

Practice of Epidemiology

Multiple Imputation With Large Data Sets: A Case Study of the Children's Mental Health Initiative

Elizabeth A. Stuart, Melissa Azur, Constantine Frangakis, and Philip Leaf

Initially submitted August 4, 2008; accepted for publication January 21, 2009.

Multiple imputation is an effective method for dealing with missing data, and it is becoming increasingly common in many fields. However, the method is still relatively rarely used in epidemiology, perhaps in part because relatively few studies have looked at practical questions about how to implement multiple imputation in large data sets used for diverse purposes. This paper addresses this gap by focusing on the practicalities and diagnostics for multiple imputation in large data sets. It primarily discusses the method of multiple imputation by chained equations, which iterates through the data, imputing one variable at a time conditional on the others. Illustrative data were derived from 9,186 youths participating in the national evaluation of the Community Mental Health Services for Children and Their Families Program, a US federally funded program designed to develop and enhance community-based systems of care to meet the needs of children with serious emotional disturbances and their families. Multiple imputation was used to ensure that data analysis samples reflect the full population of youth participating in this program. This case study provides an illustration to assist researchers in implementing multiple imputation in their own data.

mental health services; missing at random; missing data; multiple imputation

Abbreviations: CMHI, Children's Mental Health Initiative; MAR, missing at random; MCAR, missing completely at random; MICE, multiple imputation by chained equations; NMAR, not missing at random.

Missing data is a problem in many studies, particularly in large epidemiologic studies in which it may be difficult to ensure that complete data are collected from all individuals. Multiple imputation is one technique becoming increasingly advocated to deal with missing data because of its improved performance over alternative approaches (1–4). However, multiple imputation is still rarely used in epidemiology (2), perhaps in part because relatively little practical guidance is available for implementing and evaluating multiple imputation techniques. This is particularly true for large data sets, such as are common in epidemiology. This paper presents a case study of imputation of data from the national evaluation of the Community Mental Health Services for Children and Their Families Program (i.e., the Children's Mental Health Initiative (CMHI)) and discusses some of the issues involved in conducting a large-scale imputation.

Multiple imputation is a powerful and flexible technique for dealing with missing data. Conceived by Rubin (5) and

described further by Little and Rubin (6) and Schafer (7), multiple imputation imputes each missing value multiple times. Inferences using the multiply imputed data thus account for the missing data and the uncertainty in the imputations. Multiple imputation is relatively easy to implement and is appropriate for a wide range of data sets. Two other general techniques for dealing with missing data are single imputation methods—such as mean imputation—and maximum likelihood approaches, which directly estimate parameters by accounting for the missingness. Multiple imputation offers distinct advantages over each of these alternatives. Single imputation does not “know” that some values have been imputed and treats all values as true values. It thus does not account for the uncertainty in the missing values and so underestimates the variance of estimates, leading to inflated type I error rates. Maximum likelihood techniques can be difficult to implement for complex or nonstandard models, although special-purpose software is

Correspondence to Dr. Elizabeth A. Stuart, Johns Hopkins Bloomberg School of Public Health, 624 North Broadway, 8th Floor, Baltimore, MD 21205 (e-mail: estuart@jhsph.edu).

making it more feasible (1). In contrast, multiple imputation techniques provide valid variance estimates and are easy to implement and describe.

Although multiple imputation is often used by individual researchers performing imputation for a particular analysis, this paper focuses on situations in which a large data set will be used by multiple researchers, for diverse purposes. For example, a research group may want to create one large, multiply imputed data set rather than have each researcher deal with the missing data individually (and perhaps differently; 8, 9). This paper builds on previous papers that provide an overview of multiple imputation (3, 10, 11) by focusing on the complications encountered in implementing multiple imputation for large data sets that will be used for a range of analyses.

THE CHILDREN'S MENTAL HEALTH INITIATIVE

We illustrate the methods by using a large data set from the CMHI, funded by the Substance Abuse and Mental Health Services Administration. Since 1993, the CMHI has funded communities to develop service systems to provide a comprehensive spectrum of mental health and support services to US children with mental illness (12). A national evaluation is currently collecting descriptive data on all children referred to the CMHI. More detailed data are gathered from youth and families who agree to participate in a longitudinal outcomes study (13).

