Available online at www.sciencedirect.com
ScienceDirect
Research in Social and Administrative Pharmacy 3 (2007) 1–27
RESEARCH IN SOCIAL & ADMINISTRATIVE PHARMACY

Original Research

Missing data on the Center for Epidemiologic Studies Depression Scale: A comparison of 4 imputation techniques

Christine Bono, Ph.D.[a], L. Douglas Ried, Ph.D.[a,b,*], Carole Kimberlin, Ph.D.[a], Bruce Vogel, Ph.D.[b,c]

[a]*Pharmacy Health Care Administration, College of Pharmacy, University of Florida, Gainesville, FL, USA*

[b]*Rehabilitation Outcomes Research Center of Excellence, NF/SG Veterans Affairs Medical Center, Gainesville, FL, USA*

[c]*Department of Epidemiology and Health Policy Research, College of Medicine, University of Florida, Gainesville, FL, USA*

Abstract

Background: Missing data are widespread in the medical sciences. Given their prevalence, researchers must be prepared to address problems that arise when data are missing.

Objectives: The objectives were to (1) provide an estimate of bias for each imputation technique with known values from data engineered to be missing completely at random; (2) determine whether different Center for Epidemiologic Studies Depression (CES-D) Scale scores were obtained from item-mean, person-mean, regression, and hot-deck imputation techniques and whether they differed from the CES-D score obtained from complete cases; and (3) determine whether the variables that predicted

Supported with resources and the use of facilities at the Malcom Randall Veterans Affairs Medical Center in Gainesville, Fla. We also acknowledge partial funding by Knoll Pharmaceuticals (now Abbott Laboratories) and the University of Florida.

* Corresponding author. Pharmacy Health Care Administration, College of Pharmacy, PO Box 100496, University of Florida, Gainesville, FL 32610-0496, USA. Tel.: +1 352 273 6259; fax: +1 352 273 6270.

E-mail address: ried@cop.ufl.edu (L.D. Ried).

doi:10.1016/j.sapharm.2006.04.001

the CES-D scores were the same for the complete cases and each of the 4 imputation techniques.
Methods: Depressive symptoms were assessed in patients (N = 2,317) in an international clinical trial comparing high blood pressure treatments between April 1, 1999, and October 31, 1999. Patients were mailed surveys after randomization. Depressive symptoms were measured using the CES-D Scale. Respondents who completed all 20 items were compared with those who did not complete all 20 items, using independent *t* tests and chi-square. *Z* scores were used to determine CES-D mean differences, and multiple regression models were used to predict the CES-D scores for the 4 imputation techniques and the complete case data.
Results: Imputed CES-D mean scores ranged from 14.58 to 14.68. The 4 imputed CES-D mean scores were consistently, but not significantly, higher than the complete case CES-D mean of 14.06. Imputed mean scores were similar to each other and the complete case mean score. Four regression models predicting the imputed CES-D scores yielded similar predictions. With the exception of sex, the same variables predicted the complete case CES-D and the imputed CES-D scores.
Conclusions: All the imputed means were similar to the complete case mean, with the exception of the regression imputation. Imputing missing data did not significantly alter the conclusions regarding which factors were associated with variations in CES-D scores. Since imputation has the potential to increase statistical power, researchers dealing with missing CES-D scores should consider imputing missing data.

Keywords: Imputation; Missing data; Psychiatric rating scales; Quality of life; Depression

1. Introduction

One area of the social and medical sciences with significant patient reported outcome implications is the area of depression. However, just as with the rest of the medical and social sciences, missing data are widespread in studies that include patients' self-reports of depression or depression symptoms. In the worst case, authors of published studies may not even mention if they encountered missing data. For example, the Center for Epidemiologic Studies Depression (CES-D) Scale[1] is a widely used indicator of depression in population-based studies. Authors of 10 of 25 recent, randomly selected articles using the CES-D did not mention missing data (the list of articles is available from the author upon request). In other cases, the authors simply mentioned that data were missing and then conducted the analysis using only the complete cases. Complete case analysis, whereby only cases with complete data are included in the analysis, has its own concerns and limitations.[2] In the case of the CES-D, authors of 11 of the 25 articles noted missing data, but used only the complete cases in their analyses with no assessment of potential biases. In other words, 85% of recent studies using the CES-D for population-based estimations and clinical outcome studies either did not mention if data were missing or eliminated all persons with missing data, potentially introducing bias and influencing the results. Given the prevalence of missing data, researchers must be better prepared to address problems that arise when data are missing.

1.1. Purposes of the study

This investigation had 2 purposes. The primary purpose was to evaluate the results of the use of 4 single-value imputation strategies for the measurement of depressive symptoms as measured by the CES-D Scale and study the impact of imputing missing data on the conclusions. Information regarding the similarities and differences among these imputation techniques in the conclusions drawn will be useful to researchers who are struggling with missing CES-D data and will provide them with a starting point for conducting their own systematic comparison. It may provide them with some confidence regarding their own study's interpretations when data are missing or encourage them to implement a strategy to ameliorate the power and bias consequences of missing data.[3] However, before embarking on the primary purpose, a secondary purpose was to provide a brief rationale and a review of the methods used to impute the missing data in this study. A number of articles and textbooks are available for a complete review of the issue of missing data.[2,4-9] We focused on the rationale and a brief review on the issues we faced in making decisions regarding whether to impute missing data in our own work with the CES-D in this investigation.[10,11]

1.2. Considerations in selecting imputation strategies

A plethora of issues must be considered when making decisions regarding whether missing data should be imputed and the techniques that should be used. As mentioned earlier, this article cannot completely review these issues. Some of the more important issues include single versus multiple imputation, normal versus nonnormal data issues (eg, categorical data), models that include mixed levels of measurement (ie, ordinal, ratio), and nonparametric techniques (eg, hot deck) versus parametric techniques (eg, multiple imputation). The reader is directed to more complete discussions of these issues in published authoritative resources.[2,4-9]

However, in all studies the first consideration in determining whether imputation is appropriate is to examine the source and reason for the missing data.[2] In some cases, the reason for missing data from a study subject is that he or she did not return the survey. This situation is sometimes referred to as missing "units." This article does not examine strategies for imputing or assessing the reasons or effect of missing unit nonresponse. In other instances, the data collection instrument may have been returned but may be missing specific items within the survey. Accurately ascertaining the reasons the data are missing is important to any decision regarding whether to impute data, regardless of whether the data involve missing units or items. In either case, the primary issue is whether the nonresponse is "ignorable."

1.3. Missing completely at random—ignorable nonresponse

An item or questionnaire is said to be missing completely at random (MCAR) if the missing assessments are independent of all previous, current, and future assessments had they been observed.[12] Thus, when an item or questionnaire is MCAR, cases with complete data are indistinguishable from cases with incomplete data.[13] Missing data due to accidental death, the respondent moving out of town, or staff forgetting to administer the instrument due to a random incident may yield an item or questionnaire MCAR.[14]

1.4. Missing at random—ignorable nonresponse

Data are termed missing at random (MAR) if the missing values of the dependent variable depend on the independent variable(s) but do not depend on the dependent variable.[15] Thus, when an item or questionnaire is MAR, cases with incomplete data are different from cases with complete data, but the pattern of missing data is traceable or predictable from other variables in the database rather than being due to the specific variable on which the data are missing.[4] The cause of the missing data is some external influence. For MAR, the probability of having a missing questionnaire may depend on previous scores, but must be independent of both current and future scores.[2,16] For example, if patients with a poorer quality-of-life score at the previous assessment are more likely to have a missing questionnaire at the current assessment, then the missing questionnaire is not MCAR, but instead is MAR.[2,16] MAR is a less stringent assumption than MCAR.

