October, 2003

Imputation of missing longitudinal data: a comparison of methods

Paula Diehr, *University of Washington*
Jean Mundahl Engels, *University of Washington*

ELSEVIER

Journal of Clinical Epidemiology 56 (2003) 968–976

Journal of Clinical Epidemiology

Imputation of missing longitudinal data: a comparison of methods

Jean Mundahl Engels*, Paula Diehr

Departments of Biostatistics and Health Services, University of Washington, 1959 Northeast Pacific Avenue, Box 357232, Seattle, WA 98195, USA

Accepted 11 October 2002

Abstract

Background and Objective: Missing information is inevitable in longitudinal studies, and can result in biased estimates and a loss of power. One approach to this problem is to impute the missing data to yield a more complete data set. Our goal was to compare the performance of 14 methods of imputing missing data on depression, weight, cognitive functioning, and self-rated health in a longitudinal cohort of older adults.

Methods: We identified situations where a person had a known value following one or more missing values, and treated the known value as a "missing value." This "missing value" was imputed using each method and compared to the observed value. Methods were compared on the root mean square error, mean absolute deviation, bias, and relative variance of the estimates.

Results: Most imputation methods were biased toward estimating the "missing value" as too healthy, and most estimates had a variance that was too low. Imputed values based on a person's values before and after the "missing value" were superior to other methods, followed by imputations based on a person's values before the "missing value." Imputations that used no information specific to the person, such as using the sample mean, had the worst performance.

Conclusions: We conclude that, in longitudinal studies where the overall trend is for worse health over time and where missing data can be assumed to be primarily related to worse health, missing data in a longitudinal sequence should be imputed from the available longitudinal data for that person.

Keywords: Missing data; Imputation; Longitudinal; Depression; Cohort

1. Introduction

In longitudinal studies, it is unlikely that every person's information will be obtained at all prespecified times. Missing data can cause biased estimates, because people with missing data are usually sicker than those with complete data [1]. Additionally, power decreases if a complete case analysis is performed, because the sample size is smaller. The goal of this article is to compare 14 different methods of imputation using a real data set with real missing data patterns where the true values of the missing data are known.

Statistical methods are available that take the missing data into account at the time of analysis. These methods include likelihood-based approaches such as generalized linear models and the expectation-maximization (EM) algorithm when data are "missing at random" [2]. If the data are "missing not at random," methods that specify a model for the missing data mechanism or that are robust to missing data mechanisms must be used. A different approach is to impute the missing values so that the resulting dataset is, in a sense, complete. This is most convenient for large prospective databases that will be used for many different types of analyses by a number of researchers. Methods that account for missingness at the time of data analysis will not be considered further here. Analyses of the imputed data sets may need to be modified, because the variances of imputed values tend to be underestimated [3]. Several approaches, such as multiple imputation, have been developed to address this problem [4,5]. Although multiple imputation may be the recommended imputation approach, the current article considers only the situation where data are imputed once, and where only standard analytic methods will be used for the analysis.

2. Methods

It often is difficult to determine and compare the accuracy of different imputation methods. Unless one is able to retrieve the missing data, a method must be devised to create a dataset that mimics real life data and missing data patterns,

* Corresponding author. 2167 Watertown Road, Long Lake, MN 55356; Tel.: 651-271-9383.

E-mail address: mundahl@u.washington.edu (J.M. Engels).

0895-4356/03/$ – see front matter
doi: 10.1016/S0895-4356(03)00170-7

but where the true value of the missing datum is known. Some studies of imputation methods have used real datasets and simulated missing data patterns by deleting values. [6] With this method, the true value is known, but the missingness pattern may not be realistic. Other studies have performed imputations on existing data sets with missing data, but because the true values were not known, the accuracy of the results could not be determined [7]. Here we use real data, a real missingness pattern, and a known true value. We next describe the dataset, the method of constructing missing data, the imputation methods to be compared, and how these methods were evaluated.

2.1. Data

The data used for this article came from the Cardiovascular Health Study (CHS), a population-based, longitudinal study of coronary heart disease and stroke in adults aged 65 years and older (mean age at baseline was 73 years) [8]. Its main objective was to identify risk factors related to coronary heart disease and stroke. The study population consisted of 5,888 persons—2,495 men and 3,393 women. We used four longitudinal variables—Health Status, Weight, Depression, and Minimental score—to compare various imputation methods. These variables were assessed first in 1990 and then either semiannually or annually thereafter. Health Status was a self-rated measure of a person's perceived overall health: excellent, very good, good, fair, or poor (coded 5, 4, 3, 2, 1). Weight was measured in pounds. Depression was assessed using a modified, short version of the Center for Epidemiologic Studies Depression Scale [9], with a range of scores from 0 to 30; higher values represent more depressive symptoms. Cognitive function (Minimental) was measured by the Mini-Mental State Examination [10], with a range of scores from 0 to 100, with higher values representing better cognitive function.

