are the ones usually used in mathematical statistics: normal linear regression, binomial, poisson, multinomial, etc. Implicit models are ones that can be thought of as underlying procedures used to "fix up" specific data structures in practice; often they have a "nonparametric", "locally linear", or "nearest neighbor" flavor to them. Although explicit models are the theoretical ideal precisely justifying multiple imputation technology, often implicit models can be used in place of explicit models. Both types of models are illustrated in Herzog and Rubin (1983), where repeated imputations are created using an explicit regression model and an implicit matching model, which is a modification of the Census Bureau's hot-deck. Many other examples of both kinds of models appear in Rubin (1987). Section 2.4, here, provides a specific example of an implicit model.

2.2 Ignorable vs. Nonignorable Models

The models underlying imputation methods, whether implicit or explicit, can be classified as assuming either ignorable reasons for missing data or nonignorable reasons. The term "ignorable" is coined in Rubin (1976) and is fully explicated in the context of multiple imputation in Rubin (1987). The basic idea is conveyed by a simple example in which X is observed for all units in the data base, and Y is missing for the nonrespondents but observed for the respondents. An ignorable model asserts that a nonrespondent is only randomly different from a respondent with the same value of X. A nonignorable model asserts that even though a respondent and nonrespondent appear identical with respect to X, their Y values systematically differ (e.g., the nonrespondent's Y is typically 20% larger than the respondent's Y with the same value of X). There is no direct evidence in the data to address the veracity of any such assumption, which is a good reason to consider several models and explore resultant sensitivity whenever possible.

2.3 Proper Imputation Methods

Imputation procedures, whether based on explicit or implicit models, or ignorable or nonignorable models, that incorporate appropriate variability among the repetitions within a model are called *proper*, which is defined precisely in Rubin (1987). The essential reason for using proper imputation methods is that they properly reflect sampling variability when creating repeated imputations under a model, and as a result lead to valid inferences. For example, assume ignorable nonresponse so that respondents and nonrespondents with a common value of X have Y values only randomly different from each other. Even then, simply randomly drawing imputations for nonrespondents from matching respondents' Y values ignores some sampling variability. This variability arises from the fact that the sampled respondents' Y values at X randomly differ from the population of Y values at X. Properly reflecting this variability leads to repeated imputation inferences that are valid under the posited response mechanism. Ideal imputation methods (i.e., fully Bayesian ones) are automatically proper.

In the context of simple random samples and ignorable nonresponse, Rubin and Schenker (1986) study hot-deck imputation (i.e., simply randomly drawing imputed values from respondents), which is *not* proper, and a variety of proper imputation methods based on both explicit and implicit models, including a fully normal model, the Bayesian Bootstrap (Rubin, 1981), and an approximate Bayesian Bootstrap. The Approximate Bayesian Bootstrap (ABB) can be used to illustrate how an intuitive imputation method based on an implicit model, such as the simple random hot-deck, can be modified to be proper.

2.4 Example of a Proper Imputation Method - The ABB

Consider a collection of n units with the same value of X where there are n_R respondents and $n_{NR} = n - n_R$ nonrespondents. The ABB creates M ignorable repeated imputations as follows. For $\ell = 1, \ldots, M$ create n possible values of Y by first drawing n values at random with replacement from the n_R observed values of Y, and second drawing the n_{NR} missing values of Y at random with replacement from those n values. The drawing of n_{NR} missing values from a possible sample of n values rather than the observed sample of n_R values generates appropriate between imputation variability, at least assuming large simple random samples at X, as shown by Rubin and Schenker (1986). The ABB approximates the Bayesian Bootstrap by using a scaled multinomial distribution to approximate a Dirichlet distribution.

The ABB can be made nonignorable in many ways. For example, at the first step, instead of drawing n values of Y at random from the n_R observed values, independently draw n values of Y with probability proportional to Y^2 (or some other function of Y). This will skew the nonrespondents to have typically larger values of Y than respondents with the same values of X.

