Comparative methods for handling missing data in large databases

Antonia J. Henry, MD, MPH,[a,b] Nathanael D. Hevelone, MPH,[b] Stuart Lipsitz, ScD,[b] and Louis L. Nguyen, MD, MBA, MPH,[a,b] *Boston, Mass*

Objective: **Analysis of complex survey databases is an important tool for health services researchers. Missing data elements are challenging because the reasons for "missingness" are multifactorial, especially categorical variables such as race. We simulated missing data for race and analyzed the bias from five methods used in predicting major amputation in patients with critical limb ischemia (CLI).**

Methods: **Patient discharges with fully observed data containing lower extremity revascularization or major amputation and CLI were selected from the 2003 to 2007 Nationwide Inpatient Sample, a complex survey database (weighted n = 684,057). Considering several random missing data schemes, we compared five missing data methods: complete case analysis, replacement with observed frequencies, missing indicator variable, multiple imputation, and reweighted estimating equations. We created 100 simulated data sets, with 5%, 15%, or 30% of subjects' race drawn to be missing from the full data set. Bias was estimated by comparing the estimated regression coefficients averaged over 100 simulated data sets (β_{miss}) from each method vs estimates from the fully observed data set (β_{full}), with relative bias calculated as ($\beta_{full} - \beta_{miss}/\beta_{full}$) × 100%.**

Results: **Our results demonstrate that reweighted estimating equations produce the least biased and the missing indicator variable produces the most biased coefficients. Complete case analysis, replacement with observed frequencies, and multiple imputation resulted in moderate bias. Sensitivity analysis demonstrated the optimal method choice depends on the quantity and type of missing data encountered.**

Conclusions: **Missing data are an important analytic topic in research with large databases. The commonly used missing indicator variable method introduces severe bias and should be used with caution. We present empiric evidence to guide method selection for handling missing data. (J Vasc Surg 2013;58:1353-9.)**

Analysis of administrative databases is an important tool for health services researchers in vascular surgery. Large numbers and complex survey sampling methodology offer opportunities to address clinical questions with nationally representative data. Although diagnosis and procedure codes are audited and algorithms enrich clinical detail, missing data must be addressed by end-users. Missing demographic data are problematic because reasons for "missingness" are multifactorial. Critically ill patients and certain demographic groups may be less likely to report this information at hospital registration.[1] Analytic methods for handling missing data, especially categorical variables, may introduce bias if the methods do not account for complex survey sampling design. We simulated missing data for race and analyzed the bias from five missing data methods to predict major amputation in patients with critical limb ischemia (CLI). These methods may not be necessary for all research with large databases, but several post hoc methods for handling missing data will be illustrated here.

Missing data mechanisms are defined as missing completely at random (MCAR), missing at random (MAR), and not missing at random (NMAR). Under the MCAR mechanism, the probability of missingness is unrelated to the unknown value of the variable or to other variables in the dataset.[2-4] MAR assumes that the probability that data are missing is not related to the unknown value of the variable but is related to other variables. If the probability that data are missing is related to the unknown value of the variable, then the data are NMAR. The missing data mechanism must be modeled to obtain valid parameter estimates, and this requires detailed a priori knowledge of the missing data mechanism that is not usually available to end-users.[2] To focus on biases that can occur even under MCAR and MAR, we will not simulate data that are NMAR.

There are four common methods and one novel method for handling missing data in large databases. These include complete case analysis (CCA), replacement with observed frequencies (RF), the missing indicator variable method (MIV), multiple imputation (MI), and reweighted estimating equations (RWEE). CCA deletes records with any missing values and is the default in most software

From the Division of Vascular & Endovascular Surgery[a] and the Center for Surgery and Public Health,[b] Brigham & Women's Hospital, Harvard Medical School.

This work was supported by a National Institutes of Health K23 Research Career Development Award (HL-084386) to L.L.N., a National Research Service Award T32 Institutional Research Training Grant (HL-007734) to A.J.H, and a National Cancer Institutes Award (NCI CA-60679) to S.L.

Author conflict of interest: none.

Additional material for this article may be found online at www.jvascsurg.org.

Reprint requests: Louis L. Nguyen, MD, MBA, MPH, Brigham & Women's Hospital, Division of Vascular & Endovascular Surgery, 75 Francis St, Boston, MA 02115 (e-mail: llnguyen@partners.org).

The editors and reviewers of this article have no relevant financial relationships to disclose per the JVS policy that requires reviewers to decline review of any manuscript for which they may have a conflict of interest.

0741-5214/$36.00

http://dx.doi.org/10.1016/j.jvs.2013.05.008

packages.[2,3,5] RF replaces missing values with the sample frequencies or means for the variable of interest calculated from complete cases. In MIV, records with missing data on multilevel categorical variables are designated with indicator variables.

MI was designed to address missing data in complex survey data sets where the database constructor and end-users are distinct entities and there is no singular defined missing data mechanism.[6] The missing values are imputed based on a model relating the missing variable to observed variables, generating multiple completed data sets. The multiple parameter estimates and standard errors are analyzed separately and combined to produce a single parameter estimate and standard error representing the uncertainty of the imputation process.[4-6]

RWEE is a novel method developed specifically for complex survey data.[7-9] The survey "weights" are the inverse of the probability that a person was drawn from the population to be sampled. RWEE adjusts the survey weights of persons with completely observed data. The original survey weights are multiplied by the inverse of the probability that a person was drawn from the survey to be a "complete case." A logistic regression model estimates the probability that the variable of interest is observed and includes variables that are completely observed and related to the variable of interest. The original survey weights are multiplied by the inverse of the estimated probability that the variable is observed. Records with missing data are deleted, and the usual complex survey analysis incorporating stratification, clustering, and weighting is performed with the new adjusted weights.

Health services researchers may find that these methods are applicable to work with large databases. Correlating the missing data mechanism to the assumptions required by each method will guide selection of an appropriate method. Using simulated data sets, we aim to provide practical examples of the advantages and disadvantages of five methods of handling missing data.

