Research in Nursing & Health, 2002, 25, 76–84
DOI 10.1002/nur.10015

Focus on Research Methods

Multiple Imputation for Missing Data

Patricia A. Patrician*

Walter Reed Army Medical Center, Washington, DC
Received 15 May 2001; accepted 19 October 2001

Abstract: Missing data occur frequently in survey and longitudinal research. Incomplete data are problematic, particularly in the presence of substantial absent information or systematic nonresponse patterns. Listwise deletion and mean imputation are the most common techniques to reconcile missing data. However, more recent techniques may improve parameter estimates, standard errors, and test statistics. The purpose of this article is to review the problems associated with missing data, options for handling missing data, and recent multiple imputation methods. It informs researchers' decisions about whether to delete or impute missing responses and the method best suited to doing so. An empirical investigation of AIDS care data outcomes illustrates the process of multiple imputation. © 2002 John Wiley & Sons, Inc. Res Nurs Health 25:76–84, 2002

Keywords: missing data; multiple imputation; nonresponse; cumulative logit models

Missing data are universally problematic in survey and longitudinal research. In surveys individuals may not respond to certain questions. Participants inadvertently skip questions, may not have the requested information at hand, or choose not to respond. In longitudinal studies participants relocate, die, or drop out for other reasons. The uninformed researcher may analyze these incomplete data inappropriately or not at all. Recent theoretical and computational advances, most notably multiple imputation (MI) methods, enable the researcher to use the existing data to generate, or impute, values approximating the "real" value, while preserving the uncertainty of the missing values (Schafer, 1997/2000). This article outlines the problems associated with missing data, reviews standard methods of handling absent values, and highlights the usefulness of multiple imputation. An example illustrates how and why MI works.

The author thanks Linda Aiken, Douglas Sloane, and Julie Sochalski for reviewing earlier drafts and Paul D. Allison for his thoughtful critique of this paper.

The opinions or assertions contained here are the private views of the author and are not to be construed as official or as reflecting the views of the Department of the Army or the Department of Defense.

Contract grant sponsor: National Institute for Nursing Research; contract grant number: NR02280.

Contract grant sponsor: Agency for Healthcare Research and Quality; contract grant number: HS08603.

Contract grant sponsor: Army Nurse Corps.

Correspondence to Patricia A. Patrician, Lieutenant Colonel, Army Nurse Corps, Nursing Research Service, P.O. Box 59645, Walter Reed Station, Washington, DC 20012.

*Assistant Chief, Nursing Research Service.

 © 2002 John Wiley & Sons, Inc.

Although a few researchers recommend statistical consultation for complex methods of handling missing data (Ferketich & Verran, 1992), all investigators should have a basic understanding of these analytical methods. Mechanisms for resolving missing data are a valuable addition to any nurse researcher's methodological toolkit. Nurse researchers who analyze complex multivariate data must decide how to account for incomplete information. As nurse researchers also increase their use of secondary data, they must be prepared to judge the database constructor's method of resolving missing data.

MISSING DATA—A PERVASIVE PROBLEM

Missing data are problematic for a number of reasons. First, most statistical procedures rely on complete-data methods of analysis (Allison, 2000; Rubin, 1987). That is, computational programs require that all cases contain values for all variables to be analyzed. Thus, as a default, most statistical software programs exclude from analysis cases that have missing data on any of the variables (listwise deletion). The analyst proceeds as if all remaining cases comprise the entire data set. This can lead to two potentially serious problems: compromised analytic power and nonresponse bias.

Depending on the percentage of missing information, analytic power may be significantly reduced if the researcher excludes all cases missing data for one or more variables (Allison, 2000; Heitjan, 1997). In multivariate analysis incomplete data for any one of the study variables renders the entire case useless. A significant proportion of the original sample could be excluded from analysis. Discarding cases can bias a study severely. Analytic power also may be reduced because of the decreased sample size. Researchers tend to oversample based on expected response rates, but in surveys it is highly unlikely that each individual will respond to each question, even with high overall response rates.

Another problem with missing values is nonresponse bias (Barnard & Meng, 1999). When respondents who do not answer a particular question differ from those who answer, a systematic pattern or bias characterizes the missing data (Tabachnick & Fidell, 2000). Nonresponders may decide to omit certain questions for very distinct reasons that researchers may never know. To assess for nonresponse bias, one can determine whether respondents differ from nonrespondents. Of course, data are needed on variables that relate to the nonresponse. Therefore, this is best determined when there is nonresponse on some but not all the questions. One can simply use dummy variables for the responders versus nonresponders and test for differences between the groups.

