MULTIPLE IMPUTATION OF NHANES III

Meena Khare, Roderick J. A. Little, Donald B. Rubin and Joseph L. Schafer *
Joseph L. Schafer, Department of Statistics, 325 Classroom Building, University Park, PA 16802

Key Words: Bayesian simulation, Gibbs sampling, missing data, sample surveys.

1 INTRODUCTION

The National Health and Nutrition Examination Survey (NHANES), conducted by the National Center for Health Statistics (NCHS), collects important nutritional and health-related data on the civilian noninstitutionalized U.S. population and important subgroups. The ongoing NHANES III consists of two national samples collected over a six-year period (1988–94) with a total sample size of approximately 40,000 persons. High rates of unit nonresponse in NHANES, together with some residual item nonresponse, lead to high rates of missingness on key survey variables. In previous NHANES surveys, unit nonresponse was handled by weighting-class adjustments, with little or no compensation made for item nonresponse. Similar weighting-class adjustments were also planned for NHANES III (Ezzati and Khare, 1992).

In 1992, NCHS initiated a project to investigate alternatives to the current NHANES nonresponse adjustment methodology. One of these alternatives was multiple imputation (MI) (Rubin 1987). In this paper, we describe our preliminary efforts to multiply impute a portion of the data from NHANES III, Phase 1 (1988–91). A data file consisting of 27 key variables for 12,392 sampled adults was multiply imputed for both item and unit nonresponse using techniques of iterative Bayesian simulation via Markov chains described by Schafer (1991). Exploratory analysis of the imputed values suggests that both the marginal distributions of variables and important relationships between them were accurately preserved. MI interval estimates for scalar quantities of interest (means, subdomain means, etc.) were, in some cases, dramatically wider than corresponding intervals that ignored the missing-data uncertainty. This project represents the first successful implementation of proper MI methodology in a large multivariate survey, and consequently gives useful insight into the feasibility of multiply imputing NHANES and other large multipurpose sample surveys on an ongoing basis.

*Meena Khare, National Center for Health Statistics. Roderick J. A. Little, University of Michigan. Donald B. Rubin, Harvard University. Joseph L. Schafer, Pennsylvania State University.

2 NONRESPONSE IN NHANES III

Patterns of nonresponse in NHANES III are heavily influenced by the process of data collection, which occurred primarily in three stages:

1. *Household screening.* When a household was selected into the NHANES III sample, a brief screening interview was conducted to determine household size and the age, sex, and race of every household member. This information was required for the final stage of sampling in which individuals were selected within households. As a byproduct of this screening procedure, the basic demographic characteristics—age, sex, and race—are known for each sampled person; no data are missing for these items.
2. *Personal interview.* After the household screening and final stage of sampling, in-depth interviews were conducted. In Phase 1 of NHANES III, 14% of the sampled persons could not be interviewed at all. Among those that were interviewed, refusal or inability to answer specific interview questions led to some additional item nonresponse at typical rates of 1–5% per item.
3. *MEC examination.* Upon completion of the personal interview, sampled persons were requested to report to the MEC for the physical examination. Among interviewed persons in Phase 1, 9% did not report to the subsequent physical examination, resulting in an overall examination rate among sampled persons of 78%. In the examinations, not all items were successfully recorded for all examinees. Data recording errors and other mistakes by personnel also caused single items or groups of items to be missing. Among examinees, nonresponse rates for individual MEC items were on the order of 5–8%.

At the end of this data collection process, many key variables from the MEC examination were missing at rates relative to the entire sample of 30% or more.

It is common practice to compensate for unit nonresponse by weighting-class adjustments. Respondents and nonrespondents are grouped together into a relatively small number of classes or cells. The nonrespondents are assigned survey weights of zero, and the weights of the remaining respondents are proportionately inflated so that the total weight

of the units within cells is preserved. In previous NHANES surveys, survey weights for examinees were inflated up to the level of the full sample within broad classes defined by geography, demographic variables, and family income. After reweighting, little or no imputation was used to compensate for the remaining item nonresponse; missing items were left blank and omitted from further analyses.

One disadvantage of this weighting-class approach is that very little information obtained from the personal interview was used in the nonresponse adjustment for the non-examined. With the exception of income, none of the interview information was used in the formation of weighting classes. Yet, the personal interview provided many variables that are potentially powerful predictors for some of the MEC examination items. There seem to be significant potential gains, both in reducing nonresponse bias and increasing precision, from including more of the interview variables in the nonresponse adjustment.