METHODS

Creation of the fully observed data set. The 2003 to 2007 Nationwide Inpatient Samples (NISs) were queried for adult discharges containing a diagnosis of chronic CLI and lower extremity revascularization (LER) or major amputation.[10] A weighted sum of 958,120 discharge records was included in this initial cohort. Subject inclusion and exclusion criteria have been reported elsewhere.[9] Briefly, patient discharges were selected if their record contained International Classification of Diseases-9th Edition-Clinical Modification codes for chronic CLI and a procedure to treat CLI, including major lower extremity amputation or LER (lower extremity bypass or angioplasty). The primary outcome was major amputation vs LER. Records with missing data were excluded. The fully observed data set included a weighted sum of 684,057 records.

As in our prior work, bivariate and multivariate logistic regression were performed in the fully observed data set to examine relationships between socioeconomic status, comorbidities, hospital-level factors, and the outcome of major amputation.[9] To use a more parsimonious multivariate model, we excluded several comorbidity variables from our original work that were weakly associated with the outcome and not significant confounders of the primary predictors, which were pulmonary circulatory disorders, chronic pulmonary disease, uncomplicated diabetes, hypothyroidism, liver disease, metastatic cancer, coagulopathy, chronic blood loss anemia, drug abuse, psychoses, and depression.

Missing data mechanism simulation. Race was simulated to be missing for 5%, 15%, and 30% of the records in the fully observed data set according to the MCAR and MAR mechanisms. To estimate the potential bias introduced by each method, 100 simulated data sets were created for each scenario. To simulate MCAR, a random number generator was used to delete the value for race in 5%, 15%, and 30% of records.

To simulate missingness under MAR, multivariate weighted logistic regression in the initial data set identified predictors of missing race. These predictors included metropolitan residential area type, hospital bed size, hospital region, and several Agency for Healthcare Research and Quality comorbidities (deficiency anemia, complicated diabetes mellitus, obesity, and cardiac valvular disease). Although the primary outcome of major amputation was not significantly associated with missing values for race, this variable was empirically included in the model to satisfy the condition that missing data on the variable of interest be associated with the outcome. Parameter estimates from this regression model determined the probability for deleting race in the simulations. Predictors of missing race were included in the multivariate logistic regression model for the outcome of major amputation to produce unbiased parameter estimates. To specify the percentage of records with race deleted, the intercept for the regression model was adjusted.

Analytic methods. Five missing data methods were used to find predictors of major amputation in each simulation scenario using the same weighted multivariate logistic regression model for the outcome of major amputation. The estimated regression coefficients from the fully observed data set were set as reference values. Estimated regression coefficients from the various missing data methods were compared with these reference values to estimate the bias of each missing data method.

CCA excluded all records with missing values on race in the simulated data sets. For RF, six continuous variables were created for race. For complete cases, the variable for a race category was set to 1 based on the NIS value for race and 0 in the other categories. The race variables for all records where race had been deleted were filled in with the group frequencies from the fully observed data set, which were: white, 0.701; black, 0.165; Hispanic, 0.089; Asian/Pacific Islander, 0.012; Native American, 0.006; and other race, 0.019. Records with missing data on race were treated as having a probability of representing patients from each race category. In MIV, six binary variables for each level of race were created in addition to a seventh variable for records with missing data on race.

For MI, five replacement data sets were created for each simulation, and the missing values for race were imputed using the multinomial logistic regression imputation model for monotone missing data in SAS 9.2 software (SAS Institute, Cary, NC), which fit a different probability of being in each race group.[11,12] SAS PROC MI does not incorporate survey weights into the imputation procedure. Each data set was analyzed separately. The parameter estimates and standard errors were combined in the SAS MIANALYZE procedure to output parameter estimates for each level of race with white as the reference group.[11,12]

In RWEE, a weighted logistic regression model was run in the simulated data sets to estimate parameters for the outcome that race was observed given three variables that were observed for all records in the fully observed data set: median income quartile, hospital region, and discharge year. Records with race observed were included in the analysis using the RWEE-adjusted survey weights.

Statistical analysis. The estimated regression coefficients from the multivariate model were averaged across 100 simulated data sets under the six scenarios of MCAR, 5%, 15%, and 30% missing; and MAR, 5%, 15%, and 30% missing. To determine the amount of bias in the regression coefficients from each missing data method, we compared the average of the estimated regression coefficients (β_{miss}) to the reference values from the fully observed data set (β_{full}). The percentage difference between the mean of the estimated coefficients from the simulated data sets and the coefficients from the fully observed data set was defined as the estimated relative bias and calculated as ($\beta_{full} - \beta_{miss}/\beta_{full}$) $\times$ 100%. The magnitude of the relative bias was graded for clearer interpretation based on expert consultation with our group's statistician. The grades are defined as negligible (0%-5%), minimal (5%-10%), moderate (10%-20%), heavy (20%-30%), and severe (>30%). Student's *t*-test was used to determine if the means of the estimated regression coefficients were significantly different from the reference coefficients estimated from the fully observed data set. To account for multiple testing, we set the criterion for significance at $\alpha < .001$. All database linkages and analyses were performed with SAS 9.2 software.

RESULTS

CCA produced minimally biased regression coefficients in the MCAR scenarios. Data from the 5% and 15% missing scenarios are presented in the Supplementary Tables I-VI (online only) due to similar results between MCAR and MAR. CCA performed well in the MCAR scenarios where few mean regression coefficients were minimally to moderately biased (Table I; Supplementary Tables I-III, online only). In the MAR scenarios, CCA performed less favorably because many of the estimated coefficients had moderate to severe bias (Table II; Supplementary Tables IV-VI, online only). With increasing missingness, the estimated coefficients were more likely to have heavy to severe relative bias.

RF resulted in many estimated coefficients that were moderately to severely biased. In the MCAR-5% missing scenario, three estimated coefficients were heavily to severely biased, and this number increased in the 15% and 30% missing scenarios. In the MAR scenarios, heavy to severe bias was found frequently. RF performed less favorably than CCA across all scenarios.

MIV introduced the most bias compared with the other four methods. Most of the estimated coefficients had heavy to severe relative bias across all scenarios. In the MCAR-5% missing scenario, >75% of the estimated coefficients were severely biased. This method continued to perform poorly as the percentage of missing data increased and under MAR.

The results of the MI method were more biased than those from CCA or RWEE. In the MCAR-5% missing scenario, six mean regression coefficients carried moderate to heavy relative bias. The relative bias increased along with the percentage missing and under MAR. In the MAR-30% missing scenario, nine estimated coefficients carried heavy to severe relative bias, and >90% of the estimated coefficients were significantly different from the reference coefficients.