Here, we focus on the sample of children who were referred to the CMHI between 1997 and 2005 and who agreed to participate in the longitudinal study ($N = 9,186$). These data potentially provide a wealth of information on children's mental health services, with longitudinal data on more than 9,000 children across 45 sites. However, substantial data are also missing. While basic demographics are available for approximately 99% of children in the study sample, most variables have missing values for 30%–70% of the children. The percentage of variables missing for each child ranges from 1% to nearly 100%, although 50% of the children have at least 87% of the variables observed and 75% have at least 73% of the variables observed. In this paper, we focus on the baseline data, with approximately 400 variables, including measures of behavior problems, family resources, and service receipt.

Many of the end users of the CMHI data are trained in psychology and other applied fields and likely have limited experience with methods for missing data. The goal of the imputation process described here was to create a general use, multiply imputed data set that a broad range of researchers could use to answer their research questions, similar to a situation that may be encountered by research groups that aim to create imputed data sets for all of their members to use.

MISSING DATA

Types of missing data

Nearly all studies have some missing data; the question is, How much of a problem is it? The answer depends on the mechanism that caused the missingness. There are 3 types of missing-data mechanisms (14): missing completely at ran-

dom (MCAR), missing at random (MAR), and not missing at random (NMAR).

MCAR occurs when the missingness is unrelated to the variables under study. In other words, the missingness is purely random, and the individuals with missing data are a simple random sample of the full sample. MAR means that the probability of an observation being missing may depend on observed values but not on unobserved values. Finally, NMAR means that the probability of missingness depends on both observed and unobserved values.

Most commonly, missing data are assumed to be MAR. The MCAR assumption is generally unrealistic, as can be observed in the data if the missingness is related to any of the observed characteristics. In these cases, MAR, while empirically unverifiable, is often a reasonable assumption to make unless substantive knowledge about the data or data collection process indicates that the missingness may depend on unobserved values. In that case, a NMAR model should be posited. The MAR assumption is also sometimes made more reasonable by including "auxiliary variables" that are related to the missingness but may not be of interest in the analyses themselves; in fact, this strategy can greatly improve the imputations (15).

Assessing missing data

Understanding when and why variables are missing is crucial. This step can be accomplished by calculating the rate of missingness for each variable as well as examining the patterns of missingness. Finally, the observed characteristics of individuals with observed and missing values for some key variables of interest should be compared.

In the CMHI baseline data, the percentage of missing values across variables ranges from 1% to 99%. The missingness is scattered, with no clear pattern, except that variability is large across sites. Across the 45 sites, the percentage of variables with more than 50% missing ranged from 20% to 69%. Sites were responsible for their own data collection, and this variability likely reflects variation in the time and resources available as well as possibly differences in the populations served.

In the CMHI data, the missingness on key variables is related to a number of observed characteristics of the children, confirming that the data are not MCAR. For example, children with a missing value on the internalizing problems scale were more likely to be eligible for Medicaid and to have conduct disorder and were less likely to be white or Hispanic (Table 1). Unfortunately, there is no way to empirically confirm that the data are not NMAR; however, for our purposes, we are comfortable making that assumption here.

MULTIPLE IMPUTATION

Multiple imputation has emerged as an appropriate and flexible way of handling missing data. Complete-case methods, which simply discard observations with any missing data, generally make the usually unrealistic assumption that the data are MCAR, or at least MAR within categories defined by the variables included in the analysis model (16). Some researchers avoid imputation approaches because of fears of "making up data." In fact, complete-case analyses require stronger assumptions than does imputation.