Imputation, although typically more difficult to carry out in practice than weighting-class adjustments, offers some potentially important advantages including the reduction of variance and the opportunity to use more covariate information. Moreover, through the technique of multiple imputation (Rubin, 1987), it is possible to assess the impact of missing-data uncertainty on the variances of estimators and revise variance estimates to reflect this additional uncertainty. Applications of multiple imputation to large surveys such as NHANES have been previously hampered by the difficulty of generating proper multiple imputations in multivariate settings. Recent advances in techniques of Bayesian computation, however, now make it possible to generate proper multiple imputations in multivariate settings under a variety of useful models for both continuous and categorical data (Schafer, 1991). Multiple imputations can now be routinely generated using iterative simulation schemes based on Markov chains, including the Gibbs sampler and the Metropolis algorithm (Geman and Geman, 1984; Gelfand and Smith, 1990).

3 THE MI DATA FILE

To test the applicability of MI methodology to NHANES, a data file was prepared consisting of approximately 30 key variables from the screener questionnaire, personal interview, and MEC examination in Phase 1 of NHANES III. Major attention focused on just twelve variables from three MEC components—body measurements, blood pressure, and lipids. Because the personal interview and examination procedures were substantially different for adults and children, this study was restricted to the 12,391 adults (age 17 years and older) in the Phase 1 sample. Of these, 72.5% completed both the interview and the examination, 12.8% were interviewed but not examined, and 14.9% were neither interviewed nor examined. Everyone who missed the interview also missed the examination. The twelve MEC variables with their overall rates of missingness reflecting both unit and item nonresponse are listed in Table 1.

Table 1: MEC variables in the NHANES III multiple imputation data file.

Name	*% missing*	*Description*
Body measurements		
HT	31.5	height
WT	33.2	weight
WST	32.9	waist circumference
BUT	32.8	buttocks circumference
Blood pressure		
BP1K1D	33.2	first systolic pressure
BP1K5D	33.3	first diastolic pressure
BP2K1D	33.4	second systolic pressure
BP2K5D	33.5	second diastolic pressure
BP3K1D	33.4	third systolic pressure
BP3K5D	33.6	third diastolic pressure
Lipids		
TCRES	33.3	total serum cholesterol
HDRES	33.9	HDL cholesterol

In addition to these twelve MEC variables, we included in the data file a number of additional variables from the screening and personal interviews. These pre-MEC variables were judged to contain potentially valuable information for imputing the missing MEC items. Imputation methods that condition on auxiliary variables have some well known benefits, including the reduction of nonresponse bias and sampling variance. For reducing mean squared error of prediction, it is usually beneficial to make maximal use of whatever covariate information is available. Another important, but less well known, benefit of including auxiliary variables arises when attempting to reflect missing-data uncertainty. MI will provide valid inferences only if the imputations exhibit enough variability to represent our true state of knowledge about the missing values in a conditional or *a posteriori* Bayesian sense. Omitting an auxiliary variable from the imputation procedure is equivalent to specifying, say, a regression model in which the coefficient of the auxiliary variable is set to zero *a priori*. Fixing parameters of the imputation model to zero, when the data do not provide strong evidence that they are truly zero, will tend to produce MIs having too little variability.

One set of auxiliary variables that requires careful consideration is the set that conveys information about the sample design. Surveys with complex sampling plans have important features—unequal probabilities of selection, stratification, and clustering—that distinguish them from simple random samples. The observational units in complex surveys are typically not exchangeable and cannot be appropriately described by simple probability models that assume units are independent and identically distributed. In order to guarantee that MI inferences are valid, essential information about the sample design must be included in the analysis.

Finally, apart from considerations of mean squared error and variance estimation, we also felt it was essential to include auxiliary variables to preserve important statistical relationships in the dataset, especially those relationships that may be of interest to potential secondary users of the data.

All of the above arguments tend to favor large imputation models over small ones, encouraging us to include as many auxiliary variables as possible. Balanced against these considerations were computational limitations that prevented us from fitting a model as large as we would have liked. Starting with about twenty candidate auxiliary variables, we reduced the list, on the basis of exploratory regression analyses (described below) and *a priori* considerations, to the fifteen variables shown in Table 2. Information on the stratified cluster design of NHANES was reflected in STAND, a 44-level categorical variable indicating the mobile examination location or primary sampling unit (PSU) to which a person belonged. Further information pertinent to the final stage of sampling was contained in the demographic variables AGE, SEX, and RACE. Eleven variables from the personal interview were included because they were found to have statistically significant and scientifically important relationships to one or more of the twelve MEC items.