The variables chosen were a mix of both continuous and discrete variables, and had varying distributions (not shown here). Although Weight had a reasonably normal distribution that stayed fairly constant over time, the distribution of Health Status changed over time as persons got sicker. The distribution of Depression was skewed right, while Minimental was highly skewed to the left. Using these different types of variables allowed us to evaluate whether the performance of the imputation methods was dependent on the unique properties of the variables.

As it turned out, results were consistent across the four variables, and therefore, this article reports detailed results for only one variable, Depression. Results for the other variables are summarized at the end, and complete results can be found elsewhere [11]. We also used additional baseline variables that were required in certain imputation methods, such as regression imputation; these variables included age, gender, and baseline values for the four study variables.

2.2. Construction of "missing values"

In CHS, unlike many longitudinal studies, almost no one was lost to follow-up, except to death. Persons who had one or several missing values usually returned later in the study. We took advantage of this feature to study the performance of the imputation techniques. An observation for a person who had just missed one or several observations had a high probability of being missing as well. For example, a Depression value following a missing Depression value was about 10 times as likely to be missing as a Depression value following an observed value. (This was true whether or not we included values missing because of death.) Thus, the first valid observation following a string of missing observations had a high probability of being missing, even though it was, in fact, observed. Our approach was to set such values to "missing," impute them using different methods, and then compare the imputed values to the values that were actually observed. Thus, although data are restricted to persons who returned after having had an absence, we used real data with a real missingness pattern, and the true value was known.

Depression data were available annually from 1990 to 1999. We defined a "missing value" as a known value following a missing value. Furthermore, some of the imputation methods required that there be at least one known value before and/or after the "missing value." The earliest possible "missing" data would thus be in 1992 (where 1990 was known, 1991 was missing, 1992 was known, and where at least one value from 1993–1999 was known), and the latest possible "missing value" would be in 1998 (where at least one value from 1990–1996 was known, 1997 was missing, 1998 was known, and 1999 was known). If a person had more than one string of missing values followed by a known value, all such possibilities were used. In addition, we did not estimate data that were missing due to death because the person had to be alive for the "missing value" to be known in the first place.

2.3. Imputation methods

Various imputation methods were used to estimate the "missing values." Some methods used only information pertaining to the person whose data were missing, while some used the values of other persons. We expected methods that capitalized on the existence of longitudinal data for the person with a missing value to perform better than other methods. The following methods assume a dataset where each row corresponds to a person's data, and the columns correspond to sequential years for the variable being imputed. The imputation methods are divided into four categories for comparison purposes according to the type of data they use to make the estimate. The four groups are: Population, Baseline, Before, and Before and After (B/A). The methods are defined below and summarized in Table 1.

2.3.1. "No person data" methods (population group)

These methods use no information specific to the person. The *column mean* and *column median* use the mean/median (of Depression) for that year to estimate the "missing value."

Table 1
Imputation methods defined

Group	Method	Definition
Population	Column mean	Mean of all persons in the dataset for a particular year
Population	Column median	Median of all persons in the dataset for a particular year
Baseline	Class mean	Mean of other persons in corresponding class
Baseline	Class median	Median of other persons in corresponding class
Baseline	Hot deck	Value of a random person in corresponding class
Baseline	Regression	Predicted value from a regression model
Baseline	Regression with error	Same as Regression with an error term added
Before	Previous row mean	Mean of person's previous known values
Before	Previous row median	Median of person's previous known values
Before	LOCF	Last observation carried forward
B/A	Row mean	Mean of person's values before and after
B/A	Row median	Median of person's values before and after
B/A	NOCB	Next observation carried backward
B/A	Last & next	Average of the last known and next known values

2.3.2. "Baseline data + other covariates" methods (baseline group)

These methods use the person's baseline Depression value along with other baseline covariate information. No longitudinal information is used. One way that covariates are used is to define imputation classes that group persons according to their covariates [5]. The imputation classes for Depression were formed by grouping persons according to their baseline Depression score (three groups), baseline health status, and gender. When data were missing for one of these classifying variables, the persons with missing data formed their own class. Additionally, when there were only a few persons in a class, we combined the classes with a neighboring class to be able to impute the missing values. After the imputation classes were constructed, several class imputation methods were performed. The *class mean* and *class median* impute the missing value as the mean/median of known Depression values at that time point, in the appropriate class. The *hot deck* [12] method, which is used to impute labor force items in the Current Population Survey [13], involves a recipient (the person with missing data) and a donor (another person in the same imputation class, whose value is known). We selected a donor by sampling without replacement from the recipient's class, and replaced the recipient's missing value with the donor's value at the appropriate time point. These class imputations imply that persons with missing data are a random sample of the persons in their class [14].