RWEE produced the least biased parameter estimates across all scenarios. In the MCAR scenarios, this method performed similarly to CCA. In the MAR scenarios, RWEE surpassed CCA in producing the least biased results. Under the most challenging scenario, MAR-30% missing, only three estimated regression coefficients were moderately to severely biased.

Income quartile and hospital region were highly sensitive to the amount of missing data, the missing data mechanism, and the analytic method. The relative bias for the income variable levels increased twofold when the percentage of missing data increased from 15% to 30% in the MCAR and MAR scenarios when RF or MI was used. RWEE included this variable in the reweighting process, potentially contributing to minimal relative bias. In the MAR scenarios, the probability of missing race was based on several predictors, including region. In the 30% missing data scenarios, the relative bias for region introduced by CCA increased from 0.05% to 1.6% under MCAR to 25.7% to 1103.3% under MAR. Again, the inclusion of hospital region in the RWEE model resulted in a more modest increase in the range of relative bias, 3% to 37%.

Table III summarizes the results across the different simulation scenarios by converting the results into a linear score and grading system. The relative bias categories were assigned a 5-point scale, from 5 for negligible to 1 for severe. The percentages of variables in each relative bias category were multiplied by the point value. The sum of these scores for each simulation scenario ranged from 100 to 500. A score of 100 to 149 was designated as

Table I. Missing completely at random: 30% missing

	Relative bias[a]				
Variable	*CCA*	*RF*	*MIV*	*MI*	*RWEE*
Age, years					
Q2: 62-70	0.29	1.36[b]	1.37[b]	2.24[b]	0.29
Q3: 71-78	0.50	6.42[b]	6.41[b]	6.70[b]	0.49
Q4: >78	0.04	18.75[b]	18.67[b]	14.38[b]	0.05
Race					
Black	0.04	6.55[b]	6.07[b]	16.24[b]	0.05
Hispanic	0.34	10.29[b]	9.62[b]	6.40[b]	0.34
Asian/PI	1.46	4.49	3.89	41.66[b]	1.46
Native American	0.28	5.30[b]	4.99[b]	1.82	0.29
Other	1.47	11.59[b]	11.10[b]	104.83[b]	1.48
Female sex	0.45	8.61[b]	215.28[b]	7.66[b]	0.44
Income					
Q1	1.48	48.33[b]	191.55[b]	24.73[b]	1.49
Q2	1.43	28.66[b]	118.42[b]	14.60[b]	1.43
Q3	5.12	54.11[b]	403.97[b]	25.92[b]	5.17
Insurance					
Private	0.65	4.40[b]	116.11[b]	3.21[b]	0.66
Medicaid	1.57	8.21[b]	252.38[b]	4.26[b]	1.54
Uninsured	10.06	4.23[b]	610.09[b]	47.93[b]	10.24
Small metropolitan	0.57	24.45[b]	80.68[b]	14.12[b]	0.56
Micropolitan	4.46	133.55[b]	158.65[b]	62.27[b]	4.41
Nonmetropolitan	9.80	630.19[b]	231.92[b]	328.23[b]	9.82
Congestive heart failure	0.02	0.58[b]	105.80[b]	0.16[b]	0.03
Cardiac valve disease	0.18	4.27[b]	286.19[b]	2.01[b]	0.18
Complicated diabetes	0.23	0.82[b]	65.68[b]	0.69[b]	0.23
Hypertension	0.15	5.66[b]	383.93[b]	3.35[b]	0.16
Electrolyte disorders	0.35	0.14[b]	152.16[b]	0.06	0.35
Neurologic disorder	0.21	1.71[b]	32.04[b]	1.57[b]	0.20
Paralysis	0.03	1.47[b]	21.46[b]	1.06[b]	0.03
Vascular disease	0.05	1.07[b]	21.48[b]	1.02[b]	0.04
Renal failure	0.93	14.85[b]	533.67[b]	6.23[b]	0.93
Weight loss	0.00	0.77[b]	77.46[b]	0.49[b]	0.01
Obesity	0.43	5.50[b]	474.78[b]	2.16[b]	0.44
Deficiency anemia	0.35	2.68[b]	154.79[b]	1.64[b]	0.36
Diagnostic angiogram	0.05	0.19[b]	124.25[b]	0.06[b]	0.05
Elective	0.09	1.18[b]	373.33[b]	0.84[b]	0.09
LER volume/y					
Q1: 0-11	0.25	0.51[b]	71.72[b]	0.04	0.25
Q2: 12-71	0.08	0.94[b]	169.11[b]	0.35[b]	0.08
Q3: 72-248	0.25	2.05[b]	90.09[b]	0.99[b]	0.25
Hospital size					
Small	0.88	0.65[b]	282.13[b]	0.54[b]	0.89
Medium	1.42	0.55[b]	154.33[b]	1.82[b]	1.44
Midwest	0.05	111.21[b]	1553.02[b]	101.16[b]	0.54
South	0.62	10.92[b]	108.66[b]	9.04[b]	0.62
West	1.59	8.09[b]	289.80[b]	66.48[b]	1.58

CCA, Complete case analysis; *LER,* lower extremity revascularization; *MI,* multiple imputation; *MIV,* missing indicator variable; *PI,* Pacific Islander; *Q,* quartile; *RF,* replacement with observed frequencies; *RWEE,* reweighted estimating equations.
[a]Relative bias: 0%-5% = negligible, 5%-10% = minimal, 10%-20% = moderate, 20%-30% = heavy, >30% = severe.
[b]$P < .001$.

one star, 150 to 249 as two stars, 250 to 349 as three stars, 350 to 449 as four stars, and 450 to 500 as five stars.