1.5. Missing not at random—nonignorable nonresponse

An item or questionnaire is said to be missing not at random (MNAR) if the missing values are not predictable from other variables in the database but are predictable from the variable on which the data are missing.[2,16] For example, if a participant in a depression study does not return a questionnaire because of increased feelings of depression, these data are MNAR. Thus, data are MNAR if the data missingness is explainable and only explainable by the very variables on which the data are missing. Bias associated with the missing unit or items is "nonignorable" in this situation. Whenever the probability of dropout depends on at least one unobserved score, then the process is termed MNAR.[2,16] With MNAR the probability of having a missing questionnaire depends on scores in current and future unobserved assessments.[2,16] With MNAR, the missingness may be influenced by values on the missing variable and on its relationships with other variables. In this case, strategies are available to assess the degree of bias, including instrumental variables,[17,18] propensity score calibration,[19,20] and Heckman's correction.[21] Examples of data that are MNAR in a typical clinical trial of a pharmaceutical agent are data associated with increased toxicity, progressive disease, or death.[14]

1.6. General purpose approaches: deletion of observations

Next, after determining the reason for the missing data and whether it is ignorable or nonignorable, the decision regarding whether or not to impute missing data and what strategy to use to impute the data should be examined (Fig. 1).[6] Until recently, the complete case approach to analyzing the data was the standard, whether the reasons for missing data were ignorable or not. The complete case approach discards all cases with unit or item nonresponse, and the analysis is conducted using only the complete cases whether partial information is available or not. For example, in the case of the CES-D, all of an individual subject's responses would be discarded if the response to one item was missing, even though the subject had completed the other 19 items of the scale. However, as noted earlier, the complete case approach requires the missing data to be MCAR to yield unbiased results. The complete case approach is best suited to situations where the extent of missing data is small, the sample size is large enough to allow for the deletion of the cases with missing data without severely

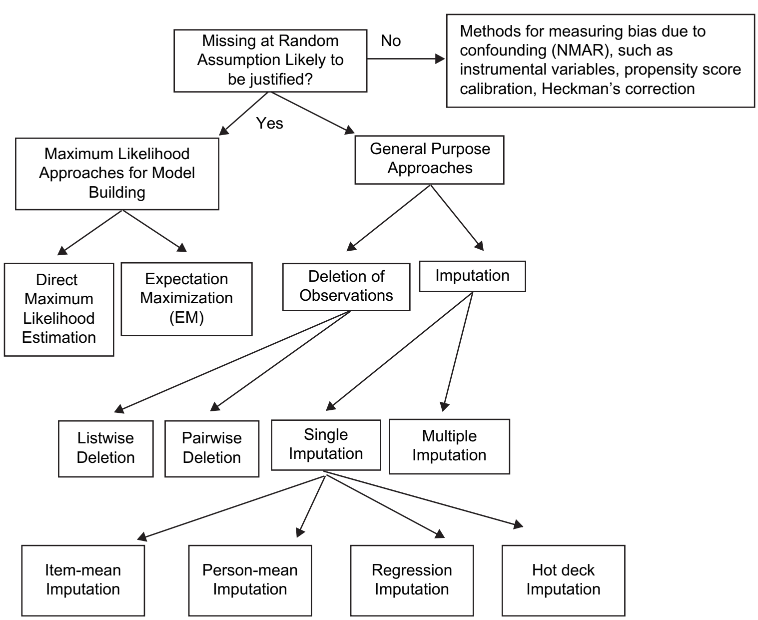

Fig. 1. Options for handling missing data. Adapted from Croy and Novins.[6]

jeopardizing statistical power, and the relationships in the data are strong so as not to be biased by any missing data process.[12,15] The complete case approach is simple[2,16] and requires little effort from the researcher because it is the default in most statistical software packages. A smaller sample size is the most apparent disadvantage. With the complete case approach, the remaining sample size may be inadequate for conducting statistical analyses with sufficient power. Another disadvantage of the complete case approach is the loss of other information for an individual respondent on all other variables when incomplete cases are discarded due to item nonresponse on a single variable.[2,16] Finally, unless the data are MCAR (ie, a random subsample of all cases), discarding incomplete cases will result in a biased sample, such as possible systematic differences between the complete cases and the incomplete cases.[22] If the data are not MCAR, then other means of handling the missing data should be used, including data imputation.

1.7. General purpose approaches: missing data imputation

Various single-value imputation approaches, mixed effects and pattern mixture models, and multiple imputation methods are available to the researcher (Fig. 1).[6] Imputation is the process of estimating the missing value(s) based on valid values of other variables and/or cases in the sample.[15] The objective of imputation is to use known relationships that can be identified in the valid values of the sample to assist in estimating the values that are missing.[15] The complete case approach (listwise deletion) removes the problem of missing data by removing these observations, whereas the imputation approach removes the problem of missing data by filling in values for the unknown or missing data.[16] Data missing at random, either MCAR or MAR, are more amenable to imputation techniques.[23] Standard analytic methods can be used once missing values have been imputed.

A variety of imputation techniques have been proposed in the literature. Some techniques are mathematically and computationally difficult to apply because missing values are estimated based on valid values of other variables and/or cases in the sample.[15] Methods of imputation include item-mean imputation, person-mean imputation, cold-deck imputation, hot-deck imputation, last value carried forward imputation, worst value imputation, best value imputation, random imputation, regression imputation, Monte Carlo and other stochastic methods, and multiple imputation plus their combinations.

Although a number of single-value and multiple imputation strategies are available,[6] for this investigation we chose to focus on 4 single-value imputation techniques, including item-mean, person-mean, regression, and hot-deck imputation, because we determined the data to be MAR.[24] These 4 imputation techniques vary in terms of complexity, but all are commonly used and were used in the present study for a number of reasons. Item-mean imputation was chosen because it is the most commonly used imputation technique for missing data on the CES-D and is included in most statistical software. Person-mean imputation was chosen because it is the only other imputation technique used with missing data on the CES-D to date. Regression imputation was chosen because this technique has not been used with missing CES-D data, and it purposefully uses existing relationships to predict the missing value, whereas the other 3 imputation techniques chosen do not. Finally, metric-matching hot-deck imputation was chosen because it is used with ignorable nonresponse (MAR) and when predictor variables are both categorical and metric.[5]

1.8. Description of the 4 single-value imputation methods

Item-mean imputation replaces an individual's missing values on an item with the mean for that item calculated by using the scores of all study respondents who completed that item.[23] An advantage of item-mean imputation is its simplicity with respect to its application.[15] Very little programming is required to use this imputation technique. Compared with other imputation methods, its major disadvantage is that the actual distribution of values is distorted by substituting the mean for all missing values, thereby causing the variance to be artificially reduced.[3,16] Another disadvantage is that it depresses the observed correlations because all missing values are replaced by a constant.[15] Given these disadvantages, item-mean imputation may be prone to producing biased estimates of the missing values.[7]

Person-mean imputation requires substitution of the mean of all of an *individual's* completed items for those items that were not completed on a given scale.[23] This differs from item-mean where the mean response of the *whole sample* that responded to the item is substituted. Person-mean imputation could result in different substitutions for each person with missing items. On the plus side, because it does not substitute a constant value, it does not artificially reduce the measure's variability and is less likely to attenuate the correlation. A disadvantage is that it tends to inflate the reliability estimates as the number of missing items increases.[23] However, when the numbers of either respondents with missing items or items missing within scales are 20% or less, both item-mean imputation and person-mean imputation provide good estimates of the reliability of measures.[23]

Regression imputation substitutes a predicted value based on the regression of other variables on the missing variable from individuals with complete data. It assumes that other variables selected from the data are related to the missing variables.[23] For example, when applying regression imputation to a missing value in a depression scale, the depression item is regressed on determinants of depression such as age and sex. Its major advantage is an unbiased point estimate of the missing value. The resulting rectangular data set is larger, and the bias in the depression score variable is reduced because the multivariate distribution of known variables is used in determining estimates for missing values.[22] Although regression imputation has the appeal of using relationships already existing in the sample as the basis of prediction, it also has disadvantages. First, it reinforces already existing relationships in the data; thus, the resulting data become more characteristic of the sample and less generalizable to the population. Second, regression imputation assumes that the variable with missing data has substantial correlations with the other variables, an assumption that may or may not hold true.