DESCRIBING MISSING DATA

To determine how to handle missing data, one must identify whether data are missing completely at random (MCAR) or missing at random (MAR). Responses are said to be missing *completely* at random if the probability of missing data on one variable is not related to the value of that variable or to other variables in the data set (Allison, 2000). If, for example, nonresponses to a survey item asking about a respondent's weight was neither a result of a respondent's weight nor responder's gender, we would conclude that the values for weight were MCAR. On the other hand, if we were to discover that overweight people or females tended not to report their weight, the responses on weight would *not* be MCAR. Whether data are MCAR can be verified partly by comparing the nonresponders to the responders on all other variables (Allison, 2000). However, it would be nearly impossible to determine whether the probability of missing data was a result of the value of the variable itself.

Data can be missing at random, a less restrictive notion than MCAR. MAR occurs when the probability of a missing value is not dependent on the value itself but may depend on the values of other variables in the data set (Allison, 2000). In the above case, for example, if the missingness of the variable *weight* was not dependent on weight, we would say that the weight values are MAR, regardless of whether other variables are associated with the missing weight values. By contrast, according to Rubin (1987), individuals at either very high or very low income levels generally tend not to report their income in census surveys. In this case nonresponse is very dependent on income. We would conclude that census income data are *not* MAR. The MAR assumption is impossible to test directly because there is no way of knowing the values of the missing data (Allison, 2000). However, it is important to determine whether a proxy can serve as the variable with missing data. For example, zip codes or job information may serve as proxy measures for income. The researcher might explore then

whether individuals in certain job categories omit income responses.

Rubin (1987) distinguished between item and unit nonresponse. Item nonresponse occurs when an individual fails to respond to a question or questions on a survey. Unit nonresponse exists when an individual fails to return the survey. For example, if a hospital is the unit of analysis and the nurse-executive at each hospital is the respondent, those who fail to return the survey would be unit nonresponsive. If, however, the nurse-executive completes the survey except for several items, this would be a case of item nonresponse. This distinction is important for some of the approaches to handling missing data.

APPROACHES TO HANDLING MISSING DATA

Missing data traditionally have been handled by deletion from analysis of cases that contain absent values, by single imputation, and, most recently, by using multiple imputation techniques. In this section the benefits and drawbacks of these traditional methods are reviewed.

Listwise Deletion

Before 1980 statisticians viewed missing data as "something to be gotten rid of" (Schafer, 1998, p. 3). The technique known as listwise deletion simply discards cases with missing data, restricting analysis to those cases with a full complement of values. Virtually all statistical software programs incorporate listwise deletion as a default because it accommodates any type of statistical analysis (Allison, 2000). If the data are MCAR, deletion yields unbiased parameter estimates but larger standard errors because the sample size declines. However, listwise deletion can lead to misleading results if a large proportion of the data is discarded. This is also problematic if the data are not MCAR. If the data are MAR, but not completely at random, listwise deletion may lead to biased estimates, that is, regression coefficients that are erroneously too large or too small. If a large portion of the data are missing, analytic power may be severely diminished.

Pairwise Deletion

Pairwise deletion is another case exclusion method used for correlations and for many linear models, such as factor analysis and linear regression (Allison, 2000). This technique eliminates cases if one or both of the values are unavailable for correlation (SPSS Manual, 1999). A correlation matrix may yield varied sample sizes for different bivariate correlations depending on the percentage of missing information. To compute the coefficients for regression analyses, only complete data for pairs of correlated variables are used (Allison, 2000). Like listwise deletion, pairwise deletion may create significant biases when the missing data are not MCAR (Allison, 2000).

Regardless of the deletion method, getting the correct standard error calculations and evaluating potential bias are major issues. Deletion techniques erode efficiency such that the variation around the true estimate (i.e., the standard error) is too large (Allison, 1999). If the data are not MCAR, bias may also be a serious issue. Case deletion methods assume that deleted cases are a random subsample of the data set (Schafer, 1999), which of course is erroneous.