2.5 Practical Imputation Methods

In practice, I believe that proper implicit models, both ignorable and nonignorable, will be the most useful from the perspective of the data-base constructor. In fact, I believe that in common practice, one can often "cheat" and use intelligently designed but inexpensive variants of existing single imputation techniques.

A very important point here is that the existence of missing data generally makes neat analysis impossible, and we should not waste a major portion of our resources fixing up a relatively minor problem (e.g., don't spend 80% of the budget fixing up the 30% of information that is missing). A sloppy argument, which can be made more formal, suggests that sensible but not pristine methods of multiple imputation will usually be adequate. Suppose we have a multiple imputation method that is, in some sense, 80% ok and we apply it to a problem with 30% missing information; then we are 20% deficient (not ok) on only 30% of the information, or 6% deficient overall, or 94% ok overall.

Work is needed to produce a modern multiple-imputation replacement for the workhorse "hot-deck" -- a "new-wave" hot-deck, and I see a variety of possibilities for real progress in this direction in the next few years.

3. GENERAL PROCEDURES FOR ANALYZING MULTIPLY-IMPUTED DATA SETS

The proper multiple imputations within each model are called repetitions and are combined (in ways to be described shortly) to create one inference under each model. The inferences across models are not combined but are contrasted to reveal sensitivity of inference to assumptions about the reasons for the missing data. The critical issue then is how to analyze the repetitions within one model to yield a valid inference under the posited reasons for missing data. The key idea is that M repetitions yield M completed data sets, each of which can be analyzed by standard complete-data methods just as if it were the real data set. The M complete-data analyses based on the M repeated imputations are then combined to create one repeated-imputation inference.

3.1 The Repeated-Imputation Inference For Point and Interval Estimation

Let $\hat{\theta}_\ell$, U_ℓ, $\ell = 1, \ldots, M$ be M complete-data estimates and their associated variances for a parameter θ, calculated from the M data sets completed by repeated imputations under one model for nonresponse. For instance, for a regression analysis, $\theta=\beta$, $\hat{\theta}_\ell$=the least squares estimate of β, and U_ℓ=(residual mean square) x $(X^TX)^{-1}$, in the standard notation. The final estimate of θ is

$$\bar{\theta} = \sum_{\ell=1}^{M} \hat{\theta}_\ell/M \quad .$$

The variability associated with this estimate has two components: the average within-imputation variance,

$$\bar{U} = \sum_{\ell=1}^{M} U_\ell/M \quad ,$$

and the between-imputation component,

$$B = \sum (\hat{\theta}_\ell - \bar{\theta})^2 \;/\; (M - 1)$$

where with vector θ, $(\bullet)^2$ is replaced by $(\bullet)^T(\bullet)$. The total variability associated with θ is then

$$T = \bar{U} + (1 + M^{-1})B \quad .$$

With scalar θ, the approximate reference distribution for interval estimates and significance tests is a t distribution:

$$(\theta - \bar{\theta})\; T^{-1/2} \sim \; t_v \quad ,$$

where the degrees of freedom,

$$v = (M - 1)\;\{1 + [(1 + M^{-1})B/\bar{U}]^{-1}\}^2 \quad ,$$

is based on a Satterthwaite approximation (Rubin and Schenker 1986, and Rubin 1987). The within to between ratio, $r=\bar{U}/B$, estimates the population quantity $(1 - \gamma)/\gamma$, where γ is the fraction of information about θ missing due to nonresponse. In the case of ignorable nonresponse with no covariates, γ equals the fraction of data values that are missing, but typically γ is less than this because of dependence between variables with the attendant ability to improve prediction of missing values from observed values.

Although interval estimation based on this t reference distribution works quite well in most cases of practical importance, slightly better interval estimates are obtainable when M is small and γ is large by using a special case of the Behrens-Fisher distribution as a reference distribution. Current joint work with Raghunathan addresses this improvement.