Table 1. Comparison of Characteristics of Children With Observed and Missing Values on the Internalizing Symptoms Scale

Characteristic	Children With Missing Internalizing Scale Values, %	Children With Observed Internalizing Scale Values, %	P Value (2-Sided)
American Indian	18.3	6.3	0.00
Caucasian	49.5	61.0	0.00
Hispanic	10.9	13.0	0.01
Conduct disorder	15.8	8.9	0.00
Eligible for Medicaid	74.5	69.2	0.00
ADHD	36.3	42.0	0.00
Currently receiving services	59.6	65.8	0.00
Parental history of substance use treatment	50.2	56.0	0.05
Parental history of psychiatric hospitalization	38.9	42.3	0.05
Convicted of a crime	38.5	33.0	0.05
Percentage of day spent in special education classes	35.3	36.4	0.76

Abbreviation: ADHD, attention deficit hyperactivity disorder.

Multiple imputation methods work by imputing (or filling in) the missing values with reasonable predictions multiple times. This step creates a set of “complete” data sets with no missing values. The analysis is then run separately on each data set, and the results are combined across data sets by using the multiple imputation combining rules (5). The resulting estimates account for both within- and between-imputation uncertainty, reflecting the fact that the imputed values are not the known true values. Doing so results in correct standard error estimates and coverage rates, as compared with single imputation methods or simply including a missing data indicator for each variable in the model (17, 18).

The original approaches to creating multiple imputations generally assumed a large, joint model for all of the variables, for example, multivariate normality (6, 7). More recently, a more flexible method called multiple imputation by chained equations (MICE) has been developed (19). MICE cycles through the variables, modeling each conditional on the others. The imputations themselves are predicted values from these regression models, with the appropriate random error included. The procedure is as follows: first, the variable with the least missingness (variable 1) is imputed conditional on all variables with no missingness. The variable with the second least missingness is then imputed conditional on the variables with no missing values and variable 1, and so on. After all of the variables have been cycled through in this way (one “iteration”), there are no longer any missing values in the data. This process is then repeated using this data set with no missing values.

Raghunathan et al. (20) recommend 10 iterations for each imputation. The idea is that, at the end of 10 iterations, the imputations should have stabilized such that the order in which variables were imputed no longer matters. The im-

puted values at the end of the 10th iteration, combined with the observed data, constitute one imputed data set. This entire process is then repeated to create multiple imputed data sets, such that, to create 10 complete data sets, a total of 10×10 iterations are performed.

A strength of MICE is that each variable can be modeled by using a model tailored to its distribution, such as Poisson, logistic, or Gaussian. MICE can also incorporate additional data challenges, such as bounds or variables defined for only a subset of the sample. Generally, 5–10 imputations are created, resulting in 5–10 “complete” data sets, although recent work has indicated that more imputations may be beneficial (21). For the CMHI, we created 10 imputed data sets. We implemented MICE by using the IVEWare package for SAS software (20); IVEWare is also available as a stand-alone package, and MICE packages also exist for Stata (22) and S-Plus (23) software.

Although MICE is very useful in practice, it does lack the theoretical justification of some other imputation approaches. In particular, a drawback of MICE is that fitting the series of conditional models does not necessarily imply a proper joint distribution, which could lead to inconsistencies across models, where, for example, the model for variable 2 given variable 1 may not be consistent with the model for variable 1 given variable 2. Initial research has indicated that this drawback is not generally an issue in applied problems (3, 24), but this is an area of ongoing statistical research. Another drawback, discussed further below, is the need to include many interactions to preserve associations in the data; for example, to preserve all 3-way interactions, all of the 2-way interactions must be included in all regression models, which is often not feasible.

Complications in implementing MICE

A number of complications are encountered when actually implementing MICE, particularly with large data sets. These complications include model selection and computing limitations. Ideally, the model for each variable to be imputed should fit the data well and be as general as possible, in the sense of including as many predictors and interactions as possible, as discussed above. In practice, this step is sometimes difficult to accomplish.

One strategy is to use stepwise selection to choose the model for each variable at each iteration (20). This process will include in the regression models those variables most predictive of the variable being imputed, for example, including a certain number of variables (those most predictive) or those leading to some minimum additional *R*-squared value (20). The exact model for each variable may change across iterations but should stabilize as the imputations themselves stabilize.