Table 2: Pre-MEC auxiliary variables in the NHANES III multiple imputation data file.

Name	*% missing*	*Description*
PSU identifier		
STAND	0.0	exam location (1–44)
Demographics		
AGE	0.0	17–39, 40–59, 60+
SEX	0.0	male, female
RACE	0.0	Black, Mex-Amer, Other
Interview		
ACTV	20.6	activity status
AD1	18.5	diabetes diagnosed?
AE2	19.3	hypertension diagnosed?
AE7	62.4	high cholesterol diagnosed?
AF10	20.6	heart attack diagnosed?
AR3	18.3	smoke cigarettes now?
ALCO	18.6	beer/wine/liquor?
AHT	22.7	self-reported height
AWT	21.6	self-reported weight
ASYS	21.9	interview systolic b.p.
ADIAS	21.9	interview diastolic b.p.

4 THE IMPUTATION MODEL

The most straightforward way to generate proper multiple imputations in a multivariate setting is to specify a parametric model for the complete data along with a prior distribution for the parameters, and then simulate values from the conditional distribution of the missing data given the observed data. We chose to work with a special case of the model for mixed continuous and categorical multivariate data introduced for discriminant analysis by Krzanowski (1982) and applied to incomplete multivariate data by Little and Schluchter (1985).

Let Y denote the matrix of complete data, which can be partitioned as $Y = (W, Z)$, where W is an $n \times p$ matrix of categorical variables and Z is an $n \times q$ matrix of continuous variables. Let $W_1, W_2, \ldots, W_p$ and $Z_1, Z_2, \ldots, Z_q$ denote the variables in W and Z, respectively. Suppose that the categorical variable W_j takes d_j possible levels, so that each row of W can be classified into a cell of a p-dimensional contingency table with total number of cells equal to $D = \prod_{j=1}^{p} d_j$. Let $\{x_{ijk\cdots t}\}$ denote the cell counts of this contingency table, where $x_{ijk\cdots t}$ is the number of rows of W for which $W_1 = i, W_2 = j, \ldots, W_p = t$. It is notationally convenient to index the cells of the contingency table by the single subscript d, ranging from 1 to D, so that the cell frequencies may be written as $\{x_d\}$.

The multivariate distribution for Y is most easily described in terms of the marginal distribution of W and the conditional distribution of Z given W. We assume that the marginal distribution of W is multinomial on the cell counts $\{x_{ijk\cdots t}\}$, with cell probabilities denoted by $\pi = \{\pi_{ijk\cdots t}\} = \{\pi_d\}$. Conditionally upon W, the rows of Z are assumed to be multivariate normal with means that vary between cells of the contingency table, but with a common covariance structure for all cells. Given that an individual's categorical variables determine that he should be placed into cell d, then his values of $(Z_1, Z_2, \ldots, Z_q)$ are assumed to be $N(\mu_d, \Sigma)$ independently of all other individuals. Letting $\mu = (\mu_1, \mu_2, \ldots, \mu_D)^T$ denote the $D \times q$ matrix of conditional means, we can write the unknown parameters as $\theta = (\pi, \mu, \Sigma)$.

As the number p of categorical variables grows, the contingency table typically becomes too sparse to estimate the probabilities π_d for the individual

cells, much less the mean vectors μ_d within cells. For this reason, we reduce the dimensionality of the parameter by allowing loglinear constraints on the cell probabilities π, and/or ANOVA-like constraints on the cell means μ. Loglinear constraints are a well known device for fitting parsimonious models to contingency tables (e.g., Bishop Fienberg, and Holland, 1975) and will not be described here. Let A be a $D \times r$ design matrix that relates the within-cell means μ to an $r \times q$ matrix of regression coefficients β in the manner $\mu = A\beta$, where $\text{rank}(A) = r \leq D$. In other words, we allow the means μ_d to vary from cell to cell, but require that each column of μ lie in the r-dimensional linear space spanned by the columns of A.

Among the 27 variables listed in Tables 1 and 2, sixteen are continuous while the remaining eleven consist of ordered or unordered categories. Attempts to fit a model with eleven categorical variables proved futile, because the contingency table became much too sparse to allow for stable estimation of the within-cell means unless undesirably strong restrictions were introduced on μ through the design matrix A. Further elimination of categorical variables to reduce the dimensionality of the contingency table was undesirable, because we considered all eleven to be important. In particular, retention of the 44-level classification by STAND, even though this variable was one of the main causes of sparseness, was considered essential to ensure that sample-design information was properly reflected in the MIs.