Finally, hot-deck imputation selects a value from similar respondents with completed items, usually within the same data (vs cold-deck, which uses data from another source using the same item), and substitutes the selected value for the respondent's missing value.[16] The "deck" refers to responses from those with complete items from which the researcher may select a value.[16,25] The value may be selected simply at random or by using an elaborate scheme, such as selecting a value from only those respondents who have similar characteristics, such as sex, age, or treatment group.[25] For example, if a 35-year-old woman was missing a data point on item 1, a matrix of values of all 35-year-old female respondents with completed data would be created. A random value would be selected from the matrix of values, and that value would be substituted for the missing item. One advantage of the hot-deck method is its conceptual simplicity. Another advantage is that it maintains the proper measurement level of variables. For example, item-mean imputation and person-mean imputation substitute an "average" for missing nominal variables, such as sex or race. Among several disadvantages of the hot-deck method is the extensive programming required to implement this technique if it is not randomly selected from the deck because it requires customized syntax to perform the selection of values from similar respondents. Another disadvantage is the ambiguity surrounding the definition of "similar," which may vary from one researcher to another, creating uncertainty around the results. Carefully articulated definitions of "similar" substantially reduce the ambiguity. In our study, hot-deck values were obtained by randomly selecting a single value from a matrix of persons with similar health status, sex, education, race, and age and then substituting that value for the missing value of the respondent in question.

We acknowledge the availability of numerous methods of imputing data, including single-value and multiple imputation techniques. However, as our primary purpose was to compare (1) various imputation methods with the complete case analysis and (2) a single-value imputation method (person-mean) used in the CES-D literature with other alternatives, we chose to compare 4 single-value item imputation models with each other and with the complete case approach for data we considered to be MAR.[24] We chose not to conduct multiple imputations or investigate potential biases because we considered the data to be MAR and the reasons for missing data to be ignorable. Moreover, we chose to heed the advice of Rubin,[26] who stated that "my own assessment is that unless a user has the resources … the iterative versions of software for creating multiple imputations are not yet ready for reliable applications by the typical user." The specific objectives of this investigation were to:

(1) provide an estimate of bias for each imputation technique with known values from data engineered to be missing completely at random;
(2) determine whether different CES-D scores were obtained from item-mean, person-mean, regression, and hot-deck imputation techniques and whether they differed from the CES-D score obtained from complete cases; and
(3) determine whether the variables that predicted the CES-D scores were the same when using complete cases and each of the 4 imputation techniques.

2. Methods

2.1. Study design

The Study of Antihypertensive Drugs and Depressive Symptoms (SADD-Sx)[10,11] is a substudy of INVEST.[27,28] Briefly, INVEST was a randomized, open-label, blinded endpoint study of 22,576 patients with hypertension and coronary artery disease (CAD) aged $\geq$50 years conducted from September 1997 to February 2003. Patients were randomized to antihypertensive treatment with either a verapamil SR- or atenolol-based strategy to achieve blood pressure (BP) control according to the sixth report of the Joint National Committee on Prevention, Detection, Evaluation, and Treatment of High Blood Pressure (JNC VI).[29] Most subjects had angina pectoris and/or a remote myocardial infarction and needed hypertension drug therapy. A detailed description of INVEST and its patients, randomization procedures, intervention, and implementation has previously been reported.[27,28]

Once the subjects were enrolled in the INVEST study, their names and addresses were automatically forwarded to the SADD-Sx study center via a computerized system. The purpose of the SADD-Sx study was to compare depressive symptoms among patients assigned to the verapamil-SR- and atenolol-based BP treatment strategies. Consecutively randomized INVEST patients residing in the US were mailed SADD-Sx questionnaires, including the CES-D, which measures current depressive symptoms, between April 1, 1999, and October 31, 1999 (N = 2,317). Patients were mailed surveys the next working day following randomization. Subjects were provided with self-addressed, postage-paid envelopes to return the survey. If surveys were not returned within 10 working days, a second survey was mailed. If patients did not respond to the second survey, no further attempts were made to contact them for the purposes of SADD-Sx. Follow-up surveys were sent 6 weeks, 6 months, and 1 year after enrollment. Only the results of the initial survey are reported.

Information on depressive symptoms, education level, living status, and psychiatric history was obtained by the survey. SADD-Sx was conducted according to the principles of the Declaration of Helsinki. The University of Florida Institutional Review Board approved the study protocol.

2.2. Self-rated depressive symptoms

The CES-D Scale was used to assess depressive symptoms.[1] The CES-D is a 20-item self-reported rating scale designed to measure current levels of depressive symptoms. It is widely used and is a reliable[1,30] and valid instrument.[1,31-33] In the CES-D, respondents are asked how often they had experienced each of the 20 symptoms during the past week. Responses are rated on a 4-point scale: 0, rarely or none of the time (less than once a week); 1, some or a little of the time (1-2 days a week); 2, occasionally or a moderate amount of time (3-4 days a week); and 3, most or all of the time (5-7 days a week). A sum of the responses to all 20 items constitutes a single summary score for each respondent. Higher scores are indicative of greater depressive symptoms. Scores range from 0 to 60. As a screening tool, the CES-D has been used to estimate the presence of clinically significant depression.[31,33]

2.3. Predictor variables

Several demographic, social, and health status variables have been found to be associated with depression, including race, sex, education, living arrangement, self-reported health status, history of depression, family history of depression, age, and high BP treatment. A combination of these variables representing the subjects' characteristics was used (1) to predict the item value in the regression imputation method; (2) as the characteristics used to create the hot deck for that imputation technique; and (3) to predict the level of depressive symptoms after the CES-D scores were imputed to compare the obtained regression models based on the different imputation techniques in objective 2 and objective 3.

2.3.1. Race

Kaelber et al[34] found a somewhat higher prevalence of depression among whites, followed by Hispanics and blacks. Others have found the prevalence of depression to be greatest among Hispanics and least among whites, with blacks and others in-between.[35] In our own work, we found that respondents describing themselves as white were significantly different on the CES-D from respondents who described themselves as black, Asian, Hispanic, or other.[10] However, respondents from the multiracial groups were not different from one another. Consequently, we dichotomized the race variable, and respondents who described themselves as white were coded as 1; respondents who were black, Asian, Hispanic, or other were coded as a 0 for this variable.

2.3.2. Sex

Females have been found to be nearly twice as likely to be depressed as males.[35-37] Respondents who were male were coded as 1; respondents who were female were coded as a 0 for this variable.

2.3.3. Education

Lower education level has been found to be associated with increased levels of depression among women living in urban areas.[38] Education level and physical illness have been associated with an increased risk of depression for those aged 18 to 39.[39] The following request was made to the respondents: "Please tell us the highest grade in school that you completed." Respondents had the following categories to choose from: (1) eighth grade or less; (2) some high school; (3) high school graduate; (4) some college; and (5) college graduate. Respondents who stated that they were a high school graduate, had some college, or were a college graduate were coded as 1, and respondents who stated that they had eighth grade or less or some high school education were coded as 0. Respondents who left this item blank were coded as 1, which was the modal category.

2.3.4. Living arrangements

Living alone has been found to be associated with a more than twofold increase in relative risk of depression.[38,40] Respondents were asked, "What is your current living arrangement?" Respondents had the following categories to choose from: (1) live alone, (2) live with spouse, (3) live with children, (4) live with another relative, and (5) live with nonrelative. Respondents who reported living alone were coded as 1; respondents who reported living with their spouse, children, or relatives/nonrelatives were coded as a 0 for this variable. Respondents who left this item blank were coded as 0.

2.3.5. Self-reported health status

Several studies have found a positive relationship between self-reported health status and depressive symptoms as measured by the CES-D.[41,42] Respondents were asked, "In general, would you say that your health is 1) excellent; 2) very good; 3) good; 4) fair; or 5) poor?" Respondents who stated that they were in excellent, very good, or good health were coded as 1, and respondents who stated that they were in fair or poor health were coded as 0. All respondents completed this item.