An important consideration in implementing MICE procedures is that the model used to create the imputations should be more general than the analysis model in terms of including all interactions that will be examined in the analyses (1, 8). This step will prevent the analyses from missing associations that actually exist. For example, if there is particular interest in the relation between gender and internalizing symptoms, then that relation should be included in the imputation model. In contrast, if the

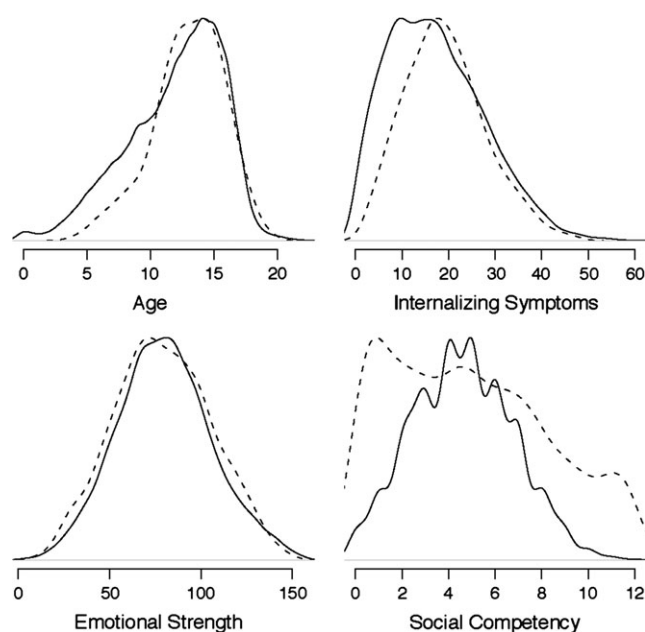


Figure 1. Comparison of observed and imputed values for 4 representative variables. For each variable, the solid line shows the density plot of observed values and the dashed line the density plot of imputed values. Age is expressed in years. The other measures are the Child Behavior Checklist (CBCL) internalizing syndrome score (33), the Behavioral and Emotional Rating Scale (34), and the CBCL total social competency score.

variables are assumed to be independent (i.e., gender is not used to impute internalizing symptoms), then the analysis may find a lack of relation simply because the imputations were generated by assuming there was none. This is not a large issue for bivariate associations because variables that have associations in the observed data will be selected by the stepwise selection procedures. However, for crucial 3-way interactions (e.g., between race, gender, and internalizing symptoms in the CMHI study), it is important to include the three 2-way interactions between those 3 variables as possible predictors.

Because of computational limitations, it was not possible to include a large number of interactions in the CMHI imputation models. There is particular interest in disparities in care and in interactions between race and gender with mental health needs and services. We thus included interactions between race, age, sex, income, and referral source as potential predictors. The stepwise selection models used a relatively liberal inclusion criterion of incorporating variables or interactions that added at least 0.01 to the *R*-squared value, resulting in approximately 6–10 variables in each regression model. Because interactions are included as if they are separate variables (i.e., IVEWare does not know that they are interactions), the models are not necessarily hierarchical in that the individual variables may not be selected even if their interaction is.

An additional issue with the CMHI data was how to handle sites. At one extreme, we could perform the imputations