DISCUSSION

Most databases have some missing data, and categorical variables with missing data add another layer of difficulty. Many researchers simplify the handling of missing data and risk, introducing bias. Our evaluation of four common and one novel method for handling missing data may serve as a useful guide for other researchers. The results of the simulations demonstrated that MCAR, CCA, and RWEE perform well, introducing the least amount of bias, followed by RF and MI, and lastly, MIV. When applied to data that are MAR, RWEE resulted in less biased parameter estimates than other methods. MIV produced the most severely biased results across all scenarios. The bias we observed when we used MIV, RF, and MI was found in other predictor variables

Table II. Missing at random: 30% missing

	Relative bias[a]				
Variable	*CCA*	*RF*	*MIV*	*MI*	*RWEE*
Age, years					
Q2: 62-70	2.12[b]	1.52[b]	1.41[b]	2.11[b]	1.28
Q3: 71-78	5.07[b]	7.51[b]	7.21[b]	7.16[b]	0.77
Q4: >78	9.24[b]	20.81[b]	19.59[b]	14.55[b]	4.31
Race					
Black	1.56[b]	14.87[b]	9.33[b]	19.88[b]	0.14
Hispanic	0.99	18.86[b]	13.82[b]	2.16[b]	0.15
Asian/PI	0.72	6.18[b]	7.31[b]	40.59[b]	1.70
Native American	2.99	13.76[b]	8.99[b]	5.01[b]	0.53
Other	11.56[b]	24.81[b]	22.73[b]	92.72[b]	3.12
Female sex	0.47	9.72[b]	468.24[b]	7.89[b]	0.41
Income					
Q1	4.89[b]	54.06[b]	191.06[b]	29.69[b]	0.11
Q2	2.02	26.30[b]	122.96[b]	15.53[b]	0.43
Q3	19.74[b]	41.92[b]	387.23[b]	25.27[b]	6.72
Insurance					
Private	1.95[b]	5.06[b]	114.37[b]	3.59[b]	0.14
Medicaid	7.98[b]	9.03[b]	253.09[b]	4.08[b]	0.07
Uninsured	14.98	1.91	610.37[b]	52.01[b]	5.47
Small metropolitan	46.33[b]	25.84[b]	80.38[b]	16.60[b]	0.39
Micropolitan	136.88[b]	151.58[b]	1.76	71.04[b]	2.88
Nonmetropolitan	378.96[b]	752.32[b]	915.44[b]	422.77[b]	44.32
Congestive heart failure	2.63[b]	0.65[b]	112.40[b]	0.20[b]	0.01
Cardiac valve disease	2.51	5.01[b]	286.19[b]	2.66[b]	0.35
Complicated diabetes	2.25[b]	0.70[b]	62.63[b]	0.76[b]	0.26
Hypertension	3.26[b]	6.80[b]	388.54[b]	4.08[b]	0.65
Electrolyte disorders	0.11	0.17[b]	151.81[b]	0.03	0.00
Neurologic disorder	2.55[b]	2.15[b]	32.04[b]	1.94[b]	0.44
Paralysis	2.49[b]	1.65[b]	21.26[b]	1.10[b]	0.38
Vascular disease	4.00[b]	1.28[b]	221.46[b]	1.08[b]	0.02
Renal failure	11.64[b]	17.59[b]	534.45[b]	7.90[b]	0.79
Weight loss	2.19[b]	1.13[b]	77.02[b]	0.73[b]	0.06
Obesity	14.75[b]	6.71[b]	475.23[b]	2.59[b]	1.41
Deficiency anemia	2.55[b]	2.86[b]	148.43[b]	1.96[b]	0.15
Diagnostic angiogram	0.89[b]	0.16[b]	125.15[b]	0.01	0.13
Elective	0.31	1.19[b]	372.74[b]	0.81[b]	0.03
LER volume/y					
Q1: 0-11	2.43[b]	0.50[b]	71.77[b]	0.01	0.14
Q2: 12-71	1.41[b]	1.08[b]	169.39[b]	0.14[b]	0.42
Q3: 72-248	0.84[b]	2.05[b]	90.03[b]	1.12[b]	0.52
Small hospital	20.35[b]	1.63[b]	281.88[b]	0.79[b]	0.15
Medium hospital	52.97[b]	4.94[b]	104.88[b]	1.08[b]	0.34
Midwest	1103.13[b]	165.59[b]	676.05[b]	118.39[b]	17.50
South	25.81[b]	15.78[b]	164.26[b]	12.77[b]	3.02[b]
West	25.70[b]	6.54[b]	258.14[b]	54.16[b]	37.14[b]

CCA, Complete case analysis; *LER,* lower extremity revascularization; *MI,* multiple imputation; *MIV,* missing indicator variable; *PI,* Pacific Islander; *Q,* quartile; *RF,* replacement with observed frequencies; *RWEE,* reweighted estimating equations.
[a]Relative bias: 0%-5% = negligible, 5%-10% = minimal, 10%-20% = moderate, 20%-30% = heavy, >30% = severe.
[b]$P < .001$.

in the model, including potential confounders of race and socioeconomic status.

Understanding the data source and database structure is a first step. We drew a cohort from the NIS, which represents a 20% stratified probability sample of United States community hospitals. The reasons for data to be missing in the NIS depend on the type of variable, data collection, and recoding of sensitive demographic and disease-related information to prevent subject identification. The 2007 NIS contains data from 40 participating states, 10 of which do not report data for race.[10] Of the 25% of discharges with missing data on race, 20% are from states that do not report this information. In these cases, missingness is clustered by state and unrelated to the intrinsic characteristics of the patients. The other 5% are missing for unknown reasons. Given the differences in state reporting patterns in the NIS, the missing data mechanism for race was assumed to meet criteria for MAR rather than NMAR. In addition, a logistic regression model for the outcome of missingness can be used to determine if the data are MCAR vs MAR.

Table III. Summary of results of linear scoring for relative bias from each method

Variable	*CCA*	*RF*	*MIV*	*MI*	*RWEE*
MCAR missing					
5%	5★	5★	2★	5★	5★
15%	5★	4★	2★	4★	5★
30%	5★	4★	2★	4★	5★
MAR missing					
5%	5★	5★	2★	5★	5★
15%	4★	4★	2★	4★	5★
30%	4★	4★	2★	4★	5★

CCA, Complete case analysis; *MAR,* missing at random; *MCAR,* missing completely at random; *MI,* multiple imputation; *MIV,* missing indicator variable; *RF,* replacement with observed frequencies; *RWEE,* reweighted estimating equations.

If the coefficients for the predictors are not significantly associated with the outcome of missingness, the data may be MCAR. If not, the data may be MAR, and RWEE may be applicable.