Another baseline approach is *regression* imputation. The variable of interest (known Depression value for the year in which the "missing value" occurred) is regressed on baseline covariates, and the resulting equation is used to estimate the missing values for that year. We regressed Depression on baseline Depression, age, gender, and health status. Data for persons missing any covariate information were not imputed. Two different types of regression imputation were performed. *Regression* assigns the person's predicted value to the "missing value." Persons with the same covariates will have exactly the same imputed value. This can lead to the variance of the imputed data set being too small, yielding inappropriately small standard errors and *P*-values at the time of analysis. To address this problem, *regression with error* [15] assigns the predicted value plus a randomly chosen value from a $N(0, s_e^2)$ distribution where s_e^2 is the residual variance from the regression.

2.3.3. "Before data only" methods (before group)

These methods use longitudinal data on a person only up to the time of the "missing" observation. These are appropriate in the common situation where a person is lost to follow-up and there are no later data. The *previous row mean* and *previous row median* are calculated from all of the person's Depression scores prior to the "missing value." Another method, the *last observation carried forward (locf)*, assigns the person's last previous known Depression score to the "missing value."

2.3.4. "Before and after data" methods (b/a group)

These methods use the Depression data on a person both before and after the "missing value." *Last & next* assigns the average of the person's last known and next known observations to the "missing value." The *row mean* and *row median* assign the mean/median of all the person's known Depression values (1990–1999) to the "missing value." The *next observation carried backward (nocb)* assigns the person's next known Depression score after the "missing" one to the "missing value." (*Nocb* does not require "before" data, but was placed in this set of methods for convenience).

2.3.5. Comparison of methods

After imputing the "missing values," we examined the performance of the estimates using four summary measures. Two measures of accuracy were used, one being the root-mean-square deviation defined as:

$$\mathrm{RMSD} = (\Sigma(y-\hat{y})^2/m)^{1/2} \quad (1)$$

where $\hat{y}$ is the imputed value, y is the true value, and m is the number of "missing values." Another measure used was the mean absolute deviation, defined as:

$$\text{MAD} = \Sigma|y-\hat{y}|/m \quad (2)$$

These two terms measure how close the estimated values are to the true values [15]. They do not always give the same results; the RMSD penalizes outliers more because the difference term is squared. We expected estimates based on the person's known values to outperform those that used only data or relationships from other persons. We also expected that methods based on data from persons without missing data would underestimate Depression. An additional summary measure was bias, assessed by computing the mean deviation, where:

$$\text{BIAS} = \Sigma(y-\hat{y})/m \quad (3)$$

We also assessed how well the imputed values preserved the variance of the true values. We expected the population methods to perform worst, because the estimate is identical for each person. *Class mean and median* and *regression* (without an error term added) also should be underdispersed because persons with the same covariates have the same estimated value. *Last & next, row mean,* and *previous row mean* should be underdispersed because means have smaller variance than individual observations, and the methods based on row medians could be underdispersed for the same reason. *Regression with error* was specifically designed to have appropriate variation. *Locf, nocb,* and *hot deck* are samples from the distribution of a person's values, and so might have appropriate variation. To assess underdispersion for each method, the sample variance of all estimated "missing values" was computed and compared to the variance of the known values that had been set to missing. A proportionate variance was calculated for each method where:

$$\text{PV} = \text{var}(\hat{y})/\text{var}(y) \quad (4)$$

2.4. Analysis

We needed more than one estimate of each summary measure to permit calculation of confidence intervals. Persons with "missing values" were put into groups (replicates) according to the year of the "missing value." We then calculated 95% confidence intervals for the mean of each performance measure, for each imputation method. An example of these intervals is depicted in Fig. 1, which shows that the average RMSD for *last & next* was approximately 4.6, with a 95% confidence interval of (4.4, 4.8). Finally, interactions between method and time were looked for by fitting general linear models with the summary measure (each measure separately) as the outcome, method as a factored variable, time as a continuous covariate, and the interaction of method and time. We found that there was not a significant time effect, nor were any interactions between method and time statistically significant for Depression or any other variable. Results of other analyses are noted briefly.

3. Results

3.1. Descriptives

There were 1,176 "missing values" for Depression among 1,023 persons. Eleven persons had three separate runs of