After considering several alternatives, we finally decided to retain only four variables—AGE, SEX, RACE, and STAND—in the categorical portion of the model, treating the other 23 variables as continuous and conditionally multivariate normal given these four. The resulting four-way table had 792 cells for 12,392 observations. Because of the sample design, this table was filled in rather nicely with only 157 empty cells. Modeling the six dichotomous variables AD1, AE2, AE7, AF10, AR3, and ALCO, and the three-point ordinal variable ACTV, as continuous and conditionally normal was only a very rough approximation at best. We considered the approximation to be acceptable, however, because these seven variables were not among the variables of primary interest in our study. The variables of greatest interest were the twelve MEC variables listed in Table 1, and the pre-MEC variables were intended to serve primarily as predictors. Moreover, some limited evidence suggests that erroneously modeling the seven discrete variables as continuous did not have a strong adverse effect on the final imputations; when the continuous imputes for these seven were rounded off to the nearest categories, the distributions of the imputed values were quite reasonable and looked very similar to the distributions actually observed in the sample.

In the final analysis, we modeled the 635 nonempty cells of the contingency table for AGE × SEX × RACE × STAND by a saturated multinomial distribution, treating the 157 empty cells as structural zeros. (This specification had no effect on the distribution of imputed values, because these four variables were never missing.) The 23 remaining "continuous" variables were then modeled as a multivariate normal linear regression. To make the normality assumption more plausible, body measurements, lipids, and systolic blood pressures were expressed on a log scale. Each of the 23 individual regressions included an intercept, 17 dummy indicators to represent the full AGE × SEX × RACE interaction, and 43 dummy indicators to represent STAND, for a total of $23(1 + 17 + 43) = 1403$ estimated regression coefficients and $23(24)/2 = 276$ residual variances and covariances. The total number of unknown free parameters in this model was thus $(635 - 1) + 1403 + 276 = 2313$.

5 MODEL FITTING AND IMPUTATION

A full description of the techniques we used to fit our model and generate MIs is beyond the scope of this article and can be found in Schafer (1991); we present only the general strategy. An EM algorithm for maximum-likelihood (ML) estimation with incomplete data under the multivariate model described above is given by Little and Schluchter (1985). To simulate missing data under an assumed value of the parameter such as $\theta = \hat{\theta}$ (the ML estimate) would be relatively straightforward. Under our model, the vector of missing observations for each person has, given his or her observed data, a multivariate normal distribution with parameters that can be calculated by applying a suitable transformation to θ. Multiple simulated versions of the missing data under $\theta = \hat{\theta}$, however, would not be proper MIs because they would ignore uncertainty about θ. Proper MIs can be most easily conceptualized as repeated draws from a Bayesian posterior predictive distribution for the missing data given the observed data. Let Y_{obs} denote the observed data and Y_{mis} the missing data. The posterior predictive density of Y_{mis} given Y_{obs}, or $P(Y_{mis} \mid Y_{obs})$, is

$$P(Y_{mis} \mid Y_{obs}) = \int P(Y_{mis} \mid Y_{obs}, \theta)\, P(\theta \mid Y_{obs})\, d\theta,$$

where $P(\theta \mid Y_{obs})$ is the posterior density of the parameters given the observed data. This distribution $P(Y_{mis} \mid Y_{obs})$ is the appropriate source of multiple imputations under the assumption that the nonresponse mechanism is *ignorable*, or that the missing data are *missing at random*, in the sense defined by Rubin (1976, 1987).

Because $P(Y_{mis} \mid Y_{obs})$ has an intractable form, we simulated draws from this distribution indirectly by constructing a Markov chain for which the limiting distribution is $P(Y_{mis} \mid Y_{obs})$. This Markov chain is defined as follows. Given a current parameter value $\theta^{(t)}$, we simulated a value of Y_{mis} conditionally upon $\theta = \theta^{(t)}$:

$$Y_{mis}^{(t+1)} \sim P(Y_{mis} \mid Y_{obs}, \theta^{(t)}). \qquad (1)$$

Then we simulated a new parameter value under a complete-data posterior assuming $Y_{mis} = Y_{mis}^{(t+1)}$:

$$\theta^{(t+1)} \sim P(\theta \mid Y_{obs}, Y_{mis}^{(t+1)}). \qquad (2)$$

Performing (1) and (2) alternately beginning from some starting value $\theta^{(0)}$ defines a Markov chain. This algorithm is a special case of the Gibbs sampler (Geman and Geman, 1984; Gelfand and Smith, 1990), and it can be shown that under very general conditions the distribution of $\theta^{(t)}$ approaches $P(\theta \mid Y_{obs})$ as $t \to \infty$. By taking t suitably large, $Y_{mis}^{(t)}$ becomes essentially a draw from $P(Y_{mis} \mid Y_{obs})$. Successive tth iterates $Y_{mis}^{(t)}, Y_{mis}^{(2t)}, Y_{mis}^{(3t)}, \ldots$ are essentially proper MIs.