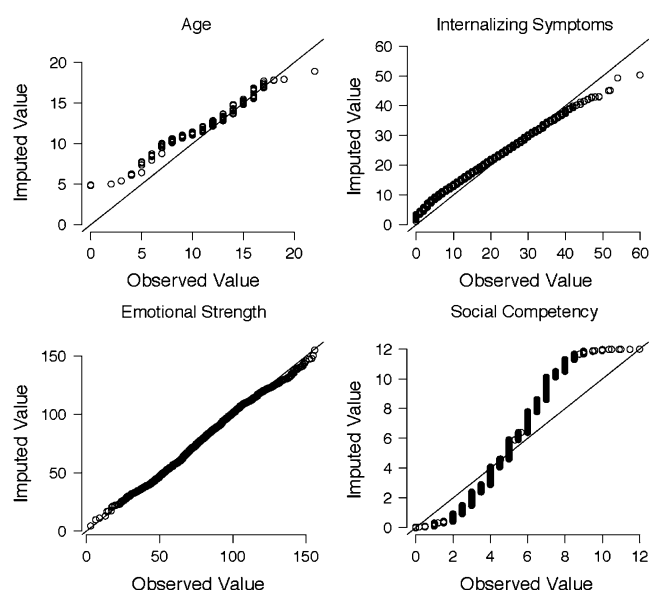


Figure 2. Comparison of observed and imputed values for 4 representative variables. For each variable, the quantile-quantile plot compares the distribution of observed values (x-axis) and of imputed values (y-axis). Age is expressed in years. The other measures are the Child Behavior Checklist (CBCL) internalizing syndrome score (33), the Behavioral and Emotional Rating Scale (34), and the CBCL total social competency score.

by completely ignoring the sites. Doing so assumes that the associations between variables are the same across sites. The other extreme would be to impute separately for each site, assuming completely different models within each site. We chose a middle route, recently recommended by Graham (1). This method treated the site indicators just as any other variable in that each site indicator could be selected by a stepwise model, if it was an important predictor of the variable under consideration. The imputation process did not otherwise account for the clustering within sites; however, any analysis using the multiply imputed data can (and should) account for that clustering.

Analyzing multiply imputed data

After the imputations are created, users have a set of complete data sets. To analyze the resulting data, the analysis is run separately within each complete data set, then the results are combined by using the multiple imputation combining rules (5, 25). A strength of multiple imputation is that researchers can run any model in the complete data sets, for example, a hierarchical linear model to account for clustering within sites in the CMHI data (20). The overall estimate is the average of the estimates from each of the complete data sets. The variance of that overall estimate is a function of the variance within each complete data set and the variance across the data sets. Many multiple imputation analysis functions available in common statistical software packages perform this combining for the user, making the analysis

Table 2. Bivariate Associations Between Predictors and Outcomes: Comparison of Complete-Case and Imputation Results

Outcome/ Independent Variable	Predictor/ Independent Variable	Complete-Case Coefficient (P Value)	Imputation Coefficient (P Value)	% Missing on Independent Variable
High emotional strength	Age at first smoking	−0.02 (0.00)	0.02 (0.19)	18
Functional impairment	Age at first smoking	0.03 (0.00)	−0.06 (0.04)	18
Clinical internalizing problems	Age at first smoking	−0.01 (0.08)	−0.07 (0.00)	18
Clinical internalizing problems	Ever smoked	−0.04 (0.46)	0.25 (0.00)	18
Clinical internalizing problems	Ever drank alcohol	−0.18 (0.00)	0.09 (0.13)	18

not much more difficult than analyses on a single, complete data set (11).

IMPUTATION DIAGNOSTICS

With these complex (and numerous) models, it is important to develop diagnostics to help determine when the imputations are reasonable. While some previous researchers have argued that it is impossible to diagnose imputations because, by definition, the missing values are unobserved, recent research has identified important diagnostics that can be undertaken (1, 26). In this section, we describe graphic and numeric diagnostics. Because these diagnostics may be overwhelming with large data sets, we emphasize identifying potentially problematic variables, which can then be examined more carefully.

One issue with examining imputation diagnostics is that differences between the observed and imputed values do not necessarily imply a problem. In fact, those differences may be precisely what the imputation is trying to address. For example, if younger children are more likely to be missing the internalizing behaviors scale and age is associated with internalizing behavior, then the distribution of internalizing behaviors scores is likely to be somewhat different in the observed and imputed data. Substantive knowledge (and knowledge of which types of individuals have missing data) is crucial to determine whether the imputations are in fact reasonable or whether the procedure needs to be modified.

Each of the diagnostics described below was run by using one of the 10 imputed data sets and was then repeated with a second to confirm that similar results were obtained, which they were. To assess the sensitivity of the imputations to a particular imputation approach, we also repeated the numeric diagnostics when comparing 2 sets of imputations: the original model and one in which we modified the stepwise selection criteria to choose more variables. Large deviations across imputation settings would indicate sensitivity to the particular models, which is undesirable.