3.2 Significance Levels for Multicomponent θ from $\{\hat{\theta}_\ell, U_\ell; \ell=1,\ldots,M\}$

For θ with k components, significance levels for null values of θ can be obtained from M repeated complete-data estimates, $\hat{\theta}_\ell$, and variance-covariance matrices, U_ℓ, using multivariate analogues of the previous expressions.

A simple procedure described in Li (1985) and Rubin (1987) that works well for M large relative to k is to let the p-value for the null value θ_0 of θ be Prob $\{F_{k,v} > D\}$ where $F_{k,v}$ is an F random variable and

$$D = (\theta_0 - \bar{\theta})T^{-1}(\theta_0 - \bar{\theta})^T/k$$

with v defined by generalizing $r=B/\bar{U}$ to be the average diagonal element of $B\bar{U}^{-1}$,

$$r = \text{trace}(B\bar{U}^{-1})/k.$$

A better procedure when M is modest, advocated in Rubin (1987), is to let the p-value be given by Prob$\{F_{k,v(k+1)/2} > \tilde{D}\}$ where F and v are as previously defined, and

$$\tilde{D} = (\theta_0 - \bar{\theta})\;\bar{U}^{-1}(\theta_0 - \bar{\theta})/[(1+r)k] \quad .$$

This procedure is quite accurate, except for large k when it tends to be too conservative due to the approximate nature of the reference distribution.

An extremely accurate procedure when $M \geq 3$ is described in forthcoming joint work with Li and Raghunathan. This procedure refers the test statistic $\tilde{D}$ to an F distribution on k and w degrees of freedom where

$$w = 4 + [k(M-1)-4](1+a/r)^2$$

with

$$a = \left[\; 1 - \frac{2}{k(M-1)} \;\right] \quad (1 + 1/M)^{-1}$$

Current joint work with Raghunathan produces extremely accurate results from $\tilde{D}$ when M=2.

3.3 Significance Levels from Repeated Significance Levels

With large data sets and large models, such as occur often with multiway contingency tables in social science research, a complete-data analysis may only produce a p-value or equivalently the χ^2 statistic on each completed data set:

$$d_\ell = (\theta_0 - \hat{\theta}_\ell)\; U_\ell^{-1}(\theta_0 - \hat{\theta}_\ell)^T \quad .$$

The problem of directly combining the $\{d_\ell, \ell=1,\ldots,M\}$ is very tricky because each d_ℓ typically leads to a p-value that is too extreme (i.e., too significant). The representation that makes progress possible is to note that in large samples

$$\hat{D} \doteq \tilde{D} \quad ,$$

where

$$\hat{D} = \frac{\left[\; \dfrac{\bar{d}}{k} - \dfrac{M - 1}{M + 1}\; r \;\right]}{(1 + r)} \quad ,$$

and $\bar{d}$ is the average d_ℓ. Replacing r in $\hat{D}$ with estimates obtained from the d_ℓ yields procedures that are acceptable in many cases. Li (1985), Rubin (1987), Raghunathan (1987), and Weld (1986) show that in common situations replacing r by a method of moments estimate yields satisfactory results. Raghunathan (1987) derives several procedures that work better than this one, but these improvements are more awkward to implement in practice.

Current research by Xiao Li Meng at Harvard suggests that a better procedure is obtained by replacing r with $\hat{r}=(1+1/M)v(\surd d)$, where $v(\surd d)$ is the variance of the M values of $\surd d_\ell$, to yield

$$\hat{D} = \frac{\frac{\bar{d}}{k} - \left(1 - \frac{1}{M}\right) v(\surd d)}{1 + \left(1 + \frac{1}{M}\right) v(\surd d)} ,$$

which is referred to a translated F distribution, or even an F distribution on k and av degrees of freedom where $a=k^{(-3/M)}$ and r in the expression for v is replaced by $\hat{r}$.