Weighting Techniques

Another method of handling missing data is to weight respondents by how many units they may represent. This is where unit nonresponse is important. Rubin (1987) provided an example of those in schools (principals) that do not respond to surveys. The schools most similar demographically to the nonrespondent schools are weighted by the number of similar unit nonrespondents. For example, if one respondent school is very similar in applicable characteristics to two nonresponding schools, the responding school would be weighted by three. This assumes there is no unit nonresponse bias, which requires supplemental information about the nonrespondents. In addition, weighting techniques can become quite cumbersome if unit nonresponse is combined with item nonresponse. Little and Rubin (1989) suggested that such weighting reduces the bias that arises from case deletion methods but makes standard error calculation more difficult. Indeed, weighting decreases sample variance because multiple identical values are replacing the missing values.

Single Imputation

Imputation, yet another alternative to dealing with missing cases, involves ascribing a value to a

missing data cell based on the values of other variables or of substituting a reasonable estimate for absent data elements (Little & Rubin, 1989). With single imputation, one value is ascribed to the absent value. For example, sometimes a mean value is used to represent missing data. Although imputing the mean is easy to understand and simple to compute, it eliminates data that may be unique to a particular individual and ascribes the "usual" to that person. The mean naturally depends on the study sample. The major problem with mean imputation is that nonresponse bias is ignored. There may be very distinct reasons why individuals do not respond to certain items. As Rubin (1987) noted, persons of very high and very low socioeconomic status often do not report their income. In such cases imputing the mean leads to erroneous statistical inferences. Mean imputation decreases variability between individuals' responses and biases correlations with other variables (Tabachnick & Fidell, 2000). Single imputation techniques do not account for variability between imputations because only one value is imputed (Schafer, 1999). This results in increasing the potential sample size for analysis but decreasing the variance by use of a mean value for substitution. Decreased variances are problematic because the resulting estimates are too close to the mean (Tabachnick & Fidell, 2000).

The Census Bureau uses the "hot deck" single imputation method in its Current Population Survey (Reilly & Pepe, 1997). This procedure matches individuals with missing data with those having similar values in a set of other variables and imputes the known value into the missing data cell. The flaw in hot deck imputation, according to Rubin (1987), is its treatment of imputed data with certainty, thus perhaps grossly underestimating variability. It also assumes no difference between respondents and nonrespondents.

Linear regression is yet another type of imputation. Predictive equations generate the imputed values using complete-case information. This method may work well when predictors are strong, but here, too, the absence of sufficient variability causes underestimation of standard errors (Little & Rubin, 1989). Stochastic regression is an alternative that adds a residual error term to the predicted score (Landerman, Land, & Pieper, 1997). However, both of these techniques can result in biased parameter and standard error estimates.

There are advantages and disadvantages to single imputation. It is advantageous for a researcher to use complete-data methods of analysis, which single imputation allows (Rubin, 1987). Single computation and incorporation of the analyst's knowledge base is an additional advantage—the analyst, who may have access to data elements not available in public use files such as geographic locale, chooses variables for the imputation process. On the other hand, single imputation treats imputed values as if they were true and thus overstates precision (Little & Rubin, 1989). As mentioned previously, variability decreases, which affects the plausibility of the parameter estimates and error terms.

MULTIPLE IMPUTATION

To correct the problems with single imputation, Rubin (1977, 1987) developed a new method of imputation. Allison (2000) and Schafer (1997/2000) further explicated Rubin's multiple imputation (MI). MI is a predictive approach to handling missing data in multivariate analysis. MI blends both classical and Bayesian statistical techniques and relies on specific iterative algorithms to create several imputations. MI aims to create plausible imputations of the missing values, to accurately reflect uncertainty, and to preserve important data relationships and aspects of the data distribution (Freedman & Wolf, 1995; Schafer, 1998). MI requires that the analyst specifies an imputation model, imputes several data sets, analyzes them separately, and then combines results. MI yields a single set of test statistics, parameter estimates, and standard errors.

Assumptions and Constraints

Multiple imputation is not without certain assumptions. First, missing data should be MAR. Second, the imputation model must match the model used for analysis (Allison, 2000). Rubin (1987) termed this a "proper" imputation model. Schafer (1999) explained that the imputation model must preserve all important associations among variables in the data set, including interactions if they will be part of the final analysis. Likewise, the dependent variable must be included in the imputation model (Schafer, 1997/2000). Finally, the algorithm used to generate imputed values must be "correct," that is, it must accommodate the necessary variables and their associations. Allison (1998) illustrated this, comparing disparate results of two algorithms for producing multiple imputations. The first

algorithm considered only the variables associated with the missingness of the data and the second included other variables and their associations. Allison's results clearly support Rubin's (1989) contention that good imputation methods use all information related to missing cases.