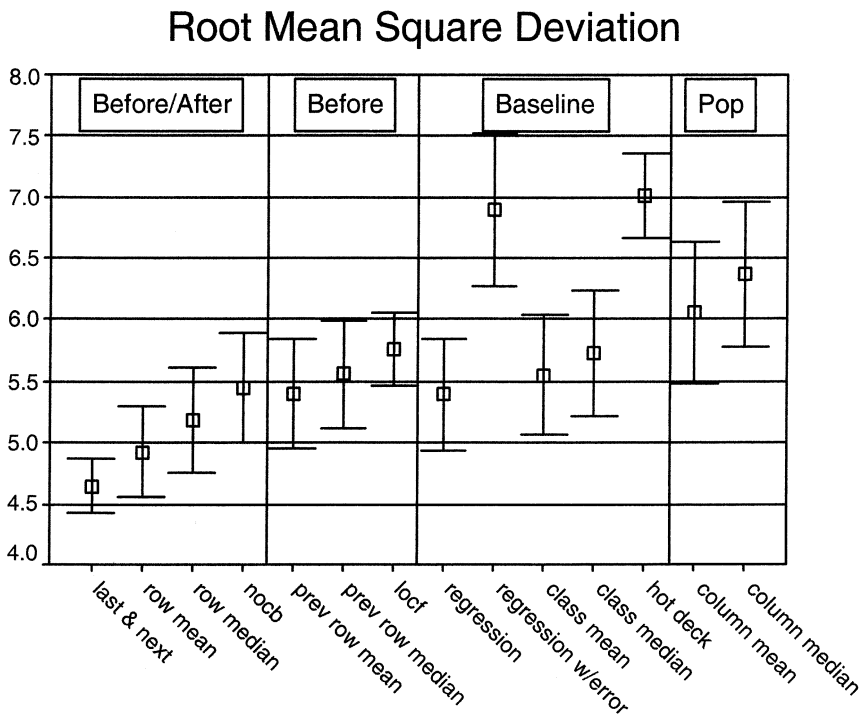

Fig. 1. 95% confidence intervals for the mean RMSD.

missing values and were included three times, and 142 were included twice. Sixty-five percent of the persons were missing four or fewer values before their imputed value. The average "missing value" was about 1.5 points higher than the mean for nonmissing values for the same year, and the median was 2.5 points higher. Persons returning after a missing value were thus more depressed on average than people who had not missed the previous assessment. Table 2 gives descriptive statistics for Depression and the rate of missing data by year. As can be seen, the missing data rate increased over time, in part due to the death of some people. The mean Depression score also increased somewhat over time, even though the persons who responded were a favorably selected subgroup. The true increase in Depression is probably greater.

Table 2
Descriptive statistics and missing data rate for Depression by year

Depression	N	Mean (SD)	Minimum	Maximum	% Missing
1990	5878	4.71 (4.60)	0.00	29.00	0.17
1991	5472	5.20 (4.76)	0.00	29.00	7.07
1992	5183	5.12 (4.94)	0.00	30.00	11.97
1993	4917	5.50 (4.96)	0.00	28.00	16.49
1994	4529	5.29 (4.92)	0.00	30.00	23.08
1995	4337	5.79 (5.03)	0.00	29.00	26.34
1996	4170	5.90 (5.12)	0.00	29.00	29.18
1997	3382	5.95 (5.07)	0.00	30.00	42.56
1998	3235	5.86 (4.99)	0.00	29.00	45.06
1999	2943	5.59 (4.91)	0.00	30.00	50.02

3.2. Accuracy results

The RMSD and MAD were used to compare the accuracy of the different imputation methods. Figs. 1 and 2 show the average value (calculated from the seven replicates) of the RMSD and MAD, respectively. The *last & next* method had the smallest mean RMSD and MAD among all methods, 4.65 (±0.24) and 3.43 (±0.19), respectively. Performance became gradually worse for methods farther to the right on the graph. *Hot deck, regression with error,* and the *column mean and median* had substantially worse accuracy than the other methods, with an average RMSD greater than 6 and an average MAD greater than 4.

3.3. Bias results

The mean deviation was used to assess the bias, with a MD of zero indicating no bias. A positive bias indicates that on average, the imputed value underestimated the true value. Fig. 3 shows that the *last & next* and *nocb* methods had little bias. Mean biases for these two methods were 0.12 (±0.46) and −0.31 (±0.50), respectively. The most biased method was the *column median* (2.43 ± 0.74). Because the majority of the bias estimates were above zero, most of the imputed values were biased toward showing less Depression. There was a similar finding for Health Status and