Graphic diagnostics

A helpful first diagnostic is to compare the distributions of the observed and imputed values of each variable through

histograms, density plots, or quantile-quantile plots (26). By “imputed values” we mean only those values actually imputed, as opposed to all values in the imputed data sets (which includes both observed and imputed values). Figures 1 and 2 show density plots and quantile-quantile plots, respectively, for 4 variables in the CMHI data. The observed and imputed distributions are quite similar, especially for age (2% missing) and emotional strength (22% missing). For internalizing symptoms (19% missing), the imputed values tend to be slightly higher than the observed values; for social competency (27% missing), the imputations are more spread out than the observed values. These differences are likely a result of differences in the types of children with missing and observed values on those scales (Table 1).

Numeric diagnostics

With large numbers of variables, it may be difficult to carefully examine a histogram for each variable. Numeric diagnostics can summarize and highlight those variables that may be of concern. In particular, we identify all variables with large differences between the observed and imputed values, in particular 1) an absolute difference in means between the observed and imputed values greater than 2 standard deviations, or 2) a ratio of variances of the observed and imputed values that is less than 0.5 or greater than 2. Again, these differences may not indicate a problem, but they provide a way to quickly determine which variables should be further investigated. Another approach is to flag any variable with a significant Kolmogorov-Smirnov test when comparing the observed and imputed values (26).

In the CMHI data, a few variables showed potential problems in terms of having different distributions of observed and imputed values, primarily those that assess use of illegal substances. Our conclusion regarding those variables was that, because of their rarity, the observed data did not provide sufficient information to reliably impute the missing values; thus, any analyses using those variables should be conducted with caution. The density plots also indicated a few variables for which extreme outlying values were obtained in the imputation process. Simplifying the imputation models for those variables solved that problem. When comparing imputations across 2 different imputation models, we found very few differences, which is consistent with

results showing that analyses using multiply imputed data are generally not very sensitive to whether the “correct” imputation model is used (1, 25).

Assessing the impact of imputation

A question of primary interest to many applied researchers is how performing multiple imputation may affect the conclusions of their study. In some cases, very similar results may be obtained; in others, the conclusions may be very different (27, 28). Davey et al. (29) show a complete-case analysis that implies that spending more time in poverty increases the psychosocial adjustment of African-American children. After multiple imputation, the more credible finding that duration of poverty has detrimental effects for all children (of every race) is revealed. Compared with simpler approaches such as complete-case analysis, multiple imputation techniques will generally provide more accurate estimates of associations in the data (5, 17, 30).

In the CMHI data, we found a number of differences when comparing complete-case analyses and analyses that use the multiply imputed data. Table 2 summarizes some of these differences. For example, the complete-case analysis implies that later initiation of smoking is negatively associated with emotional strength and positively associated with functional impairment. However, the imputed data reveal the more expected associations: that starting smoking later is not associated with emotional strength and is negatively associated with both functional impairment and internalizing problems. Similarly, the complete-case analysis implies that drinking alcohol is negatively associated with internalizing problems; after imputation, that relation is insignificant and there is evidence of a significant positive association between smoking and internalizing problems.

CONCLUSIONS AND LESSONS LEARNED

Multiple imputation involves more work than many of the other methods of dealing with missing data. So, why should a researcher invest all of this time and effort? The primary reason is to obtain more accurate results regarding the associations of interest. It also prevents a loss in power from having to exclude any observation with a missing value for even just one of the variables used in the analysis. Finally, multiple imputation improves the generalizability of the results to the population sampled by allowing inclusion of all individuals in analyses.

We have focused on a setting in which the multiply imputed data set will be used for a variety of analyses and there is interest in having consistency in the data used. In conducting this complex imputation process using the CMHI data, we also learned a few lessons regarding implementing multiple imputation approaches with large data sets.

1. Start small. Begin by building the imputation model and setting up the imputation code on a subset of the data set (both a subset of the variables and a subset of the observations). This procedure makes it easier to identify problems with the imputation process.

2. Provide clear documentation for data users. Describe how multiple imputation works, how the imputations were created, any problems encountered in that process (such as variables for which the imputations may be problematic), and how multiply imputed data should be analyzed.