If the missing data mechanism can be determined, researchers must choose a post hoc method that incorporates the missing data into the analysis. We tested four common and one novel method in six simulation scenarios that may be encountered by end-users of large databases. Under MCAR, CCA produced regression coefficients with minimal relative bias that were not significantly different from reference coefficients. These findings confirmed that CCA will produce unbiased inferences if the data are MCAR. In the MAR scenarios, CCA introduced less bias than MI. One important disadvantage of CCA is the reduction in statistical power caused by excluding individuals with missing values for some variables.[3,4]

RF and MIV are attractive options for many researchers because both allow for conservation of sample size. However, there are several important disadvantages. First, neither method incorporates the survey sampling design. Second, replacing the missing values with the mean frequencies or assigning an indicator variable for missing data assumes homogeneity of the records with missing values and ignores associations between variables, introducing greater bias. Furthermore, RF fails to account for the uncertainty of the replaced value and underestimates the standard errors.[6] In our analysis, most of the estimated regression coefficients in MIV were significantly different from the reference values. MIV introduced the most severe bias into the results in all simulation scenarios and should be used with great caution.

MI introduced more bias into the estimated coefficients than expected. One potential explanation is that the MI procedure we used does not incorporate survey weights into the imputation procedure, leading to bias. In another simulation of imputation of categorical variables with PROC MI, Allison[5] found that CCA and logistic imputation performed equally well and introduced similar degrees of bias into the analysis. In the simulations presented there, sensitivity analyses were not performed with increasing amounts of missingness, as we did. In a study of patients receiving cancer care, CCA, MIV, and MI were compared in identifying predictors of having a discussion about hospice care. Most variables had some missing data and were MAR or NMAR. The author determined that CCA deleted records in a nonrandom fashion, leading to biased parameter estimates and standard errors and, ultimately, to misidentification of significant predictors of the outcome. That analysis found the results of MIV were similar to those from MI.[4]

RWEE provided the least biased results across all scenarios and also incorporated the complex survey design into the analysis step where missing data are handled. In the MAR scenarios, RWEE introduced the least bias into the estimated regression coefficients. RWEE remained robust to increasing amounts of missingness. Although this method is a flexible option for analysts, it is easiest to implement when only one variable has missing values. A solution has been suggested for applying weighted estimating equations to data sets with more complicated patterns of missing data.[13,14] If multiple variables have missing data, but there are enough fully observed variables in the data set that are associated with the probability of having missing data to develop a good predictive model for being a complete case, then RWEE will give unbiased estimates.

This work has several important limitations. First, our reference data set represents a complete case analysis of the original data set where we assumed most of the records with missing race were MAR due to state reporting practices. We acknowledge that end-users of large databases are unable to determine if the missing data mechanism is MAR or NMAR. This assumption potentially carried bias into the simulated data sets.

Second, in our MAR simulations, we deleted race based on associations that were present in the initial data set. The estimated coefficients for the associations between predictors and the outcome of amputation may have been biased using this approach. However, we assert that applying these associations from the initial data set to the simulated data sets may be more realistic and less biased than those drawn from a different database. By using the best available real data, our simulations represent a realistic situation that would be encountered by health services researchers in vascular surgery.

Third, we limited our simulation to one categorical variable with missing values given the challenges presented to researchers. The application of these methods to binary or continuous variables will produce similar results.[8]

CONCLUSIONS

These limitations notwithstanding, the simulation scenarios presented here are unique in that we modeled plausible situations that end-users of large databases encounter and present evidence to support selection of the most appropriate method to handle missing data. This work is novel because we used real data as the basis for empiric simulations and drew our sample from a complex survey database. These five methods for handling missing data can be applied to nonsample survey databases with similar results. RWEE was initially developed for non-survey-weighted data as an extension of nonresponse weighting.[13] In nonsample surveys, each record has a weight of 1.

Missing data is an important analytical topic in health outcomes research because most databases have some missing data and mishandling the data can introduce bias. When the amount of missing data is small and the mechanism is MCAR, CCA and RWEE perform well. If the percentage of records with missing values approaches 30% or if the data are MAR, we recommend RWEE. MI may produce biased results if the multiple imputation procedure does not incorporate survey weights.

These methods may not be necessary for all research with large databases, but if they are, we offer recommendations based on an empiric simulation to assist others in making an informed decision about handling missing data. If the variables of interest are critical to the analysis and associated with the outcome and a substantial proportion of records are incomplete, then one of these post hoc methods may be applicable. Method selection should be guided by the proportion of missing data, the missing data mechanism, and the relationship of the variable to the outcome. RWEE is a novel and flexible option that provides less biased parameter estimates. Because of its simplicity of implementation, MIV may be the most widely used approach; however, MIV introduced the most bias and should be used with caution. These simulation scenarios may serve as a useful guide for other users of complex survey databases.

AUTHOR CONTRIBUTIONS

Conception and design: AH, SL, LN
Analysis and interpretation: AH, NH, SL, LN
Data collection: AH, NH, LN
Writing the article: AH, LN
Critical revision of the article: AH, NH, SL, LN
Final approval of the article: AH, NH, SL, LN
Statistical analysis: AH, NH, SL, LN
Obtained funding: AH, LN
Overall responsibility: LN