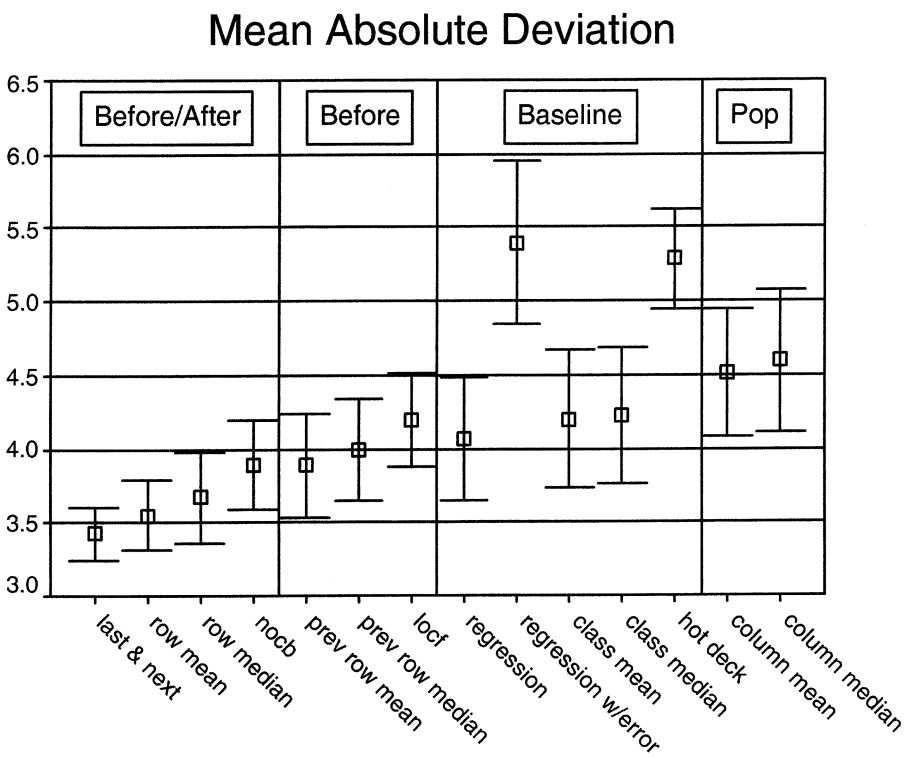

Fig. 2. 95% confidence intervals for the mean MAD.

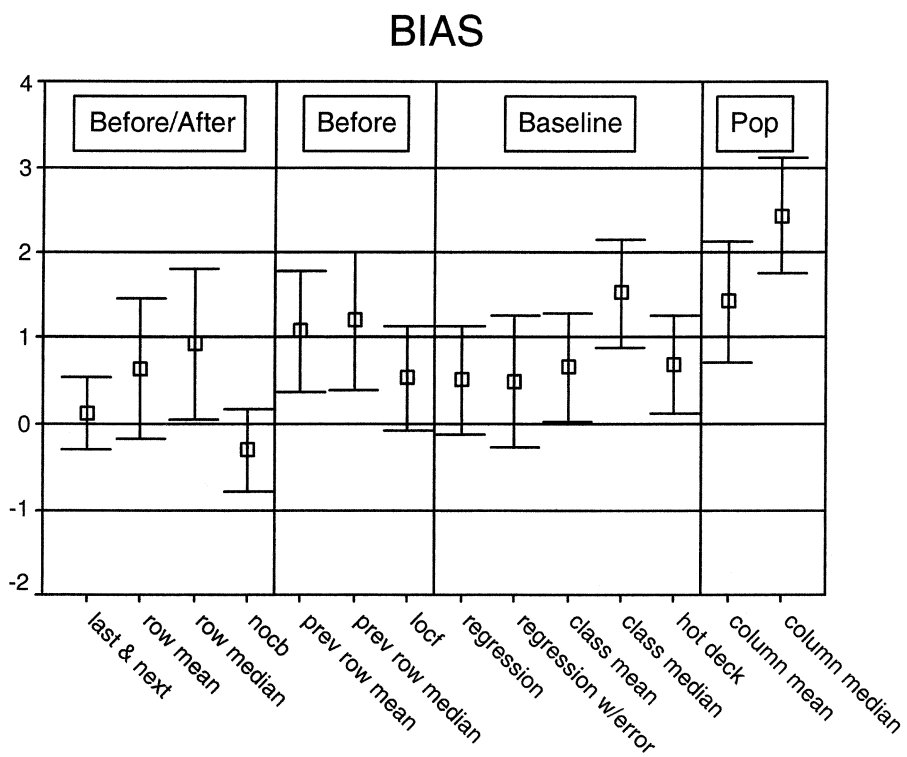

Fig. 3. 95% confidence intervals for the mean bias.

Minimental, with the imputed values being "too healthy." Weight was usually overestimated. Given that very low weight is associated with worse health for older adults [16], health may have been overestimated on this variable as well.

3.4. Proportionate variance results

The proportionate variance measure was used to assess each method's effectiveness in preserving the variance of the true values. A PV of 1 implies that the variance of the imputed values was equal to the variance of the true values; a PV less than 1 corresponds to underestimation of the true variance. Fig. 4 shows that all estimates except *nocb* had a variance that was too small (underdispersion). The mean PV for *nocb* was 1.06 ($\pm$0.16), while the means for all other methods were less than 1. As expected, *nocb, locf, hot deck,* and *regression with error* did well, and *column mean and median* had the worst performance. We had not expected the good performance of the estimators that involved means of longitudinal data, because means have lower variance than individual variables. For example, the variance of the mean of two variables is σ^2 $(1 + \rho)/2$, where σ^2 is their common variance and ρ is their correlation. The variance of *last & next* would have been $\sigma^2/2$ if the prior and following measures were independent, corresponding to a PV of 0.5. The correlation between Depression measures 2 years apart was about 0.6 (data not shown), suggesting the mean would have a variance of $\sigma^2(1 + 0.6)/2$, or a PV of about 0.8, which is close to what was observed. The estimates based on longitudinal means were thus not as underdispersed as expected, due to the high correlation over time.