In our analysis, we used the improper prior distribution $p(\pi, \beta, \Sigma) \propto \mid \Sigma \mid^{-(q+1)/2}$. Under this prior, the complete-data posterior distribution becomes the product of a Dirichlet for π, an inverted Wishart for Σ, and a matric-variate normal for β given Σ, all of which are straightforward to simulate (Schafer, 1991).

Beginning our simulation at the ML estimate $\hat{\theta}$, we simulated 400 iterations of the Markov chain. Every 40th value of Y_{mis} in the Markov chain was taken to be an independent draw from the stationary distribution. In this way, ten sets of imputations, which we shall call $\mathrm{MI}_1, \mathrm{MI}_2, \ldots, \mathrm{MI}_{10}$, were produced. The entire simulation took approximately 30 hours on a dedicated Sun SPARCstation ELC. The sequence $\mathrm{MI}_1, \mathrm{MI}_2, \ldots, \mathrm{MI}_{10}$ can be considered proper MIs only if the Markov chain achieves approximate stationarity (independence of the starting value) by 40 steps. Convergence to stationarity is difficult to assess, especially because of the high dimensionality of Y_{mis} and θ, but we informally monitored convergence by inspecting time-series plots of a few selected scalar functions of the parameter.

6 ANALYSIS OF THE IMPUTED DATA

A good imputation method should accurately preserve both the marginal distributions of the variables involved and relationships between them. Graphical displays (histograms, scatterplots, etc.) revealed that the marginal and bivariate distributions of the imputed data, for the most part, mimicked the observed data quite well. This was true both for the entire sample and within demographic subclasses. In some cases, certain non-normal features of the observed data (e.g. skewness and outliers) were not fully reproduced in the imputed values. The imputation model relied on assumptions of multivariate normality, and even after transformation the observed data were not entirely normal. For the most part, however, the imputed values appeared entirely plausible and consistent with the observed data, and secondary users of the data would not find gross anomalies that sometimes occur when ad hoc imputation schemes (e.g. hot-deck methods) are employed. Some of our graphical displays are reproduced in Schafer *et. al* (1993).

Using techniques described by Rubin (1987) for MI inference about scalar estimands, we calculated standard errors and interval estimates for a number of quantities of interest based on the ten imputations MI_1–MI_{10}. Let Q denote a scalar quantity to be estimated. Let $\hat{Q}_i$ and U_i denote a complete-data point estimate and variance estimate for Q, respectively, calculated from the ith imputed dataset, $i = 1, 2, \ldots, m$. The point estimate for Q is $\bar{Q}_m = m^{-1} \sum_{j=1}^{m} \hat{Q}_j$, the average of the m complete-data point estimates. The variance estimate associated with $\bar{Q}_m$ has two components. The within-imputation component is $\bar{U}_m = m^{-1} \sum_{j=1}^{m} U_j$, the average of the complete-data variance estimates, and the between-imputation component is $B_m = (m-1)^{-1} \sum_{j=1}^{m} (\hat{Q}_j - \bar{Q}_m)^2$, the sample variance of the m point estimates. The total variance is $T_m = \bar{U}_m + (1 + m^{-1}) B_m$. A $100(1-\alpha)\%$ interval estimate is formed by taking $\bar{Q}_m \pm t_\nu (1 - \alpha/2)\, T_m^{1/2}$, where $t_\nu(p)$ denotes the $100p$th percentile of the t distribution with ν degrees of freedom. An expression for ν is given by Rubin (1987), along with expressions for the relative increase in variance due to nonresponse r_m and the fraction of missing information γ.

Results for means of six MEC variables within categories of race/ethnicity are summarized in Table 3. The complete-data point estimates Q_i were

Table 3: MI estimates for means of six MEC variables within categories of race/ethnicity.