REFERENCES

1. McAlpine DD, Beebe TJ, Davern M, Call KT. Agreement between self-reported and administrative race and ethnicity data among Medicaid enrollees in Minnesota. Health Serv Res 2007;42:2373-88.
2. Little RJA. Regression with missing X's: a review. J Am Stat Assoc 1992;87:1227-37.
3. Allison PD. Missing data. Thousand Oaks, CA: Sage Publications; 2002. p. 1-12, 27-88.
4. He Y. Missing data analysis using multiple imputation: getting to the heart of the matter. Circ Cardiovasc Qual Outcomes 2010;3:98-105.
5. Allison PD. Imputation of categorical variables with PROC MI. Presented at: SAS Users Group International, 30th Meeting (SUGI 30). Philadelphia: University of Pennsylvania; 2005. p. 1-14.
6. Rubin DB. Multiple imputation after 18+ years. J Am Stat Assoc 1996;91:473-89.
7. HCUP Databases. Healthcare Cost and Utilization Project (HCUP). 2003-2007. Rockville, MD: Agency for Healthcare Research and Quality. Available at: http://www.hcup-us.ahrq.gov/nisoverview.jsp. Accessed November 15, 2010.
8. Moore CG, Lipsitz SR, Addy CL, Hussey JR, Fitzmaurice G, Natarajan S. Logistic regression with incomplete covariate data in complex survey sampling: application of reweighted estimating equations. Epidemiology 2009;20:382-90.
9. Henry AJ, Hevelone ND, Belkin M, Nguyen LL. Socioeconomic and hospital-related predictors of amputation for critical limb ischemia. J Vasc Surg 2011;53:330-9.e1.
10. HCUP Nationwide Inpatient Sample (NIS). Healthcare Cost and Utilization Project (HCUP). 2003-2007. Rockville, MD: Agency for Healthcare Research and Quality. Available at: http://www.hcup-us.ahrq.gov/nisoverview.jsp. Accessed November 15, 2010.
11. Berglund PA. An introduction to multiple imputation of complex sample data using SAS v9.2. Ann Arbor: University of Michigan, SAS Global Forum; 2010. p. 1-12.
12. Graham JW. Missing data analysis: making it work in the real world. Annu Rev Psychol 2009;60:549-76.
13. Zhao LP, Lipsitz S, Lew D. Regression analysis with missing covariate data using estimating equations. Biometrics 1996;52:1165-82.
14. Robins JMRAJ, Rotnitzky A, Zhao LP. Estimation of regression coefficients when some regressors are not always observed. J Am Stat Assoc 1994;89:846-66.

Submitted Feb 18, 2013; accepted May 3, 2013.

Additional material for this article may be found online at www.jvascsurg.org.

Supplementary Table I (online only). Missing completely at random: 5% missing

	CCA			*RF*			*MIV*			*MI*			*RWEE*		
Variable	*RB*[a]	*SE*	P	*RB*	*SE*	P	*RB*	*SE*	P	*RB*	*SE*	P	*RB*	*SE*	P
Age, years															
Q2: 62-70	0.28	0.16	.080	0.25	0.02	<.0001	0.25	0.02	.0001	0.48	0.02	<.0001	0.28	0.16	.078
Q3: 71-78	0.58	0.20	.004	1.08	0.03	<.0001	1.08	0.03	<.0001	1.30	0.03	<.0001	0.58	0.20	.004
Q4: >78	0.50	0.46	.282	3.23	0.07	<.0001	3.20	0.07	<.0001	2.78	0.07	<.0001	0.50	0.46	.279
Race															
Black	0.10	0.06	.090	1.05	0.06	<.0001	0.97	0.06	<.0001	2.74	0.06	<.0001	0.10	0.06	.095
Hispanic	0.02	0.12	.897	1.87	0.13	<.0001	1.75	0.13	<.0001	1.10	0.14	<.0001	0.01	0.12	.910
Asian/PI	0.74	0.50	.144	0.38	0.50	.443	0.28	0.50	.576	8.56	0.51	<.0001	0.73	0.50	.146
Native American	0.30	0.35	.390	0.56	0.35	.117	0.50	0.35	.156	0.85	0.36	.021	0.30	0.35	.395
Other	0.71	0.74	.338	1.11	0.74	.135	1.02	0.74	.168	19.61	0.77	<.0001	0.71	0.74	.341
Female sex	0.24	0.20	.240	1.48	0.03	<.0001	220.41	1.79	<.0001	1.46	0.03	<.0001	0.23	0.20	.245
Income															
Q1	0.29	0.29	.315	8.44	0.09	<.0001	198.65	0.03	<.0001	4.66	0.09	<.0001	0.29	0.29	.323
Q2	0.99	0.48	.043	5.00	0.07	<.0001	59.89	0.13	<.0001	2.70	0.07	<.0001	0.99	0.48	.044
Q3	0.95	1.66	.568	9.17	0.24	<.0001	311.61	0.28	<.0001	4.46	0.25	<.0001	0.94	1.66	.572
Insurance															
Private	0.25	0.17	.145	0.76	0.03	<.0001	111.42	0.03	<.0001	0.59	0.03	<.0001	0.25	0.17	.144
Medicaid	0.65	0.34	.059	1.47	0.06	<.0001	247.08	0.04	<.0001	0.84	0.06	<.0001	0.65	0.34	.057
Uninsured	0.32	2.79	.908	0.81	0.48	.093	566.24	0.40	<.0001	10.15	0.49	<.0001	0.31	2.79	.911
Small metropolitan	0.25	0.24	.295	4.27	0.06	<.0001	81.30	0.09	<.0001	2.59	0.05	<.0001	0.26	0.24	.292
Micropolitan	1.25	1.19	.299	23.28	0.30	<.0001	227.20	0.20	<.0001	11.46	0.29	<.0001	1.24	1.20	.302
Nonmetropolitan	5.34	5.42	.326	111.46	1.35	<.0001	208.94	1.22	<.0001	62.46	1.27	<.0001	5.23	5.41	.336
Congestive heart failure	0.02	0.04	.699	0.10	0.01	<.0001	100.12	0.01	<.0001	0.02	0.01	.001	0.02	0.04	.675
Cardiac valve disease	0.07	0.28	.810	0.83	0.03	<.0001	288.04	0.02	<.0001	0.42	0.03	<.0001	0.07	0.28	.801
Complicated diabetes	0.04	0.05	.463	0.15	0.01	<.0001	64.46	0.01	<.0001	0.13	0.01	<.0001	0.04	0.05	.465
Hypertension	0.09	0.13	.471	1.00	0.02	<.0001	382.03	0.02	<.0001	0.67	0.02	<.0001	0.09	0.13	.474
Electrolyte disorders	0.04	0.09	.624	0.03	0.01	.030	154.71	0.01	<.0001	0.01	0.01	.696	0.04	0.09	.637
Neurologic disorder	0.09	0.09	.330	0.30	0.01	<.0001	32.11	0.01	<.0001	0.32	0.01	<.0001	0.08	0.09	.340
Paralysis	0.05	0.10	.605	0.29	0.01	<.0001	22.54	0.01	<.0001	0.24	0.01	<.0001	0.06	0.10	.595
Vascular disease	0.07	0.04	.076	0.19	0.01	<.0001	220.07	0.02	<.0001	0.21	0.01	<.0001	0.07	0.04	.074
Renal failure	0.39	0.22	.078	2.62	0.04	<.0001	529.88	0.02	<.0001	1.22	0.04	<.0001	0.39	0.22	.081
Weight loss	0.10	0.08	.251	0.14	0.01	<.0001	79.85	0.01	<.0001	0.09	0.01	<.0001	0.10	0.08	.252
Obesity	0.54	0.34	.113	0.99	0.04	<.0001	472.42	0.05	<.0001	0.41	0.04	<.0001	0.54	0.34	.113
Deficiency anemia	0.00	0.08	.970	0.48	0.01	<.0001	152.46	0.02	<.0001	0.32	0.01	<.0001	0.01	0.08	.941
Diagnostic angiogram	0.02	0.03	.370	0.03	0.00	<.0001	123.73	0.00	<.0001	0.01	0.00	.001	0.02	0.03	.362
Elective	0.06	0.04	.152	0.21	0.01	<.0001	373.77	0.01	<.0001	0.17	0.01	<.0001	0.06	0.04	.153
LER volume/y															
Q1: 0-11	0.04	0.05	.436	0.09	0.01	<.0001	71.99	0.00	<.0001	0.01	0.01	.096	0.04	0.05	.443
Q2: 12-71	0.01	0.06	.834	0.16	0.01	<.0001	167.98	0.02	<.0001	0.08	0.01	<.0001	0.01	0.06	.824
Q3: 72-248	0.04	0.08	.601	0.36	0.01	<.0001	88.62	0.02	<.0001	0.17	0.01	<.0001	0.04	0.08	.620
Small hospital	0.12	0.19	.539	0.16	0.02	<.0001	279.13	0.02	<.0001	0.14	0.03	<.0001	0.12	0.19	.533
Medium hospital	0.34	0.34	.312	0.08	0.05	.113	153.11	0.06	<.0001	0.33	0.05	<.0001	0.34	0.34	.313
Midwest	5.12	6.64	.442	20.69	1.19	<.0001	1560.90	0.80	<.0001	18.75	1.20	<.0001	5.06	6.65	.448
South	0.00	0.22	.986	1.91	0.04	<.0001	104.95	0.05	<.0001	1.65	0.04	<.0001	0.01	0.22	.979
West	0.26	1.08	.811	1.56	0.25	<.0001	258.29	0.16	<.0001	14.06	0.27	<.0001	0.28	1.08	.795