3.5. Summary results for all health variables

We also summarized the overall performance of the imputation methods by examining all four health-related variables. To do this, we fit one-way ANOVAs for each variable and summary measure with imputation method as the factored variable. Post hoc tests were then done using Bonferroni's adjustment to determine which methods differed from one another, and each method was categorized as performing "well," "acceptably," or "poorly" using the results from these pairwise comparisons for each variable. A count was given to each method for how many times it performed "well," etc. (out of four total possible times corresponding to the four variables). Roughly speaking, a method was categorized by determining what other methods it was and was not significantly different from. If a method was significantly better (or worse) than most other methods, then this method was classified as performing "well" (or "poorly"). If it did not perform significantly better or worse than any other method, then it was classified as performing "acceptably." The methods that were different from the same (other) methods were put into the same performance category. Table 3 denotes situations where each method performed well for each performance measure. For example, the "D" in the first row indicates that for Depression, *last & next* performed well

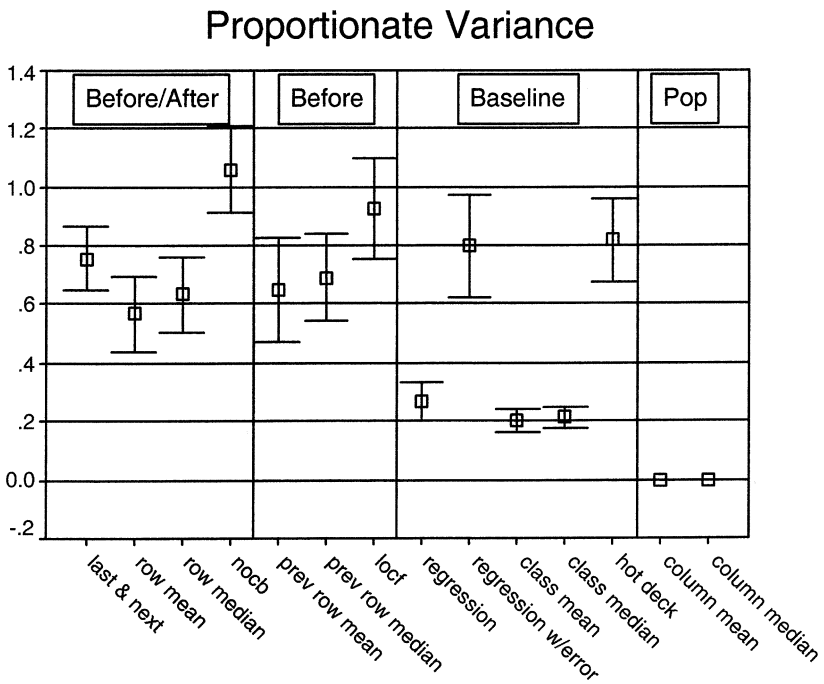

Fig. 4. 95% confidence intervals for the mean PV.

with respect to RMSD, MAD, and Bias, but was not significantly better than most other measures with respect to PV (after adjusting for multiple comparisons), which can be seen in Figs. 1–4. The only methods that performed well were the methods that used data before and after the "missing value," with *last & next* and *nocb* performing the best. Most methods performed at least acceptably on all performance measures (data not shown). The *hot deck, class mean and median,* and *column mean and median* methods had the worst performance overall.

3.6. Relaxing the restrictions

The subjects of this study all had to have at least two known values after their sequence of missing values, one to be set to "missing" and the second to permit calculation of the "before/after" estimates. This may have led to a healthier group of subjects. To investigate this possibility, we performed a second analysis requiring only that people have one known value after the sequence of missing values. We found that the order of the methods was about the same as before. (The before/after measures were, of course, not evaluated in this subset.)

Table 3
Number of times each method performed well

	RMSD	MAD	BIAS	PV
Last & next	DWMH	DWMH	DM	W
Row mean	DH	DH		
Row median		DH	H	H
Nocb	WH	WH	DMH	DWM
Previous row mean				
Previous row median				
Locf				D
Regression				
Regression with error				D
Class mean				
Class median				
Hot deck				DH
Column mean				
Column median				

Abbreviations: D, depression; H, health status; M, minimental; W, weight.

4. Discussion

Persons with missing data are usually different from those with known data, and it is likely that any imputation method that allows them to be kept in the study is an improvement over the complete case method. The best method depends on the goal of the analysis. For example, if the goal were to study whether one value could predict the value at the next time point, it would not be appropriate to estimate the missing data at all. Although there were some exceptions, the results agreed with our expectation that methods using a person's own longitudinal data would be superior to methods that used less or no information about the person with missing data. Although the Baseline and Population methods assume the data are missing at random, the Before/After and After methods also in a sense imply missing at random data, in that they assume the missing data can be estimated from data that are available for that individual. We think it likely that our data are not missing at random, which could explain their less than perfect performance.