Advantages

The advantages of MI build on the benefits of single imputation. MI allows use of complete-data methods for data analysis and also includes the data collector's knowledge. Moreover, MI incorporates random error because it requires random variation in the imputation process. Because repeated estimations are used, MI produces more reasonable estimates of standard errors than single imputation methods. MI can accommodate any model and any data and does not require specialized software. MI simulates proper inferences from data; it also increases efficiency of the estimates because MI minimizes standard errors (Rubin, 1987). It also allows randomly drawn imputations under more than one model.

Disadvantages

According to Rubin (1987), the three disadvantages of multiple imputation compared with other imputation methods are more effort to create the multiple imputations, more time to run the analyses, and more computer storage space for the imputation-created data sets. Hard-disk storage capacity is hardly an issue now. Macros, discussed later, simplify the creation and analyses of multiple data sets. These disadvantages lessen and pose fewer problems as time and technology advance.

MI has been criticized because of novices' reactions to it. Rubin (1996) said that initially MI was thought to be unacceptable because it used simulation and added random noise to the data. MI was once considered by many to be a form of "statistical alchemy" (Schafer, 1999). But Schafer (1999) claimed that MI accurately represents the observed information; it is not simply conjured data.

A final disadvantage of MI is its not producing a unique answer. Because randomness is preserved in the MI process, each data set imputed will yield slightly different estimates and standard errors. Therefore, the reproducibility of exact results may be problematic.

MULTIPLE IMPUTATION—THE PROCESS

The MI procedure is fairly straightforward. MI involves generating several data sets and analyzing them separately. The resulting estimates and standard errors are then combined through averaging formulas into one set of parameters. The extent of missing data determines the number of data sets to impute; most statisticians in this field recommend 3–5 sets (Allison, 2000; Freedman & Wolf, 1995).

Approximate Bayesian bootstrap (Rubin & Schenker, 1986) and predictive mean matching (Landerman et al., 1997) are two of several algorithms with corresponding software available for MI calculations. However, Allison (1998) found the manner in which the Bayesian bootstrap method was used was less than optimal. Allison (2000) and Schafer (1997/2000) both discussed a shareware package called NORM, available at http://www.stat.psu.edu/~jls. Allison (1999) also has developed macros for the SAS program that mimic the algorithms of NORM. These are available at http://www.ssc.upenn.edu/~allison. Allison's MISS macro generates the imputations, and the COMBINE macro calculates a single set of estimates and standard errors from the imputed data sets. The advantage of using Allison's SAS macros is not having to switch between a freestanding imputation program and a data analysis package. Therefore, the following discussion will center on Allison's and Schafer's recommended MI method—the expectation–maximization (EM) algorithm and data augmentation.

To do MI, an imputation model first must be selected. The most common model for MI is the multivariate normal (Allison, 2000; Schafer, 1997/2000). The multivariate normal model assumes all variables have normal distributions, are linearly related, and have a normal homoscedastic error term (Allison, 2000). The model should include all variables in the desired analysis plus others predictive of the missing information. The model should also include variables that describe special aspects of the sample (Schafer, 1998). It is best, according to Allison (2000), to err on the inclusive side when selecting variables. Highly skewed variables must be transformed to approximate normality (Tabachnick & Fidell, 2000).

Next, maximum likelihood estimates of the means and covariance matrix must be generated using the EM algorithm (Dempster, Laird, & Rubin, 1977). The EM algorithm has two steps, expectation and maximization. These steps

reoccur until the iterative process converges and yields a maximum likelihood estimate. The expectation step is akin to regression imputation of missing data. The maximization step uses the imputed values to calculate new values for the means and covariance matrix. After this, the expectation step resumes and, using the new mean and correlations, calculates new imputations. Convergence occurs when the estimates barely change from one iteration to the next (Allison, 2000).

In the next step data augmentation generates the multiple data sets. Data augmentation is similar to EM, but the EM algorithm is deterministic, whereas data augmentation is stochastic or probabilistic (Allison, 2000). Data augmentation, a Bayesian method (Iversen, 1984), recomputes the covariance matrix. As a member of the class of Markov Chain Monte Carlo algorithms (Schafer, 1998), data augmentation treats parameters as random variables and randomly draws these parameters from their posterior distribution. The posterior distribution is a Bayesian probability distribution created from both a prior and a conditional distribution (Iversen, 1984; Kennedy, 1992).