	$\bar{Q}_{10}$	$\bar{U}_{10}^{1/2}$	$T_{10}^{1/2}$	$100r_{10}$	100γ
Height (cm)					
White/other	168.25	0.201	0.206	5.7	5.5
Black	168.22	0.192	0.200	8.2	7.7
Mex-Amer	163.04	0.195	0.205	11.0	10.1
Weight (kg)					
White/other	73.85	0.364	0.380	9.2	8.6
Black	77.58	0.506	0.542	14.8	13.2
Mex-Amer	72.20	0.451	0.465	6.3	5.9
Systolic BP					
White/other	121.46	0.487	0.515	11.8	10.7
Black	124.18	0.709	0.766	16.6	14.6
Mex-Amer	117.94	0.647	0.675	8.8	8.2
Diastolic BP					
White/other	72.66	0.395	0.416	10.8	10.0
Black	74.63	0.473	0.503	13.2	11.9
Mex-Amer	71.17	0.519	0.536	6.4	6.1
Total cholesterol					
White/other	206.04	1.036	1.118	16.3	14.4
Black	201.28	0.960	1.075	25.3	20.9
Mex-Amer	200.28	2.447	2.490	3.6	3.5
HDL cholesterol					
White/other	51.00	0.388	0.414	14.1	12.7
Black	55.95	0.462	0.514	23.7	19.8
Mex-Amer	50.03	0.505	0.524	7.6	7.1

calculated using basic survey weights (i.e., inverse probabilities of selection) without adjustments for poststratification. Complete-data variance estimates U_i were calculated with SUDAAN software (Shah et al., 1991) using a linearization method appropriate for the sample design. The degrees of freedom ν (not shown) ranged from 221 to 7526, suggesting that the between-imputation components of variance tend to be well estimated. In contrast, we have good reason to suspect that the within-imputation components of variance are estimated rather poorly. Design effects (not shown) provided by SUDAAN displayed erratic behavior across the ten sets of multiple imputations (Little and Rubin, 1992), suggesting that the design-based variance estimation method is inherently unstable, due perhaps to the small number of primary sampling units. Keeping this in mind, we interpret the results in Table 3 only with caution.

The fractions of missing information γ are quite small, ranging from 20.9% down to 3.5%, even though the MEC variables in this dataset were missing at rates in excess of 30%. This suggests that the gains in precision from a good imputation model, which makes intelligent use of information about Y_{mis} available in Y_{obs}, can be substantial. The relative increases in variance due to nonresponse r_m range from 4 to 25%, so the total variances T_m tend to be not much larger than the within component $\bar{U}_m$. The multiply-imputed interval estimates for these quantities, at least, are not much wider than single-imputation intervals. Within smaller subdomains, however, we found that the values of r_m could be much higher. For most uses of this dataset that we can imagine, it appears that $m = 10$ imputations are more than enough to permit accurate and efficient inferences (Little and Rubin, 1992). With most fractions of missing information in the range 5–15%, it seems that $m = 5$ or even $m = 3$ would be adequate.

7 References

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

Ezzati, T. and Khare, M. (1992), "Nonresponse adjustments in a national health survey," *Proceedings of the Survey Research Methods Section of the American Statistical Association*, 339–344.

Gelfand, A.E. and Smith, A.F.M. (1990), "Sampling-based approaches to calculating marginal densities," *Journal of the American Statistical Association*, 85, 398-409.

Geman, S. and Geman, A. (1984), "Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 6, 721-740.

Krzanowski, W.J. (1982), "Mixtures of continuous and categorical variables in discriminant analysis: a hypothesis testing approach," *Biometrics*, 38, 991–1002.

Little, R.J.A. and Rubin, D.B. (1992), "Assessment of Trial Imputations of NHANES: Final Report," Waban, MA: Datametrics Research.

Little, R.J.A., and Schluchter, M.D. (1985), "Maximum-likelihood estimation for mixed continuous and categorical data with missing values," *Biometrika*, 72, 492–512.

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

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

Schafer, J.L. (1991), "Algorithms for multiple imputation and posterior simulation from incomplete multivariate data with ignorable nonresponse," Ph.D. Dissertation, Department of Statistics, Harvard University, Cambridge, MA.

Schafer, J.L., Khare, M. and Ezzati-Rice, T. (1993), "Multiple imputation of missing data in NHANES III," *Proceedings of the Annual Research Conference*, U.S. Bureau of the Census, to appear.

Shah, B.V., Barnwell, B.G., Hunt, P.N., LaVange, L.M. (1991), *SUDAAN User's Manual: Professional Software for Survey Data Analysis for Multistage Sample Designs, Release 5.0*, North Carolina: Research Triangle Park.