All methods except *nocb* were biased toward making the persons appear healthier than they actually were. The

apparent superiority of *nocb* to *locf* is due in part to our definition of a "missing value" as an observed value following a missing value. Locf was based on a preceding value that was obtained at least 2 years before the "missing value," while *nocb* could have been based on a value only 1 year after the missing value. *Locf* and *nocb* would probably be more similar in the general situation. The *nocb* method was also favored by the general trend toward more Depression over time, which would make *nocb* "too depressed" and counter somewhat the positive bias found in all other methods. If the overall trend was for improving health, most of the methods would still overestimate health, and *locf* would then be less biased.

All of the estimates but *nocb* tended to be underdispersed, but the problem was not very large for measures based on the individual's longitudinal data, which performed about as well as *hot deck* and *regression with error*. This good performance is probably due to the correlation over time among measures.

4.1. Are the "missing values" similar to real missing values?

This analysis hinges on the similarity of a known value following a string of missing values to other observations that are missing at that same time. There is some evidence that this assumption is reasonable. The values come from persons who have missed values at other times. We noted that observations following a missing value are 10 times as likely to be missing as observations following a known value. We found that the "missing" values denoted worse health than the population mean (because the *column mean* imputation was biased high) and worse health than before the string of missing values (*locf* biased high). We think it likely that our "missing" values represent a healthier subset of all persons with missing data because these persons did, in fact, return. If we knew the "true" value of missing data, average health would likely be even lower, and the biases for the population methods even larger than we have seen, whereas the biases of the person-based methods do not seem as vulnerable to having a favorable sample. Every comparison of imputation methods for missing data must depend on some assumptions. We believe that our assumption is reasonable, and that biases created by it tended to make the population methods look better rather than worse.

4.2. Generalizability

The CHS data had several unusual characteristics that permitted us to conduct this exercise. These results, however, should be relevant for a broader set of missing data problems. Results should generalize to longitudinal studies in which the overall trend is for worse health over time, where missing data can be assumed to be primarily related to worse health, where the data are collected at regular intervals, and when there is no treatment associated with the data collection opportunity. In studies where the trend is to improve over time, the performance of *nocb* and *locf* would likely be reversed. If data were collected at the time of a patient-initiated visit, or if treatment was involved, it is likely that persons with missing data would be healthier than average, rather than sicker. Again, this could reverse the direction of the bias that we found. We believe that the results pertaining to the Before methods can be generalized to typically observed datasets where subjects do not ever come back, because biases introduced by our analyzing favorable subjects who did return were probably against the direction of our main findings. The measures we studied were taken annually. It seems likely that if variables were measured more frequently, the person-based methods would have even better performance.

4.3. Limitations

The assessment of whether a method "performed well" (Table 3) was somewhat subjective because many tests were performed, and there may not have been enough replicates to guarantee that the tests had the proper size. The number of replicates ($n = 7$) was small, and assumptions of the ANOVA may not have been met. However, we thought that this analysis would give the reader an idea of the similarities in method performance across all four variables. Also, nonparametric Kruskal-Wallis *H*-tests were performed and found to be in agreement with the ANOVA results. The confidence intervals may not have been quite accurate, as well, due to the small number of replicates. Nevertheless, all methods were applied to the same data, so the relative sizes of the confidence intervals should be nearly accurate. An additional limitation was that many imputation methods were not considered. Some of these are variants of the methods used here, and their performance can be inferred from our tables.

5. Conclusions

In a large longitudinal study where numerous analyses will be performed, it is convenient to have an imputed database. If longitudinal data are available before and after the missing value, we recommend the *last & next* method. If longitudinal data are available only before the missing value, *locf* or the *previous row mean/median* should be used. If longitudinal data are not available, the Baseline methods are preferred over the Population methods. A combined strategy that imputes variables using the best possible method for the situation (e.g., *last & next* when there are values before and after, and *locf* when no after value is available, etc.) would likely have the best performance.

Acknowledgments

This work was supported by contracts N01-HC-85079 through N01-HC85086 from the National Heart, Lung, and