There are also a number of important areas for further methodological research. First, imputation diagnostics need to be further developed and incorporated into the common multiple imputation software packages. One limitation of the diagnostics we present is that they are primarily univariate and do not account for the interrelationships of variables in the data. More complex diagnostics, such as looking at residuals from regression models, could help determine which differences are reasonable. These types of diagnostics are currently under development in the statistical literature (26) and are an important area for future research. Second, given the time and energy required to implement and use multiple imputation, it is important for researchers to understand when these methods are necessary. For example, if missingness rates are relatively low, is there still a benefit to conducting multiple imputation? Can diagnostics be performed to determine when it is especially crucial to use these methods? Are there ways to determine when there is not enough information in the data to conduct reliable imputations? Third, we have shown the potential value in using the MICE procedure for large data sets. However, the method is still relatively young, and further research into its performance and limitations is needed. For example, more work should examine the consequences of not having a proper joint distribution, as well as complexities such as clustering or longitudinal data (31, 32).

In conclusion, multiple imputation is an effective, flexible, and relatively easy way to deal with missing data. This paper has outlined the use of multiple imputation, focusing on the complexities encountered in large data sets, with the aim of helping researchers use multiple imputation in their own investigations.

ACKNOWLEDGMENTS

Author affiliations: Department of Biostatistics, Johns Hopkins Bloomberg School of Public Health, Baltimore, Maryland (Elizabeth A. Stuart, Constantine Frangakis); and Department of Mental Health, Johns Hopkins Bloomberg School of Public Health, Baltimore, Maryland (Elizabeth A. Stuart, Melissa Azur, Philip Leaf).

This work was supported by the National Institute of Mental Health (grant 1R01MH075828 to P. L. and grant 1K25MH083846 to E. A. S.).

Conflict of interest: none declared.