4. FREQUENCY EVALUATIONS

Although repeated-imputation inferences are most directly motivated from the Bayesian perspective, they can be shown to possess good frequency properties. In fact, the definition of proper imputation methods means that in large samples infinite-M repeated imputation inferences will be valid. Since the finite-M adjustments are derived using approximations to Bayesian posterior distributions, however, some deficiencies can arise with finite M.

4.1 Relative Efficiency of Point Estimation

The large sample relative efficiency of the finite-M repeated imputation estimator using proper imputation methods relative to the infinite-M estimator, in units of standard errors is $(1 + \gamma/M)^{-1/2}$. Even for relatively large γ, modest values of M result in estimates $\hat{\theta}$ that are nearly fully efficient. For example, for γ=30% and M=3, the relative efficiency of $\hat{\theta}$ is approximately 95%.

4.2 Confidence Coverage

In large samples, the confidence coverage of proper imputation methods using the t reference distribution can be tabulated as a function of M, γ and the nominal level, $1 - \alpha$. Extensive results are given in Rubin and Schenker (1986) and Rubin (1987). (With single imputation, the between component of variance is automatically set to zero, since it cannot be estimated, and the reference distribution is the normal, since v cannot be estimated without B.) As stated in Section 1, three repeated imputations yield essentially valid confidence coverages, which is in striking contrast to the results using only one imputation displayed in Table 1. Even worse coverages for single imputation would have been obtained using best prediction methods, such as "fill in the mean".

4.3 Significance Levels

Work on accurately obtaining significance levels is at a relatively early stage of development, but much effort has been expended and many tables are given in Li (1985), Rubin (1987), and Raghunathan (1987), and tremendous improvements can accrue when using multiple rather than single imputation. If the repeated moments $\{\hat{\theta}_\ell, U_\ell\}$ are available, essentially perfect results are obtained for the cases of Table 1 using $\tilde{D}$ or the asymptotically equivalent D, and either the old or new improved F reference distributions.

If only the repeated χ^2 statistics are available, use of $\hat{D}$ leads to significance levels close to nominal for the cases of Table 1. For example, using the F reference distribution for this case, the rejection rates are 2%, 6%, and 10% for nominal 1%, 5%, and 10% tests, which are in stark contrast to the 25%, 45%, and 57% rates obtained with single imputation.

5. CURRENT APPLICATIONS

The large sample frequency evaluations of Section 4 clearly support the contention that multiple imputation is a very promising new tool for helping to handle nonresponse in surveys. Of course, more important to the applied researcher is whether the theory really works in applications. Fortunately, several applications both major and exploratory support this contention as well. Since two actual applications follow directly, this review will be exceedingly cursory.

5.1 Major Applications

Since 1982 I've been deeply involved with the Census Industry and Occupational Coding Project to produce public-use files with multiply-imputed codes. In fact I regard the willingness of many to give multiple imputation a try on this important problem a crucial boost for this technique. Several articles now exist indicating the success of this venture, and we are fortunate to have the current contribution by Schenker, Treiman and Weidman.

A second major application, with which I've only been peripherally involved, concerns ETS's multiply-imputing test results in the National Assessment of Educational Progress. As I understand it, the length of the full test precluded it being given to every subject, so overlapping subsets of the test items were given in random fashion to different groups of subjects, thereby intentionally creating blocks of missing data. Since data bases that could be analyzed using standard complete-data methods are considered necessary, multiple imputation is being employed.

5.2 Exploratory Applications

More than ten years ago, exploratory applications of multiple imputation were done using the CPS-IRS-SSA exact match file (Herzog and Rubin, 1983; Aziz, Kilss, and Scheuren, 1978); in fact, this project really stimulated my initial proposal for doing multiple imputation. Related work, joint with R.J.A. Little, J. Czajka, Susan Hinkins, and Fritz Scheuren, continues at IRS in the context of editing files, where only a subset of the files can be edited due to financial constraints. Another exploratory application involves multiply imputing for coarsely reported age data in a demographic survey (Heitjan and Rubin, 1986).