Blood Institute. Participating Institutions and Principal Staff: **Forsyth County, NC—**Bowman Gray School of Medicine of Wake Forest University: Gregory L. Burke, Sharon Jackson, Alan Elster, Walter H. Ettinger, Curt D. Furberg, Gerardo Heiss, Dalane Kitzman, Margie Lamb, David S. Lefkowitz, Mary F. Lyles, Cathy Nunn, Ward Riley, John Chen, Beverly Tucker; Forsyth County, NC—Bowman Gray School of Medicine-EKG Reading Center: Farida Rautaharju, Pentti Rautaharju; **Sacramento County, CA—**University of California, Davis: William Bommer, Charles Bernick, Andrew Duxbury, Mary Haan, Calvin Hirsch, Lawrence Laslett, Marshall Lee, John Robbins, Richard White; **Washington County, MD—**The Johns Hopkins University: M. Jan Busby-Whitehead, Joyce Chabot, George W. Comstock, Adrian Dobs, Linda P. Fried, Joel G. Hill, Steven J. Kittner, Shiriki Kumanyika, David Levine, Joao A. Lima, Neil R. Powe, Thomas R. Price, Jeff Williamson, Moyses Szklo, Melvyn Tockman; MRI Reading Center-Washington County, MD—The Johns Hopkins University: R. Nick Bryan, Norman Beauchamp, Carolyn C. Meltzer, Naiyer Iman, Douglas Fellows, Melanie Hawkins, Patrice Holtz, Michael Kraut, Grace Lee, Larry Schertz, Cynthia Quinn, Earl P. Steinberg, Scott Wells, Linda Wilkins, Nancy C. Yue; **Allegheny County, PA—**University of Pittsburgh: Diane G. Ives, Charles A. Jungreis, Laurie Knepper, Lewis H. Kuller, Elaine Meilahn, Peg Meyer, Roberta Moyer, Anne Newman, Richard Schulz, Vivienne E. Smith, Sidney K. Wolfson; Echocardiography Reading Center (Baseline)–University of California, Irvine: Hoda Anton-Culver, Julius M. Gardin, Margaret Knoll, Tom Kurosaki, Nathan Wong; Echocardiography Reading Center (Follow-Up)—Georgetown Medical Center: John Gottdiener, Eva Hausner, Stephen Kraus, Judy Gay, Sue Livengood, Mary Ann Yohe, Retha Webb; Ultrasound Reading Center–Geisinger Medical Center: Daniel H. O'Leary, Joseph F. Polak, Laurie Funk; Central Blood Analysis Laboratory–University of Vermont: Edwin Bovill, Elaine Cornell, Mary Cushman, Russell P. Tracy; Respiratory Sciences–University of Arizona, Tucson: Paul Enright; **Coordinating Center—**University of Washington, Seattle: Alice Arnold, Paula Diehr, Annette L. Fitzpatrick, Richard A. Kronmal, Will Longstreth, Bruce M. Psaty, Chuck Spiekerman, David S. Siscovick, Patricia W. Wahl, David Yanez; **NHLBI Project Office**: Diane E. Bild, Robin Boineau, Teri A. Manolio, Peter J. Savage, Patricia Smith.

References

[1] Fairclough DL, Peterson HF, Chang V. Why are missing quality of life data a problem in clinical trials of cancer therapy? Stat Med 1998;17:667–77.

[2] Laird NM. Missing data in longitudinal studies. Stat Med 1988;7: 305–15.

[3] Madow WG, Nisselson H, Olkin I. Incomplete data in sample surveys. Vol. 1. Report and case studies. New York: Academic Press; 1983.

[4] Rubin DB. Multiple imputation for nonresponse in surveys. New York: Wiley; 1987.

[5] Cox BG, Cohen SB. Methodological issues for health care surveys. New York: Marcel Dekker; 1985.

[6] Robertson KW, Tou A, Huff L. A study of donor pools and imputation methods for missing employment data. ASA Proc Sect Survey Res Methods; 1995.

[7] Lepkowski JM, Landis JR, Stehouwer SA. Strategies for the analysis of imputed data from a sample survey. The National Medical Care Utilization and Expenditure Survey. Med Care 1987;25(8):705–16.

[8] Fried LP, Borhani NO, Enright P, Furberg CD, Gardin JM, Kronmal RA. The Cardiovascular Health Study: design and rationale. Ann Epidemiol 1991;1:263–76.

[9] Orme J, Reis J, Herz E. Factorial and indiscriminate validity of the Center for Epidemiological Studies Depression (CES-D) Scale. J Clin Psychol 1986;42:28–33.

[10] Folstein MF, Folstein SE, McHugh PR. Mini-Mental Stat: a practical method for grading the cognitive state of patients for the clinician. J Psychiatr Res 1975;12(3):189–98.

[11] Mundahl JM. Imputation of missing longitudinal data: a comparison of methods. University of Washington, Department of Biostatistics; December 1998.

[12] Madow WG, Olkin I, Rubin D, editors. Incomplete data in sample surveys; vol. 1, theory and bibliographies. New York: Academic Press; 1983. p. 185.

[13] Brooks CA, Bailer BA. An error profile: employment as measured by the current population survey. Statistical policy working paper 3. Washington, DC: U.S. Department of Commerce, U.S. Government Printing Office; 1978.

[14] Brick JM, Kalton G. Handling missing data in survey research. Stat Methods Med Res 1996;5:215–38.

[15] Kalton G. Compensating for missing survey data. Ann Arbor, MI: Institute for Social Research; 1983.

[16] Diehr P, Bild D, Harris T, Duxbury A, Siscovick D, Rossi M. Body mass index and mortality in nonsmoking older adults: the cardiovascular health study. Am J Public Health 1998;88:623–9.