REFERENCES

1. Graham JW. Missing data analysis: making it work in the real world. *Annu Rev Psychol.* 2009;60:549–576.

2. Klebanoff MA, Cole SR. Use of multiple imputation in the epidemiologic literature. *Am J Epidemiol*. 2008;168(4):355–357.
3. Schafer JL, Graham JW. Missing data: our view of the state of the art. *Psychol Methods*. 2002;7(2):147–177.
4. Shrive FM, Stuart H, Quan H, et al. Dealing with missing data in a multi-question depression scale: a comparison of imputation methods [electronic article]. *BMC Med Res Methodol*. 2006;6:57.
5. Rubin DB. *Multiple Imputation for Nonresponse in Surveys*. New York, NY: John Wiley & Sons; 1987.
6. Little RJA, Rubin DB. *Statistical Analysis With Missing Data*. 2nd ed. Hoboken, NJ: Wiley Interscience; 2002.
7. Schafer JL. *Analysis of Incomplete Multivariate Data*. London, United Kingdom: Chapman & Hall; 1997.
8. Barnard J, Meng XL. Applications of multiple imputation in medical studies: from AIDS to NHANES. *Stat Methods Med Res*. 1999;8(1):17–36.
9. Ezzati-Rice TM, Johnson W, Khare M, et al. A simulation study to evaluate the performance of model-based multiple imputations in NCHS Health Examination Surveys. *Proceedings of the Bureau of the Census Eleventh Annual Research Conference* 1995;257–266.
10. Harel O, Zhou XH. Multiple imputation: review of theory, implementation, and software. *Stat Med*. 2007;26(16):3057–3077.
11. Horton N, Kleinman KP. Much ado about nothing: a comparison of missing data methods and software to fit incomplete data regression models. *Am Stat*. 2007;61(1):79–90.
12. Macro International Inc. *Comprehensive Community Mental Health Services for Children and Their Families Program, Evaluation Findings—Annual Report to Congress, 2002–2003*. Rockville, MD: Center for Mental Health Services, Substance Abuse and Mental Health Services Administration; 2007. (No. DHHS SMA-CB-E2002/03).
13. Holden E, Friedman R, Santiago R. Overview of the national evaluation of the Comprehensive Community Mental Health Services for children and their families. *J Emot Behav Disord*. 2001;9:4–12.
14. Rubin DB. Inference and missing data. *Biometrika*. 1976;63:581–592.
15. Collins LM, Schafer JL, Kam CK. A comparison of inclusive and restrictive strategies in modern missing-data procedures. *Psychol Methods*. 2001;6(4):330–351.
16. Davey A. Comment on chapters 11 and 12: an analysis of incomplete data. In: Collins LM, Sayer AG, eds. *New Methods for the Analysis of Change*. Washington, DC: American Psychological Association; 2001:379–384.
17. Greenland S, Finkle WD. A critical look at methods for handling missing covariates in epidemiologic regression analyses. *Am J Epidemiol*. 1995;142(12):1255–1264.
18. Vach W, Blettner M. Biased estimation of the odds ratio in case-control studies due to the use of ad hoc methods of correcting for missing values for confounding variables. *Am J Epidemiol*. 1991;134(8):895–907.
19. Raghunathan TE, Lepkowski JM, Van Hoewyk J, et al. A multivariate technique for multiply imputing missing values using a sequence of regression models. *Surv Methodol*. 2001;27:85–95.
20. Raghunathan TE, Solenberger PW, Van Hoewyk J. *IVEware: Imputation and Variance Estimation Software*. Ann Arbor, MI: Survey Methodology Program, Survey Research Center, Institute for Social Research, University of Michigan; 2007. (<http://www.isr.umich.edu/src/smp/ive/>). (Accessed December 20, 2008).
21. Graham JW, Olchowski AE, Gilreath TD. How many imputations are really needed? Some practical clarifications of multiple imputation theory. *Prev Sci*. 2007;8(3):206–213.
22. Carlin JB, Galati JC, Royston P. A new framework for managing and analyzing multiply imputed data in Stata. *Stata J*. 2008;8(1):49–67.
23. van Buuren S, Oudshoorn CGM. *Multivariate Imputation by Chained Equations: MICE V1.0 User's Manual*. The Netherlands: TNO Prevention and Health; 2000. (TNO report PG/VGZ/00.038). (<http://web.inter.nl.net/users/S.van.Buuren/mi/docs/Manual.pdf>). (Accessed December 20, 2008).
24. Brand JP. *Development, Implementation and Evaluation of Multiple Imputation Strategies for the Statistical Analysis of Incomplete Data Sets* [doctoral dissertation]. Rotterdam, The Netherlands: Erasmus University; 1999.
25. Schafer JL, Olsen MK. Multiple imputation for multivariate missing-data problems: a data analyst's perspective. *Multivariate Behav Res*. 1998;33:545–571.
26. Abayomi K, Gelman A, Levy M. Diagnostics for multivariate imputations. *Appl Stat*. 2008;57(3):273–291.
27. Cheng AL, Lin H, Kaspro W, et al. Impact of supported housing on clinical outcomes: applied analysis of a randomized trial using multiple imputation technique. *J Nerv Ment Dis*. 2007;195(1):83–88.
28. Hill JL, Waldfogel J, Brooks-Gunn J, et al. Maternal employment and child development: a fresh look using newer methods. *Dev Psychol*. 2005;41(6):833–850.
29. Davey A, Shanahan MJ, Schafer JL. Correcting for selective nonresponse in the National Longitudinal Survey of Youth using multiple imputation. *J Hum Resources*. 2001;36(3):500–519.
30. Rubin DB, Schenker N. Multiple imputation for interval estimation from simple random samples with ignorable nonresponse. *J Am Stat Assoc*. 1986;81:366–374.
31. Liu M, Taylor JM, Belin TR. Multiple imputation and posterior simulation for multivariate missing data in longitudinal studies. *Biometrics*. 2000;56(4):1157–1163.
32. Schafer JL, Yucel RM. Computational strategies for multivariate linear mixed-effects models with missing values. *J Comput Graph Stat*. 2002;11:437–457.
33. Achenbach TM. *Manual for the Child Behavior Checklist/4-18 and 1991 Profile*. Burlington, VT: Department of Psychiatry, University of Vermont; 1991.
34. Epstein MH. The development and validation of a scale to assess the emotional and behavioral strengths of children and adolescents. *Remedial Special Ed*. 1999;20(5):258–262.