CCA, Complete case analysis; *LER,* lower extremity revascularization; *MI,* multiple imputation; *MIV,* missing indicator variable; *PI,* Pacific Islander; *Q,* quartile; *RB,* relative bias; *SE,* standard error; *RF,* replacement with observed frequencies; *RWEE,* reweighted estimating equations.

[a]RB: 0%-5% = negligible, 5%-10% = minimal, 10%-20% = moderate, 20%-30% = heavy, >30% = severe.

Supplementary Table II (online only). Missing completely at random: 15% missing

	CCA			*RF*			*MIV*			*MI*			*RWEE*		
Variable	*RB*[a]	*SE*	P	*RB*[a]	*SE*	P	*RB*[a]	*SE*	P	*RB*[a]	*SE*	P	*RB*[a]	*SE*	P
Age, years															
Q2: 62-70	0.61	0.31	.051	0.68	0.03	<.0001	0.67	0.03	<.0001	1.25	0.04	<.0001	0.61	0.31	.049
Q3: 71-78	0.43	0.36	.241	3.20	0.05	<.0001	3.18	0.05	<.0001	3.60	0.06	<.0001	0.44	0.36	.224
Q4: >78	0.15	0.83	.856	9.43	0.11	<.0001	9.35	0.11	<.0001	7.73	0.12	<.0001	0.18	0.83	.829
Race															
Black	0.12	0.09	.186	3.26	0.10	<.0001	3.01	0.09	<.0001	8.15	0.08	<.0001	0.12	0.09	.193
Hispanic	0.00	0.24	.996	5.60	0.26	<.0001	5.24	0.26	<.0001	2.87	0.24	<.0001	0.00	0.24	.991
Asian/PI	1.98	0.87	.025	5.49	0.87	<.0001	5.18	0.87	<.0001	18.99	0.74	<.0001	2.00	0.87	.024
Native American	0.16	0.50	.750	2.49	0.51	<.0001	2.32	0.51	<.0001	1.31	0.49	.010	0.16	0.50	.753
Other	0.81	1.22	.508	6.22	1.23	<.0001	5.95	1.23	<.0001	54.78	1.33	<.0001	0.83	1.22	.499
Female sex	0.22	0.30	.468	4.46	0.05	<.0001	218.79	1.06	<.0001	4.24	0.05	<.0001	0.22	0.30	.458
Income															
Q1	0.27	0.49	.579	24.93	0.15	<.0001	195.68	0.05	<.0001	13.34	0.16	<.0001	0.27	0.49	.586
Q2	1.12	0.77	.147	14.81	0.12	<.0001	84.06	0.23	<.0001	7.91	0.14	<.0001	1.12	0.77	.147
Q3	1.24	3.34	.712	27.59	0.38	<.0001	349.96	0.48	<.0001	13.63	0.43	<.0001	1.24	3.34	.710
Insurance															
Private	0.06	0.32	.843	2.14	0.04	<.0001	113.35	0.04	<.0001	1.66	0.04	<.0001	0.07	0.32	.832
Medicaid	0.48	0.62	.436	4.41	0.10	<.0001	249.08	0.06	<.0001	2.15	0.11	<.0001	0.47	0.62	.448
Uninsured	2.32	5.50	.674	3.17	0.66	<.0001	585.39	0.63	<.0001	26.31	0.74	<.0001	2.28	5.50	.680
Small metropolitan	0.30	0.40	.459	12.57	0.08	<.0001	80.85	0.12	<.0001	7.51	0.09	<.0001	0.28	0.40	.482
Micropolitan	1.41	1.97	.475	68.83	0.45	<.0001	199.05	0.29	<.0001	33.14	0.43	<.0001	1.44	1.98	.469
Nonmetropolitan	15.65	9.24	.093	327.93	2.11	<.0001	27.14	1.79	<.0001	180.13	2.15	<.0001	15.48	9.24	.097
Congestive heart failure	0.01	0.08	.928	0.30	0.01	<.0001	102.49	0.02	<.0001	0.08	0.01	<.0001	0.01	0.08	.899
Cardiac valve disease	0.29	0.50	.567	2.39	0.06	<.0001	287.29	0.04	<.0001	1.27	0.07	<.0001	0.28	0.50	.583
Complicated diabetes	0.14	0.09	.132	0.45	0.01	<.0001	65.00	0.02	<.0001	0.38	0.01	<.0001	0.15	0.09	.110
Hypertension	0.11	0.22	.613	2.94	0.03	<.0001	382.86	0.04	<.0001	1.83	0.03	<.0001	0.10	0.22	.646
Electrolyte disorders	0.02	0.16	.908	0.07	0.02	.001	153.65	0.02	<.0001	0.03	0.02	.137	0.03	0.16	.866
Neurologic disorder	0.01	0.19	.959	0.89	0.02	<.0001	32.08	0.01	<.0001	0.90	0.02	<.0001	0.00	0.19	.991
Paralysis	0.20	0.16	.199	0.75	0.02	<.0001	22.09	0.02	<.0001	0.57	0.02	<.0001	0.19	0.16	.216