2.3.6. History of depression

Previous episodes of major depressive symptoms strongly predict the probability of relapse. Patients who have recovered from an episode of major depressive symptoms have greater lifetime risk of another major depressive episode.[43,44] Respondents were asked the following question: "Has a medical doctor or psychiatrist ever told you that you were depressed?" Respondents whose doctor or psychiatrist told them that they were depressed were coded as 1, and respondents whose doctor or psychiatrist did not tell them that they were depressed were coded as 0. Respondents who left this item blank were coded as 0.

2.3.7. Family history of depression

Family history of depression increases the risk of depression.[45-47] Respondents were asked, "Has anyone in your family been treated for depression by a doctor or psychiatrist?" Respondents who stated that someone in their family had been treated for depression were coded as 1, and respondents who stated that no one in their family had been treated for depression were coded as 0. Respondents who left this item blank were coded as 0.

2.3.8. Age

The literature is inconsistent concerning the effect of age on depressive symptoms. Some studies have found younger age to be associated with depression,[48,49] whereas others have found older persons to be more likely to exhibit depressive symptoms.[50,51] Age was not missing for any respondent in this study and was entered into the models as a continuous variable.

2.3.9. High BP treatment

The primary objective of the SADD-Sx was to evaluate whether there were significant differences between atenolol-based and verapamil-SR-based high BP treatment strategies with regard to the levels of depressive symptoms.[10,11] Respondents who were being treated with the verapamil-SR-led strategy were assigned a 1. Respondents who were being treated with the atenolol-led treatment strategy were assigned a 0. Treatment assignment information was not missing for any respondent.

2.4. Statistical analysis

Demographic characteristics of respondents who completed all 20 items were compared with those of respondents who did not complete all 20 items using independent *t* tests and chi-square. Two analyses for each imputation method were conducted to simulate these conditions. We compared imputation methods for (1) respondents who completed at least 16 of the 20 CES-D items (ie, excluding those completing 15 items or less) and (2) all respondents, no matter the number of completed CES-D items.

Next, we evaluated the 4 imputation strategies over these same 2 conditions (ie, imputing missing data for a limited number of items and excluding those who exceed the limit vs imputing all missing data). We compared the 4 imputation results with the complete case analytic strategy, which excludes all respondents with any missing data. The complete case strategy is the most frequent strategy used in the CES-D literature to handle subjects' missing responses to items.

Next, we evaluated the assertion that it is appropriate to impute data MCAR and evaluated the results obtained from the 4 imputation strategies. We constructed a data set that mirrored the original data with respect to the number of respondents missing one item, the number of respondents missing 2 items, and so forth (Table 1). SAS code was written so that the engineered, simulated MCAR database would randomly generate identical proportions of respondents missing 1 CES-D item, 2 CES-D items, and 3 to 20 CES-D items. Consequently, the resulting data set mirrored the actual patterns of missing CES-D items in the original data and the simulated data were engineered to be MCAR. This process was repeated a sufficient number of times so that 100 data sets that simulated data MCAR were constructed. Each of

Table 1
Number and percentage of respondents missing 0 to 20 Center for Epidemiologic Studies Depression (CES-D) Scale items at baseline

	Baseline (N = 1,635)	
Number of missing CES-D items	Number	Percent[a]
Not missing any CES-D items (complete case)	1,357	83.0
Missing 1 CES-D item	143	8.8
Missing 2 CES-D items	53	3.2
Missing 3 CES-D items	14	0.9
Missing 4 CES-D items	12	0.7
Missing 5 CES-D items	4	0.2
Missing 6 CES-D items	3	0.2
Missing 7 CES-D items	3	0.2
Missing 8 CES-D items	3	0.2
Missing 9 CES-D items	1	0.1
Missing 10 CES-D items	2	0.1
Missing 11 CES-D items	1	0.1
Missing 12 CES-D items	0	0.0
Missing 13 CES-D items	4	0.2
Missing 14 CES-D items	1	0.1
Missing 15 CES-D items	1	0.1
Missing 16 CES-D items	1	0.1
Missing 17 CES-D items	2	0.1
Missing 18 CES-D items	0	0.0
Missing 19 CES-D items	2	0.1
Missing all 20 CES-D items	28	1.7

[a] May exceed 100% due to rounding error.

the 4 imputation techniques was applied to each of the 100 simulated data sets. The mean and standard deviation parameter estimates were calculated for the CES-D Scale from the results of each simulation, as well as an overall average of the mean scores and the standard deviation of the means for each imputation type. The results are combined and the parameter estimates are reported.[52] For example, the average and 95% confidence interval are reported as the estimates in this work.

Respondents who completed all 20 CES-D items at baseline comprised the starting point for the simulation data set (N = 1,357). The purpose of the simulation was not to compare the results obtained from the simulations with the original complete case data, but rather to provide an estimate of bias for each imputation technique from known values. In this case, the results obtained from the 4 imputations would be compared with the results obtained from the known values from the complete data to estimate the bias associated with the imputation technique.

Finally, we assessed the use of these single-value imputation strategies in data that are MAR. We calculated 5 multiple regression equations to imitate a crude multiple imputation strategy. In effect, multiple imputation summarizes the results of multiple analyses based on the responses of valid respondents and generates "reasonable hypothetical responses" for nonrespondents. Systematic, multiple imputation is fast becoming a major approach designed to obtain valid statistical inferences when faced with missing data, and it is expected to become the dominant approach in practice in the near future.[26] SAS and other statistical software include modules (eg, the MI procedure in SAS) to conduct multiple imputations. Multiple imputation is a method of generating multiple simulated values for each incomplete item, then iteratively analyzing data sets with each simulated value substituted in turn. The purpose is, arguably, to generate estimates that better reflect true variability and uncertainty in the data. Multiple imputation methods yield multiple imputed replicate data sets, each of which is analyzed in turn. Methods other than imputation for estimating and correcting for bias induced by data MNAR include instrumental variables,[17,18] propensity scoring calibration,[19,20] or Heckman correction.[21] However, a comparison of these correction techniques and the obtained bias estimates is beyond the scope of this investigation.

To complete our crude analysis, we computed *z* scores to determine statistically significant differences between each of the 4 imputed CES-D means and the complete case CES-D mean (objective 2). Next, to complete objective 3, four multiple regression models were run wherein four imputed CES-D means served as the dependent variable. A fifth regression model used the complete case CES-D score as the dependent variable. Each model controlled for race, sex, education, living arrangement, self-reported health status, history of depression, family history of depression, and treatment assignment. The magnitudes of the regression coefficients for the 4 models were not statistically compared, although they were visually inspected to ascertain similarity between the 4 imputed regression models. To the best of our knowledge, no statistical test is available to test the similarity of magnitude among models using the same population. The same analyses were conducted for 4 different time periods (at baseline and 6 weeks, 6 months, and 1 year after enrollment). The patterns of the results for the 4 time periods were virtually identical with regard to any interpretation.[24] Only the baseline results are reported. Analyses were conducted using SAS version 8.12.[53] The SAS programming used to conduct the different imputation analyses is available upon request.

3. Results

A total of 1,635 respondents returned the questionnaire, representing 71% of the SADD-Sx enrollees. Eighty-three percent of the respondents returning a questionnaire completed all 20 CES-D items; questionnaires returned by 14% of the respondents had between 1 and 4 CES-D items missing; and those returned by 3% had 5 or more CES-D items missing (see Table 1).

3.1. Demographic characteristics of respondents

Seventy-eight percent of the respondents were white; 55% were male; 66% graduated from high school; 75% lived with someone; 57% reported being in excellent, very good, or good health; 18% reported a history of depression; and 21% reported having a family history of depression (Table 2). The average age of respondents was 68. Just over 48% of the respondents were assigned to the atenolol-led treatment group, and nearly 52% were assigned to the verapamil-SR-led treatment group.