Also, there is the NHTSA FARS data base with its missing values, especially on BAC, which

Heitjan and Little consider here; I also am involved in an aspect of this imputation project directly through DOT in Cambridge, MA.

But enough of this overview -- let's go on to see data from actual applications.

ACKNOWLEDGEMENTS

This work was supported by the National Science Foundation, Grant Number SES-8805433, and presents material similar to that in Rubin (1988).

REFERENCES

AZIZ, F., KILSS, B., and SCHEUREN, F. (1978). Current population survey -- administrative record exact match file. *Report for Studies from Interagency Data Linkages.* Washington D.C.: Social Security Administration.

HEITJAN, D.F., and RUBIN, D.B. (1986). Inference for coarse data using multiple imputation. *Proceedings of the 18th Symposium on the Interface of Computer Science and Statistics.*

HERZOG, T.N., and RUBIN, D.B. (1983). Using multiple imputations to handle nonresponse in sample surveys. In *Incomplete Data in Sample Surveys, Volume 2: Theory and Bibliography*, New York: Academic Press, 209-245.

LI, K.H. (1985). *Hypothesis Testing in Multiple Imputation - with Emphasis on Mixed-Up Frequencies in Contingency Tables.* Ph.D. Thesis, Department of Statistics, University of Chicago.

LITTLE, R.J.A., and RUBIN, D.B. (1987). *Statistical Analysis with Missing Data.* New York: John Wiley and Sons.

MADOW, W.G., et al. (1983). *Incomplete Data in Sample Surveys*, Vols. 1, 2, 3. New York: Academic Press.

RAGHUNATHAN, T.E. (1987). *Large Sample Significance Levels from Multiply-Imputed Data.* Ph.D. Thesis, Department of Statistics, Harvard University.

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

RUBIN, D.B. (1977). The design of a general and flexible system for handling nonresponse in sample surveys. *Report prepared for the U.S. Social Security Administration.*

RUBIN, D.B. (1978). Multiple imputations in sample surveys - a phenomenological Bayesian approach to nonresponse. *Proceedings of the Survey Research Methods Section of the American Statistical Association*, 20-34. Also in *Imputation and Editing of Faulty or Missing Survey Data*, U.S. Dept. of Commerce, 1-23.

RUBIN, D.B. (1979). Illustrating the use of multiple imputations to handle nonresponse in sample surveys. *Proceedings of the 1979 Meetings of the ISI-IASS, Manila.*

RUBIN, D.B. (1980). *Handling Nonresponse in Sample Surveys by Multiple Imuputations.* U.S. Dept. of Commerce, Bureau of the Census Monograph.

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

RUBIN, D.B. (1983). Progress report on project for Multiple imputation of 1980 codes. *Report distributed to the Bureau of the Census, The National Science Foundation, and the Social Science Research Council.*

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

RUBIN, D.B. (1986b). Statistical matching using file concatenation with adjusted weights and multiple imputations. *Journal of Business and Economic Statistics*, 87-94.

RUBIN, D.B. (1986a). Basic ideas of multiple imputation. *Survey Methodology*, 12, pp. 37-47.

RUBIN, D.B. (1988). Multiple imputation for data-base construction. *COMPSTAT-1988*, D. Edwards and N. Raun (eds.) Heidelberg: Physica Verlag.

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. (1987). Interval estimation from mulitply-imputed data: a case study using agriculture industry codes. *Journal of Official Statistics*, 3, 4, pp. 375-387.

SCHENKER, N. (1985). *Multiple Imputation for Interval Estimation from Surveys with Ignorable Nonresponse.* Ph.D Thesis, Department of Statistics, University of Chicago.

TREIMAN, D.J., BIELBY, W., and CHENG, M. (1988). Evaluating a multiple-imputation method for recalibrating 1970 U.S. Census detailed industry codes to the 1980 standard. *Sociological Methodology*, 18.

WELD, L.H. (1987). *Significance Levels from Public-Use Data with Multiply-Imputed Industry Codes.* Ph.D Thesis, Department of Statistics, Harvard University.