Data augmentation uses the initial values obtained from the EM algorithm. As an alternative, the analyst may use starting values from a listwise deletion analysis; however, data augmentation requires the analyst to specify the number of iterations for the convergence of the estimates. A good way to determine the necessary number of iterations is to examine the number of iterations required for the EM algorithm (Allison, 2000).

After all data sets have been imputed separately, the transformed variables must be back-transformed. Categorical variables, imputed under the normal model, must be rounded to the nearest category (Schafer, 1998). Then each respective data set is analyzed. In the final step in MI the results are combined into a single set of parameter estimates, test statistics, and standard errors. An average will suffice for the estimates and test statistics, but standard errors require a specific computation formula that acknowledges the number of imputations and the variance among the parameter estimates (Allison, 2000). The COMBINE macro will calculate all these values.

MULTIPLE IMPUTATION—AN EXAMPLE

To illustrate the MI procedure, data were obtained from a national study of AIDS care outcomes (Aiken, Lake, Sochalski, & Sloane, 1997). In addition to studying patient outcomes, the researchers addressed various outcomes of nurses who worked with AIDS patients. The data collection included a prospective component designed to assess nurses' exposures to needle-stick injuries. For 30 days nurses completed minisurveys at the conclusion of each shift. In the shift surveys nurses were asked about needle-stick injuries and about workload, staffing, shift worked, and satisfaction with the shift. The measures and response codes for these variables are shown in Table 1. The ultimate research aims were to determine how nurses' daily satisfaction changed in response to various situations encountered daily and how the work environment might mediate this relationship. Because the data collection proceeded in two waves and because only Wave 2 had relatively complete information on workload and staffing (different measures were used during Wave 1 data collection), the Wave 2 subsample was selected for this analysis. The purpose was to determine the ability of MI

Table 1. Shift-Specific Measures

Variable	Measure	Response Scale/Coding
Satisfaction	"How would you describe your shift?"	4-point faces scale: 1 = very sad, 2 = sad, 3 = happy, 4 = very happy
Injury	"During this shift did you have an injury from a needle/sharp?"	1 = yes, 0 = no
Workload	"How would you describe your workload this shift"	5-point numeric scale: 1 = light, 5 = heavy
Staffing	"During this shift, was your unit understaffed compared to normal?"	1 = yes, 0 = no
Evening/ night shift	"Fill in the date of your shift and its duration."	Time: from___a.m./p.m. to___a.m./p.m. (Coded as 1 = evening and night, 0 = day)

Table 2. GEE Parameter Estimates and Empirical Standard Errors (Wave 2 Subsample)

			95% CI	
	B	*SE* (B)	Lower	Upper
Intercept 1	−7.83***	.25	−8.32	−7.35
Intercept 2	−5.93***	.22	−6.36	−5.49
Intercept 3	−2.73***	.18	−3.09	−2.37
Injury	.86*	.42	.05	1.68
Workload	1.22***	.06	1.12	1.33
Staffing	.91***	.10	.71	1.11
Evening/night shift	.23***	.06	.12	.34

$N = 5{,}794$ observations.
*$p < .05$, **$p < .001$, ***$p < .0001$.

to generate appropriate parameter estimates and standard errors in this subsample from which data were randomly removed. The Wave 2 sample included 5,794 completed shift surveys from more than 400 staff nurses who worked on 20 units in 10 hospitals. An unknown number of temporary or agency nurses were also included.

All analyses were done using SAS Version 8. Models were fit using the GENMOD procedure, which allows the fitting of cumulative logit models to the data. Satisfaction, the outcome variable, had four ordered categories. In addition, one option with the GENMOD procedure is the generalized estimating equation (GEE) method, which accounts for our nonindependent measures (i.e., repeated measures on the nurses and clustering of the nurses within units). The basic SAS programming steps used in this analysis are in the Appendix.

First, the predictors of daily work satisfaction were analyzed using listwise deletion on the Wave 2 data. Missing values accounted for approximately 1,100 deleted observations. The parameter estimates, standard errors, and 95% confidence levels are shown in Table 2. Injury, workload, staffing, and evening/night shift were significant predictors of satisfaction. Note the three intercept terms. These correspond to the three nonredundant comparisons that can be made among the four dependent variable categories.