Respondents who completed 15 or fewer CES-D items were approximately 4 years older than respondents who completed 16 or more CES-D items (t test = 2.89, $P < .01$, Table 2). Respondents not completing all 20 CES-D items were more likely to report fair or poor health ($\chi^2 = 15.0$, $P < .001$); be female ($\chi^2 = 18.2$, $P < .001$); have less than a high school education ($\chi^2 = 23.9$, $P < .001$); and be nonwhite ($\chi^2 = 16.0$, $P < .001$). Respondents who did not complete all 20 CES-D items were approximately 2 years older ($t = 3.79$, $P < .01$).

3.2. Imputation of data MCAR

The complete case, ‘‘known’’ CES-D mean score was 14.06. The imputation strategies using the engineered data MCAR yielded an average CES-D score of 14.07 (95% CI = 14.06, 14.08) for the item-mean imputation; 14.07 (95% CI = 14.06, 14.08) for the person-mean imputation; 14.22 (95% CI = 14.21, 14.23) for the regression imputation; and 14.07 (95%

Table 2
Demographic characteristics of respondents completing (1) 16 or more Center for Epidemiologic Studies Depression (CES-D) Scale items versus 15 or fewer and (2) respondents completing all of the CES-D items versus having one or more missing items

	N of all respondents	Completed at least 16 CES-D items		Completed all 20 CES-D items	
Demographic characteristics	N (percent)	Percent no (N = 56)	Percent yes (N = 1,579)	Percent no (N = 278)	Percent yes (N = 1,357)
Race					
Nonwhite	358 (21.9)	32.1	21.5	31.0**	20.0**
White	1,277 (78.1)	67.9	78.5	69.0	809.0
Sex					
Female	728 (44.5)	51.8	44.3	56.1**	42.25**
Male	907 (55.5)	48.2	55.7	43.9	57.8
Education					
Less than high school graduate	563 (34.4)	46.4	34.0	47.1**	31.8**
Graduated high school or higher	1,072 (65.6)	53.6	66.0	52.9	68.2
Living arrangement					
Not alone	1,229 (75.2)	69.6	75.4	72.0	75.8
Alone	406 (24.8)	30.4	24.6	28.0	24.2
Health status					
Fair or Poor	699 (42.7)	44.6	42.7	53.2**	40.6**
Excellent, very good, or good	936 (57.3)	55.4	57.3	46.8	59.4
History of depression					
No	1,341 (82.0)	87.5	81.8	83.1	81.8
Yes	294 (18.0)	12.5	18.2	16.9	18.2
Family history of depression					
No	1,299 (79.4)	89.3	79.1	81.3	79.1
Yes	336 (20.6)	10.7	20.9	18.7	20.9
Treatment assignment					
Noncalcium channel blocker	790 (48.3)	51.8	48.2	50.0	48.0
Calcium channel blocker	845 (51.7)	48.2	51.8	50.0	52.0
Mean age (standard deviation)	68.2 (9.8)	71.9** (9.6)	68.1** (9.7)	70.2** (10.4)	67.8** (9.6)

$^{**}P < .01$.

CI = 14.06, 14.09) for the hot-deck imputation. All the imputed means were similar to the complete case mean, with the exception of the regression imputation. History of depression, family history of depression, sex, education, race, self-reported health status, living arrangement, and age predicted simulated CES-D score regardless of imputation technique (Table 3).

Table 3
Regression models predicting Center for Epidemiologic Studies Depression (CES-D) score using data missing completely at random from baseline survey

Independent variable	Complete case parameter estimate	Item-mean imputation parameter estimate	Person-mean imputation parameter estimate	Regression imputation parameter estimate	Hot-deck imputation parameter estimate
Intercept	27.67**	27.37**	28.01**	28.47**	28.48**
Race	−1.68**	−1.82**	−1.70*	−1.88**	−1.94**
Sex	−0.99	−1.04*	−1.11*	−1.05*	−1.17*
Education	−3.16**	−3.10**	−3.13**	−3.31**	−3.47**
Living arrangement	1.60**	1.50*	1.44*	1.61**	1.59**
Self-reported health status	−6.94**	−6.71**	−6.93**	−6.89**	−6.81**
History of depression	8.18**	7.81**	8.26**	8.11**	7.90**
Family history of depression	2.81**	2.79**	2.67**	2.85**	2.72**
Age	−0.11**	−0.10**	−0.11**	−0.11**	−0.11**
Treatment assignment	−0.23	−0.35	−0.36	−0.33	−0.36
F ratio	68.7	66.4	66.7	72.1	69.3
R^2 of model	.32	.31	.31	.33	.32

*$P < .05$.
**$P < .01$.

3.3. Differences in CES-D means obtained by each of the 4 imputation techniques

For respondents who were missing 0 to 4 CES-D items, the imputed CES-D means ranged from 14.58 to 14.68 (see Table 4). The lowest CES-D mean was obtained from hot-deck imputation and the highest with regression imputation. Each of the 4 imputed CES-D means was higher than the complete case CES-D mean of 14.06. None of the imputed means were significantly different from the complete case mean among respondents who were missing 0 to 4 CES-D items. No statistically significant differences were found between the 4 imputed means.

Among respondents who were missing any number (≥1) of CES-D items, the imputed CES-D means ranged from 14.66 to 14.85 (see Table 4). The lowest CES-D mean was obtained from item-mean imputation and the highest mean with regression imputation. Each of the imputed CES-D means was higher than the complete case CES-D mean. As before, none of the imputed means was significantly different from the complete case mean among those missing any number of CES-D items, nor were differences among the 4 imputed means statistically significant.

A sensitivity analysis was conducted to assess the stability of these findings. First, missing CES-D items were assigned the best possible outcome

Table 4
Mean and standard deviation of Center for Epidemiologic Studies Depression (CES-D) score by imputation technique for (1) respondents completing 16 or more CES-D items versus 15 or fewer and (2) respondents completing all of the CES-D items versus having one or more missing items

Imputation technique	Mean (1) (SD)[a]	*P* value of *z* score	Mean (2) (SD)[a]	*P* value of *z* score
Complete case mean				
(1) N = 1,357	14.06		14.06	
(2) N = 1,357	11.28[a]	—	11.28[a]	—
Item-mean imputation				
(1) N = 1,579	14.60		14.66	
(2) N = 1,635	11.24[a]	0.05	11.08[a]	0.05
Person-mean imputation				
(1) N = 1,579	14.65		14.81	
(2) N = 1,607	11.38[a]	0.05	11.49[a]	0.07
Regression imputation				
(1) N = 1,579	14.68		14.85	
(2) N = 1,635	11.30[a]	0.05	11.19[a]	0.07
Hot-deck imputation				
(1) N = 1,579	14.58		14.72	
(2) N = 1,635	11.24[a]	0.05	11.23[a]	0.06

[a] Standard deviation of mean score.

and the CES-D score was then calculated. Next, the missing CES-D items were assigned the worst possible outcome and then the CES-D score was calculated. These 2 imputed CES-D means were then compared with the complete case CES-D mean. Neither of these worst/best case imputed means was statistically significant from the complete case mean.

Finally, a frequent use of the CES-D in population-based and clinical research is to classify patients as being at high risk of current depression. Subjects with CES-D scores of $\geq$16 are considered to be at high risk of current depression. At baseline, among respondents missing 0 to 4 CES-D items, there was 99.2% agreement in risk-level classification among the 4 techniques. Only 13 of 1,579 respondents were classified differently by the 4 imputation techniques. Among respondents missing 0 to 19 CES-D items, there was 98.3% agreement in risk-level classification and only 27 of 1,607 respondents were classified differently by the 4 imputation techniques.

3.4. Regression models predicting the imputed CES-D scores

Among respondents who were missing 0 to 4 CES-D items, history of depression, family history of depression, sex, education, race, self-reported health status, living arrangement, and age were predictive of CES-D score regardless of which imputation technique was used (Table 5). Having

Table 5
Regression models predicting Center for Epidemiologic Studies Depression (CES-D) score using complete cases and each imputation technique among (1) respondents completing 16 or more CES-D items versus 15 or fewer and (2) respondents completing all of the CES-D items versus having one or more missing items

	Complete case parameter estimate	Item-mean imputation parameter estimate		Person-mean imputation parameter estimate		Regression imputation parameter estimate		Hot-deck imputation parameter estimate	
Independent variable	(1)	(1)	(2)	(1)	(2)	(1)	(2)	(1)	(2)
Intercept	27.67**	28.54**	27.91**	28.95**	28.01**	28.85**	28.57**	28.64**	28.21**
Race	−1.68**	−1.56**	−1.61**	−1.62**	−1.83**	−1.64**	−1.71**	−1.54**	−1.63**
Sex	−0.99	−1.25**	−1.27**	−1.25**	−1.28**	−1.26**	−1.33**	−1.21**	−1.31**
Education	−3.16**	−3.34**	−3.25**	−3.41**	−3.35**	−3.40**	−3.42**	−3.36**	−3.16**
Living arrangement	1.60**	1.35*	1.35**	1.34*	1.56**	1.36*	1.43*	1.40**	1.20*
Self-reported health status	−6.94**	−6.87**	−6.64**	−6.97**	−6.86**	−6.98**	−6.91**	−6.87**	−6.94**
History of depression	8.18**	8.04**	8.00	8.12**	8.16**	8.10**	8.05**	8.05**	7.99**
Family history of depression	2.81**	2.78**	2.71**	2.80**	2.68**	2.80**	2.67**	2.74**	2.71**
Age	−0.11**	−0.12**	−0.11**	−0.12**	−0.11**	−0.12**	−0.11**	−0.12**	−0.11**
Treatment assignment	−0.23	−0.23	−0.28	−0.24	−0.29	−0.24	−0.31	−0.25	−0.33
F ratio	68.7**	83.2**	83.4**	83.3**	80.2**	85.2**	87.7**	83.2**	83.6**
R^2	.32	.32	.32	.32	.31	.33	.33	.32	.32

** $P < .01$.

a history of depression, having a family history of depression, and living alone were all associated with a higher CES-D score, whereas being older, male, a high school graduate, white, and healthy by self-report were all associated with a lower CES-D score.

Among respondents with no limits to the number of missing CES-D items, the same variables were predictive of CES-D score regardless of which imputation technique was used (Table 5). Having a history of depression, having a family history of depression, and living alone were all associated with a higher CES-D score, whereas being older, male, a high school graduate, white, and healthy by self-report were all associated with a lower CES-D score.

3.5. Comparison of complete case regression models and imputed CES-D score models

The complete case regression model was similar, with respect to both the predictors of CES-D score and the direction of these predictors, to the 4 regression models using imputation for respondents missing 0 to 4 CES-D items (Table 5). However, the complete case regression did not reveal sex as a predictor of CES-D score as it was in each of the 4 imputed models.

Among respondents missing any number of CES-D items, the complete case regression model was similar, with respect to both the predictors of CES-D score and the direction of these predictors, to the 4 regression models using imputation (Table 5). In the complete case model, history of depression, family history of depression, education, race, self-reported health status, living arrangement, and age were predictors of CES-D score, just as they were in the imputed models. As before, the complete case model did not reveal sex as a predictor of CES-D score as it was for each of the 4 imputed models.

4. Discussion

Demographically, respondents completing at least 16 CES-D items were similar to those completing 15 or fewer CES-D items, with the exception of age. Respondents who completed 15 or fewer CES-D items were generally older. Because there was a significant demographic difference, it is likely that the missing CES-D data are not MCAR, but are more likely to be MAR. When the demographic characteristics of respondents completing all 20 CES-D items were compared with the demographic characteristics of respondents not completing all 20 CES-D items, several demographic differences were revealed. Respondents who did not complete all 20 CES-D items were older, were female, rated their health as poorer, were nonwhite, and were less likely to have completed high school. These findings indicate that the data were not MCAR and were more likely to be MAR. The complete case approach, which deletes all cases with any missing data, requires the missing data to be MCAR to yield completely unbiased results that are generalizable to the population. When the data are not MCAR, as in our case, imputation of missing CES-D data should be considered, rather than simply using complete case analysis. Additional analyses, not reported herein, were conducted[24] that indicated that the data were MAR versus MNAR. Consequently, bias estimation strategies, such as instrumental variable,[17,18] propensity score calibration,[19,20] or Heckman's correction[21] analyses, were not conducted.

4.1. Imputation of data MCAR

When comparing the imputed CES-D scores from the single-value imputation methods with the actual CES-D score with the data engineered to be MCAR, the item imputation, person imputation, and the hot-deck imputation yielded means that were within .01 of the actual observed mean and were not statistically significantly different from one another, as predicted if the data were MCAR. The difference in mean score between the known sample and the regression mean imputation was the most disparate but still within normal limits. Our finding supports the finding of another simulation study using 1,000 independent block runs that found person-mean and hot-deck imputations to be more efficient than regression imputation.[54] This provides a degree of confidence that if the data are deemed to be MCAR, one of these single-value imputation strategies could be used to increase the sample size and study power.

4.2. Imputed CES-D means and complete case CES-D means

When we applied each of the imputation techniques to the original data, the analysis yielded CES-D scores that were similar for respondents missing 0 to 4 CES-D items. Despite the finding that the imputed means were similar, some interesting trends emerged. Among respondents who were missing 0 to 4 CES-D items, hot-deck imputation yielded the lowest CES-D mean score and regression imputation yielded the highest CES-D mean score. The findings were similar for regression imputation among those completing 20 CES-D items versus those missing any number of CES-D items (highest CES-D score), but item imputation resulted in the lowest average CES-D score in this instance. Regression imputation may yield a slightly higher CES-D score, due to the regressors that were chosen in this investigation. If different predictor variables were used to impute missing data, perhaps the CES-D score obtained by regression imputation would be closer to the CES-D scores obtained by the remaining imputation techniques. Furthermore, it is also possible that the CES-D score obtained by regression imputation may be higher because of unmeasured variables or imputed predictor variables. Even though regression imputation yielded slightly higher

CES-D scores than the other techniques, it is important to remember that the CES-D mean obtained by regression imputation was not statistically different from the CES-D means obtained by the other 3 imputation techniques. Because the imputed CES-D means were quite similar to each other, the imputation technique did not have a profound effect on the CES-D mean. Given this finding, researchers are left with a wider range of imputation options.[55-57] The choice of which imputation technique to use is up to the discretion of the researcher; however, researchers should use an imputation technique that they are comfortable with both methodologically and programmatically.[26] Before using an imputation technique, researchers must weigh the advantages and disadvantages of each technique. Because researchers are provided with quite an array of imputation techniques, it is no longer appropriate to use complete case analysis without at least examining imputation options and their effects.

The 4 imputed CES-D means were then compared with the complete case result. In our analyses, each of the imputed CES-D means were not significantly different from the complete case CES-D mean. Therefore, conclusions of the previous studies using the complete case CES-D mean would remain relatively unchallenged by the findings of the current investigation. However, if the imputed CES-D means were different from the complete case CES-D mean in this investigation, then doubt could have been cast on the conclusions of previous studies that used only the complete cases to calculate the CES-D mean. Even though the imputed CES-D means did not differ from the complete case CES-D mean in this investigation, researchers should nonetheless use an imputation technique with respect to missing data to ascertain that using the complete case analysis does not impart selective bias into the study conclusions. In addition, imputation would provide additional observations for analysis; and these additional observations may provide the power necessary to detect a statistically significant difference. Thus, findings that were not significant because of insufficient power when using only the complete cases may become statistically significant after increasing the sample size with imputation.