Next, 25% of the values for workload and staffing were removed at random. Because of the randomly generated missing values, data were at least MAR. A cumulative logit model was fitted to these data, again using listwise deletion. The results in Table 3 show the intercepts and coefficients when only 4,500 observations of the 5,794 were used in the analysis. Next the missing values were imputed using Allison's MISS macro to create five imputed data sets and to fit cumulative logit models for each. Care was taken to exclude from analysis the cases with a missing satisfaction measure so as to avoid imputing values for the dependent variable. The estimates and standard errors for each imputed data set are shown in Table 4. Finally, the COMBINE macro was used to aggregate results for the parameter

Table 3. GEE Parameter Estimates and Empirical Standard Errors (Wave 2 Subsample With 25% Missing Data)

			95% CI	
	B	*SE* (B)	Lower	Upper
Intercept 1	−7.70***	.26	−8.20	−7.20
Intercept 2	−5.78***	.23	−6.23	−5.33
Intercept 3	−2.60***	.19	−2.98	−2.23
Injury	.74	.49	−.22	1.69
Workload	1.18***	.57	1.07	1.30
Staffing	.96***	.11	.74	1.76
Evening/night shift	.20**	.06	.09	.32

$N = 4{,}500$ observations.
*$p < .05$. **$p < .001$. ***$p < .0001$.

Table 4. Parameter Estimates and Standard Errors From Five Imputed Data Sets

	Variables							
	Injury		Workload		Staffing		Evening/Night Shift	
Data set	B	*SE*(B)	B	*SE*(B)	B	*SE*(B)	B	*SE*(B)
1	1.30	.36	1.18	.04	1.00	.08	.17	.05
2	1.07	.38	1.18	.05	1.02	.09	.20	.05
3	1.13	.37	1.18	.05	1.00	.08	.16	.05
4	1.05	.36	1.17	.04	1.04	.08	.18	.05
5	1.36	.37	1.17	.05	1.03	.08	.17	.05

estimates and the standard errors. The combined imputation results are displayed in Table 5. Comparing results of Table 5 with those of Table 2, MI yielded estimates, standard errors, and p values similar to those obtained before 25% of the workload and staffing data were removed. Additional variables could have been included in the model, but the correlations between other variables and workload and staffing were low.

MI proved appropriate in this exercise with a large data set and a moderate amount of missing information. The parameter estimates, standard errors, and test statistics generated by MI were very close to what was obtained with the complete Wave 2 data set. Listwise deletion, on the other hand, led to erroneous conclusions about one of the independent variables (note the nonsignificance of injury as a predictor variable in Table 3).

In addition to the superiority of the method as compared to listwise deletion, MI was simple to perform. Allison's two SAS macros contain explicit instructions for their basic use and adaptation to other SAS procedures, such as PROC REG, for ordinary least-squares regression, and PROC LOGISTIC, for logistic regression. The availability of software (i.e., NORM) and SAS macros for and ease of use of MI should encourage researchers to explore this statistical technique. In addition, the latest version of SAS has a procedure, PROC MI, with the necessary macros built into it.

Table 5. Imputation Results

	B	*SE* (B)
Injury	1.18*	.40
Workload	1.17***	.05
Staffing	1.02***	.09
Evening/night shift	.18***	.05

*$p < .05$, **$p < .001$, ***$p < .0001$.

CONCLUSION

Prior to any analysis, researchers must examine their data sets for the amount and pattern of missing data. Likewise, they must determine the best approach to handle the missing data, that is, one that allows the researcher to avoid incorrect analyses and the possible abandonment of potentially valuable data. MI is one of several methods that should be considered. Knowledge of this technique will better inform the researcher to make such decisions.

REFERENCES

Aiken, L.H., Lake, E.T., Sochalski, J., & Sloane, D. (1997). Design of an outcomes study on the organization of hospital AIDS care. Research in the Sociology of Health Care, 14(3), 2–26.

Allison, P.D. (1998). Multiple imputation for missing data: A cautionary tale. Retrieved January 10, 1999, from University of Pennsylvania Web site: http://www.ssc.upenn.edu/~allison

Allison, P.D. (2000). Missing data. Thousand Oaks, CA: Sage Publications.

Allison, P.D. (1999). Multiple regression: A primer. Thousand Oaks, CA: Pine Forge Press.