4.3. Independent variables predicting CES-D score in the 4 imputed models

The next objective was to assess whether predictors of CES-D scores differed with different methods of imputing the dependent CES-D score. Among respondents missing 0 to 4 CES-D items, history of depression, family history of depression, sex, education, race, living arrangement, self-reported health status, and age were significant predictors regardless of the technique used to impute the missing CES-D data. Among respondents missing any number of CES-D items, the predictors were similar regardless of whether item-mean, person-mean, regression, or hot-deck imputation was used. After imputing the missing items and determining the CES-D score, the predictors of the imputed CES-D scores were found to be the same as predictors from previous studies. This finding provides some assurance that imputing missing data in this instance did not seriously bias the interpretation. Because the predictors of CES-D score were similar regardless of how the respondents were grouped (eg, 0 to 4 missing CES-D items or any number of missing CES-D items), researchers may consider imputing all missing CES-D, thereby increasing a study's power, rather than discarding respondents with 5 or more missing items before analysis.

4.4. Comparing complete case model predictors with imputed CES-D score predictors

The complete case model was similar to the 4 imputed models, differing by one predictor—sex. Sex was not a significant predictor of CES-D in the complete case model but was a significant predictor in each of the 4 imputed models. Previous literature has demonstrated that sex is a strong predictor of CES-D score. It is therefore perplexing as to why sex was not significant. The results obtained in this investigation may have been different if the incidence of missing data in this investigation was higher. It is possible that a higher incidence of missing data would yield imputation estimates that were dissimilar. Although agreement can be noted between imputation techniques among both groups of respondents, the true classification of depressive symptoms remains unknown among this population.

Limitations to this study exist that should be kept in mind when interpreting the results. First, the "deck" used for the hot-deck imputation potentially has a significant effect on the results. Given the number of subjects included in the study, there is nearly an infinite number of decks possible and a different deck may yield different results with each iteration. The fact that hot-deck imputation, for the most part, yielded the lowest CES-D mean among respondents may be attributed, at least in part, to the deck that was used in this investigation. Without creating many decks and comparing and contrasting the results obtained by these decks, one cannot be certain that the deck used in this investigation was representative of all possible decks created from the data. Another potential influence on the results of the regression imputation is the selection of the predictor variables. The regression imputation consistently resulted in the highest CES-D score. An important variable may have been inadvertently omitted in this investigation (eg, income, physical disabilities). Despite including 9 variables previously found to be associated with depression in each of the regression models, it is possible that an additional variable may have been overlooked or simply not measured in this investigation. Hence, the imputed CES-D score obtained from the regression imputation could have been very different if another combination of variables were included as predictors. This consistent finding regarding regression imputation causes us 2 concerns: (1) that regression single-value imputation is an inefficient means of imputing data if the reasons for missing data are ignorable compared with the other methods or (2) our initial assertion that the data were MAR was not warranted and incorporating strategies to evaluate confounding bias were needed. Regression imputation may reinforce relationships already in the data, and bias was inserted into the results because we imputed data for missing data in the predictor variables in the regression imputation. As the predictor data were not MCAR, this may have influenced the outcome of the prediction of the CES-D score. This is a plausible reason for it being consistently higher than the score given by the other methods. Because of these 2 reasons, the findings may be more characteristic of the sample and less generalizable to the population. Next, the incidence of missing data was low in this investigation. The effects of imputation on CES-D score could be larger or smaller given more or less missing data. It would be of great interest to simulate different levels of missing data and then reexamine the effects that these imputation techniques have on both the CES-D mean and the factor structure. We also did not use multiple imputation strategies; instead, we chose to limit our comparisons to multiple, single-value imputation strategies. Additional comparisons using multiple imputation techniques and comparisons of the results with strategies designed to correct effect estimates for unmeasured confounding would be of value for an even more complete assessment of missing data. Finally, analyses were conducted to assess whether specific missing items within surveys that were returned were associated with the depression scores and served as the basis for our assessment that the data were MAR.[24] However, we did not conduct analyses to determine whether survey nonresponse was associated with depression or depressive symptoms. Other strategies, such as instrumental variable,[17,18] propensity score calibration,[19,20] and Heckman's correction,[21] could be used to estimate the effect of missing units. However, in this analysis, we do not impute the missing units and acknowledge that we only are imputing missing items from surveys that were returned.

Future work on missing data imputation should test the influence of imputation methods for longitudinal data. For example, are methods other than the last value carried forward imputation technique appropriate for respondents missing data who had a previous response to an item? For example, a recent study found that last value carried forward and standardized score methods performed poorly compared with closest match imputation and, especially, listwise deletion.[58] Similarly, future studies using the hot-deck imputation technique with the SADD-Sx data could use different decks rather than relying on the results from a single deck.

5. Conclusions

This investigation showed that the 4 imputation techniques yielded similar CES-D mean scores that did not differ significantly from the complete case CES-D mean. It also demonstrated that the predictors of CES-D score were similar regardless of imputation technique used and, with the exception of sex, were the same as the predictors of complete case CES-D score. Unfortunately, there is no one universal solution to the problem of missing data. The results from this investigation indicate that using imputation would not significantly alter the conclusions for our data. Given this fact, and the fact that using imputation has the potential to increase the power of the investigation, researchers should carefully consider imputing missing data as one strategy. The imputation technique did not seem to matter in this case, although cautions about the structure of the data, reasons for missing data, and ignorable versus nonignorable missing data should still be heeded.[2] The problems associated with missing data are by no means trivial and, although some solutions have begun to appear, a great deal of work remains.

References

1. Radloff L. The CES-D scale: a self-report depression scale for research in the general population. *Appl Psychol Meas*. 1977;1:385–401.
2. Piogett TD. A review of methods for missing data. *Educ Res Eval*. 2001;7:353–383.
3. Fayers PM, Curran D, Machin D. Incomplete quality of life data in randomized trials: missing items. *Stat Med*. 1998;17:679–696.
4. Allison PD, Allison SI. *Missing Data*. Thousand Oaks, CA: Sage Publications Inc.; 2001.
5. Rubin DB. *Multiple Imputation for Non-response in Surveys*. New York: JW Wiley & Sons; 1987.
6. Croy CD, Novins DK. Methods for addressing missing data in psychiatric and developmental research. *J Am Acad Child Adolesc Psychiatry*. 2005 Dec;44:1230–1240.
7. Little RJA, Rubin DB. *Statistical Analysis With Missing Data*. (2nd ed). New York: John Wiley and Sons; 2002.
8. Schafer JL. *Analysis of incomplete multivariate data*. In: *Chapman & Hall Series Monographs on Statistics and Applied Probability*, vol. 72. London: Chapman & Hall; 1997.
9. Schafer JL. Multiple imputation: a primer. *Stat Methods Med Res*. 1999;8:3–15.
10. Ried LD, Tueth MJ, Handberg E, Kupfer S, Pepine CJ. Antihypertensive drugs and depressive symptoms (SADD-Sx) in patients treated with a calcium antagonist-based or β-blocker–based hypertension strategy in the INternational Verapamil SR-Trandolapril STudy (INVEST). *Psychosom Med*. 2005;67:398–406.
11. Ried LD, Tueth MJ, Taylor M, Sauer B, Lopez L, Pepine CJ. Comparing depressive symptoms of patients at high risk for depression after antihypertensive treatment with verapamil-led versus atenolol-led strategies. *Ann Pharmacother* 2006; (published online, March 28, 2006).
12. Troxel AB, Fairclough DL, Curran D, Hahn EA. Statistical analysis of quality of life with missing data in cancer clinical trials. *Stat Med*. 1998;17:653–666.
13. Heitjan DF. Annotation: what can be done about missing data? approaches to imputation. *Am J Public Health*. 1997;87:548–550.
14. Fairclough DL, Peterson HF, Chang V. Why are missing quality of life data a problem in clinical trials of cancer therapy? *Stat Med*. 1998;17:667–677.
15. Hair Jr JE, Anderson RE, Tatham RL, Black WC. *Multivariate Data Analysis*. Upper Saddle River, NJ: Prentice Hall; 1998.
16. Curran D, Molenberghs G, Fayers PM, Machin D. Incomplete quality of life data in randomized trials: missing forms. *Stat Med*. 1998;17:697–709.