Barnard, J., & Meng, X. (1999). Application of multiple imputation in medical studies: From AIDS to NHANES. Statistical Methods in Medical Research, 8(1), 17–36.

Dempster, A.P., Laird, N., & Rubin, D.B. (1977). Maximum likelihood from incomplete data via the EM algorithm. Journal of the Royal Statistical Society, B39, 1–39.

Ferketich, S.L., & Verran, J.A. (1992). Analysis issues in outcomes research. In Patient outcomes research: Examining the effectiveness of nursing practice (pp. 174–188). Washington, DC: U.S. Department of Health and Human Services.

Freedman, V.A., & Wolf, D.A. (1995). A case study on the use of multiple imputation. Demography, 32, 459–470.

Heitjan, D.F. (1997). What can be done about missing data? Approaches to imputation. American Journal of Public Health, 87, 548–550.

Iversen, G.R. (1984). Bayesian statistical inference. (Sage University Paper series on Quantitative Applications in the Social Sciences, No. 07-043). Beverly Hills, CA: Sage.

Kennedy, P. (1992). A guide to econometrics (3rd ed.). Cambridge, MA: The MIT Press.

Landerman, L.R., Land, K.C., & Pieper, C.F. (1997). An empirical evaluation of the predictive mean matching method for imputing missing values. Sociological Methods & Research, 26, 3–33.

Little, R.J., & Rubin, D.B. (1989). The analysis of social science data with missing values. Sociological Methods & Research, 18, 292–326.

Reilly, M., & Pepe, M. (1997). The relationship between hot-deck multiple imputation and weighted likelihood. Statistics in Medicine, 16, 5–19.

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

Rubin, D.B. (1977). Formalizing notions about the effect of nonrespondents in sample surveys. Journal of the American Statistical Association, 72, 538–543.

Rubin, D.B. (1987). Multiple imputation for nonresponse in surveys. New York: Wiley.

Rubin, D.B. (1996). Multiple imputation after 18+ years. Journal of the American Statistical Association, 91, 473–515.

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

Schafer, J.L. (1998, March). The practice of multiple imputation. Paper presented at the meeting of the Methodology Center, Pennsylvania State University, University Park, PA.

Schafer, J.L. (1999). Multiple imputation: A primer. Statistical Methods in Medical Research, 8, 3–15.

Schafer, J.L. (2000). Analysis of incomplete multivariate data. Boca Raton, FL: Chapman & Hall/CRC. (Original work published 1997)

SPSS. (1999). SPSS Base 9.0 Applications Guide. Chicago: SPSS, Inc.

Tabachnick, B.G., & Fidell, L.S. (2000). Using multivariate statistics (4th ed.). New York: HarperCollins College Publishers.

APPENDIX

SAS Programming Steps Used in the Example Analysis[1]

data shift
set p.shiftden (keep = wave 2 facial injury workload staffing shiftime nurseid hospid40)
if Wave 2 = 1;
specifies data to be analyzed
if facial = . then delete;
excludes from analysis observations with missing dependent variable
run;
%include 'a:\miss.sas';
locates MISS macro
%miss(data = shift, var = facial injury workload staffing shiftime, out = impute, danum = 5);
invokes MISS macro; data = data set; var = variables used for the imputation; out = labels the output data set; danum = specifies the number of data sets to impute
proc genmod data = impute;
specifies SAS procedure to be used on the imputed data sets
class nurseid hospid40;
model facial = injury workload staffing shiftime /dist = mult link = clogit type3;
specifies model's dependent and independent variables; dist = specifies multinomial distribution; link = indicates cumulative logit model; type3 = requests Type III score statistics[2]
make 'parminfo' out = z;
labels output data set[2]
repeated subject = nurseid (hospid40)/type = ind;
invokes GEE, clustering by nurse and by unit, type = specifies correlation matrix[2] **by dsnum;**
provides 5 sets of estimates, one for each multiply imputed data set
run;
%include 'a:\combine.sas';
locates COMBINE macro
%combine(data = z, parname = injury workload staffing shiftime);
invokes COMBINE macro, data = specifies output data set, parmname = labels parameters[2]

[1]The SAS programming steps are in bold print. The author's comments and explanations are italicized.

[2]Depending on the SAS procedure used (e.g., PROC GENMOD versus PROC LOGISTIC or PROC REG), some commands will differ. Please see the text accompanying Allison's SAS macros for the correct procedures.