17. Bowden RJ, Turkington DA. *Instrumental Variables*. Cambridge, England: Cambridge University Press; 1984.
18. Angrist JD, Imbens GW, Rubin DB. Identification of causal effects using instrumental variables. *J Am Stat Assoc*. 1996;91:444–455.
19. Sturmer T, Schneeweiss S, Avorn J, Glynn RJ. Correcting effect estimates for unmeasured confounding in cohort studies with validation studies using propensity score calibration. *Am J Epidemiol*. 2005;162:279–289.
20. Rosner B, Spiegelman D, Willett WC. Correction of logistic regression relative risk estimates and confidence intervals for measurement error: the case of multiple covariates measured with error. *Am J Epidemiol*. 1990;132:734–745.
21. Heckman J. Sample selection bias as a specification error. *Econometrica*. 1979;47:153–161.
22. Whitehead JC. Item nonresponse in contingent valuation: should CV researchers impute values for missing independent variables? *J Leisure Res*. 1994;26:296–298.
23. Downey RG, King CV. Missing data in Likert ratings: a comparison of replacement methods. *J Gen Psychol*. 1998;125:175–191.
24. Bono CB. Missing Data on the Center for Epidemiologic Studies Depression (CES-D) Scale: A Comparison of Four Imputation Techniques. Dissertation Abstracts: Ann Arbor, MI; May 2003.
25. Ford BL. An overview of hot-deck procedures. In: Madlow WG, Olkim I, Rubin DB, eds, *Incomplete Data in Sample Surveys, Theory and Bibliographies*, vol. 2. San Diego, CA: Academic Press; 1983, p. 185–207.
26. Rubin DB. Software for multiple imputation. Available at: http://www.statsol.ie/solas/rubin1.htm. Accessed 30.03.06.
27. Pepine CJ, Handberg-Thurmond E, Marks RG, et al, for the INVEST Investigators. Rationale and design of the International Verapamil SR/Trandolapril Study (INVEST): an Internet-based randomized trial in coronary artery disease patients with hypertension. *J Am Coll Cardiol*. 1998;32:1228–1237.
28. Pepine CJ, Handberg EM, Cooper-DeHoff RM, et al. A calcium antagonist vs a non-calcium antagonist hypertension treatment strategy for patients with coronary artery disease. The International Verapamil-Trandolapril Study (INVEST): a randomized controlled trial. *JAMA*. 2003;290:2805–2816.
29. Joint National Committee on Prevention, Detection, Evaluation, and Treatment of High Blood Pressure. The sixth report of the Joint National Committee on prevention, detection, evaluation, and treatment of high blood pressure. *Arch Intern Med*. 1997;157:2413–2446.
30. Turner RJ, Noh S. Physical disability and depression: a longitudinal analysis. *J Health Soc Behav*. 1988;29:23–37.
31. Weissman MM, Sholomskas D, Pottenger M, Prusoff BA, Locke BZ. Assessing depressive symptoms in five psychiatric populations: a validation study. *Am J Epidemiol*. 1977;106:203–214.
32. Husaini BA, Neff JA, Harrington JB, Hughes MD, Stone RH. Depression in rural communities: validating the CES-D Scale. *J Community Psychol*. 1980;8:20–27.
33. Roberts RE, Vernon SW. The Center for Epidemiologic Studies Depression Scale: its use in a community sample. *Am J Psychiatry*. 1983;140:41–46.
34. Kaelber CT, Moul DE, Farmer MW. Epidemiology of depression. In: Beckham EE, Leber WR, eds. *Handbook of Depression*. 2nd ed. New York: Guilford Press; 1995, pp. 3–35.
35. Frerichs RE, Aneshensel CS, Clark VA. Prevalence of depression in Los Angeles county. *Am J Epidemiol*. 1981;113:691–699.
36. Sonnenberg CM, Beekman AT, Deeg DJ, van Tilburg W. Sex differences in late-life depression. *Acta Psychiatr Scand*. 2000;101:286–292.
37. Lonnqvist JK, Henriksson MM, Isometsa ET, et al. Mental disorders and suicide prevention. *Psychiatry Clin Neurosci*. 1995;49:S111–S116.

38. Carpiniello B, Carta MG, Rudas N. Depression among elderly people. A psychosocial study of urban and rural populations. *Acta Psychiatr Scand.* 1989;80:445–450.
39. George LK. Social factors and the onset and outcomes of depression. In: Schaie K, Blazer D, House J, eds. *Aging, Health Behaviors, and Health Outcomes*. Hillsdale, NJ: Erlbaum; 1992, pp. 137–78.
40. Kennedy GJ, Kelman HR, Thomas C. The emergence of depressive symptoms in late life: the importance of declining health and increasing disability. *J Community Health*. 1990;15:93–104.
41. Knight RG, Williams S, McGee R, Olaman S. Psychometric properties of the Center for Epidemiologic Studies Depression Scale (CES-D) in a sample of women in middle life. *Behav Res Ther*. 1997;35:373–380.
42. Wassertheil-Smoller S, Applegate WB, Berge K, et al. Change in depression as a precursor of cardiovascular events. *Arch Intern Med.* 1996;156:553–561.
43. Keller MB, Beardslee WR, Dorer DJ, Lavori PW, Samuelson H, Klerman GR. Impact of severity and chronicity of parental affective illness on adaptive functioning and psychopathology in children. *Arch Gen Psychiatry*. 1986;43:930–937.
44. Wells KB, Burnam MA, Rogers W, Hays R, Camp P. The course of depression in adult outpatients. Results from the Medical Outcomes Study. *Arch Gen Psychiatry*. 1992; 49:788–794.
45. Tsuang MT, Faraone SV. The Genetics of Mood Disorders. Baltimore, MD: Johns Hopkins University Press, 1990.
46. Winokur G. Unipolar depression: is it divisible into autonomous subtypes? *Arch Gen Psychiatry*. 1979;36:47–52.
47. Winokur G, Morrison J. The Iowa 500: follow-up of 225 depressives. *Br J Psychiatry*. 1973;123:543–548.
48. Regier DA, Narrow WE, Rae DS, et al. The de Facto US Mental and Addictive Disorders Service System: epidemiologic catchment area prospective 1-year prevalence rates of disorders and services. *Arch Gen Psychiatry*. 1993;50:85–94.
49. Weissman MM, Bruce M, Leaf P, Florio L, Holzer C. Affective disorders. In: Robins L, Regier E, eds. *Psychiatric Disorders in America*. New York: Free Press; 1991, pp. 53–80.
50. Blazer D, Burchett V, Service C, George LK. The association of age and depression among the elderly: an epidemiologic exploration. *J Gerontol.* 1991;46:M210–M215.
51. Ross CE, Mirowsky J. Age and depression. *J Health Soc Behav*. 1992;33:187–205.
52. Garson GD. Data imputation for missing values. Available at: http://www2.chass.ncsu.edu/garson/pa765/missing.htm. Accessed 30.03.06.
53. SAS. Statistical Analysis Systems Institute Package 2001. *SAS/STAT User's Guide, Version 8.12*. Cary, NC: Statistical Analysis Systems Institute, Inc; 2001.
54. Hawthorne G, Elliott P. Imputing cross-sectional missing data: comparison of common techniques. *Aust N Z J Psychiatry*. 2005;39:583–590.
55. Roth P. Missing data: a conceptual review for applied psychologists. *Personnel Psych*. 1994;47:537–560.
56. Little RJA. Pattern-mixture models for multivariate incomplete data. *J Am Stat Assoc*. 1993;88:125−134.
57. Graham JW, Hofer SM, MacKinnon DP. Maximizing the usefulness of data obtained with planned missing value patterns: an application of maximum likelihood procedures. *Multivariate Behav Res*. 1996;31:197–218.
58. Elliott P, Hawthorne G. Imputing missing repeated measures data: how should we proceed? *Aust N Z J Psychiatry*. 2005;39:575–582.