Vascular disease	0.05	0.07	.469	0.54	0.01	<.0001	220.62	0.03	<.0001	0.56	0.01	<.0001	0.05	0.07	.507
Renal failure	0.01	0.38	.983	7.79	0.06	<.0001	531.38	0.04	<.0001	3.57	0.07	<.0001	0.05	0.38	.886
Weight loss	0.03	0.15	.841	0.40	0.02	<.0001	78.84	0.01	<.0001	0.26	0.02	<.0001	0.03	0.15	.863
Obesity	0.34	0.57	.548	2.93	0.07	<.0001	473.40	0.08	<.0001	1.20	0.08	<.0001	0.33	0.57	.558
Deficiency anemia	0.07	0.15	.633	1.39	0.02	<.0001	153.46	0.04	<.0001	0.89	0.02	<.0001	0.07	0.15	.634
Diagnostic angiogram	0.00	0.05	.970	0.10	0.01	<.0001	123.94	0.00	<.0001	0.02	0.00	<.0001	0.00	0.05	.945
Elective	0.03	0.09	.720	0.62	0.01	<.0001	373.58	0.01	<.0001	0.47	0.01	<.0001	0.03	0.09	.721
LER volume/y															
Q1: 0-11	0.08	0.09	.339	0.27	0.01	<.0001	71.88	0.00	<.0001	0.00	0.01	.935	0.09	0.09	.325
Q2: 12-71	0.01	0.12	.942	0.47	0.02	<.0001	168.47	0.02	<.0001	0.19	0.02	<.0001	0.01	0.12	.941
Q3: 72-248	0.08	0.15	.605	1.04	0.02	<.0001	89.22	0.03	<.0001	0.51	0.02	<.0001	0.08	0.15	.606
Small hospital	0.40	0.36	.267	0.35	0.04	<.0001	280.34	0.04	<.0001	0.37	0.05	<.0001	0.38	0.36	.284
Medium hospital	0.17	0.61	.774	0.39	0.08	<.0001	153.56	0.11	<.0001	1.04	0.08	<.0001	0.16	0.60	.789
Midwest	9.26	14.17	.515	59.21	1.51	<.0001	1555.85	1.38	<.0001	57.53	1.68	<.0001	9.18	14.20	.520
South	0.15	0.44	.729	5.67	0.06	<.0001	106.53	0.06	<.0001	4.94	0.06	<.0001	0.14	0.44	.755
West	0.75	1.96	.702	5.88	0.44	<.0001	271.39	0.21	<.0001	36.36	0.43	<.0001	0.73	1.96	.712

CCA, Complete case analysis; *LER,* lower extremity revascularization; *MI,* multiple imputation; *MIV,* missing indicator variable; *PI,* Pacific Islander; *Q,* quartile; *RB,* relative bias; *SE,* standard error; *RF,* replacement with observed frequencies; *RWEE,* reweighted estimating equations.
[a]RB: 0%-5% = negligible, 5%-10% = minimal, 10%-20% = moderate, 20%-30% = heavy, >30% = severe.

Supplementary Table III (online only). Missing completely at random: 30% missing

	CCA			*RF*			*MIV*			*MI*			*RWEE*		
Variable	*RB*[a]	*SE*	P	*RB*	*SE*	P	*RB*[a]	*SE*	P	*RB*[a]	*SE*	P	*RB*[a]	*SE*	P
Age, years															
Q2: 62-70	0.29	0.46	.531	1.36	0.04	<.0001	1.37	0.04	<.0001	2.24	0.05	<.0001	0.29	0.46	.536
Q3: 71-78	0.50	0.58	.396	6.42	0.06	<.0001	6.41	0.06	<.0001	6.70	0.07	<.0001	0.49	0.58	.398
Q4: >78	0.04	1.01	.970	18.75	0.12	<.0001	18.67	0.12	<.0001	14.38	0.15	<.0001	0.05	1.01	.957
Race															
Black	0.04	0.14	.762	6.55	0.15	<.0001	6.07	0.15	<.0001	16.24	0.13	<.0001	0.05	0.14	.722
Hispanic	0.34	0.37	.353	10.29	0.37	<.0001	9.62	0.37	<.0001	6.40	0.36	<.0001	0.34	0.37	.358
Asian/PI	1.46	1.42	.307	4.49	1.44	.002	3.89	1.44	.008	41.66	1.12	<.0001	1.46	1.42	.306
Native American	0.28	0.98	.776	5.30	0.97	<.0001	4.99	0.97	<.0001	1.82	0.70	.011	0.29	0.98	.768
Other	1.47	2.12	.491	11.59	2.16	<.0001	11.10	2.17	<.0001	104.83	1.69	<.0001	1.48	2.12	.486
Female sex	0.45	0.58	.437	8.61	0.07	<.0001	215.28	0.86	<.0001	7.66	0.07	<.0001	0.44	0.57	.449
Income															
Q1	1.48	0.79	.064	48.33	0.18	<.0001	191.55	0.07	<.0001	24.73	0.23	<.0001	1.49	0.79	.062
Q2	1.43	1.21	.238	28.66	0.15	<.0001	118.42	0.26	<.0001	14.60	0.19	<.0001	1.43	1.21	.239
Q3	5.12	4.95	.303	54.11	0.48	<.0001	403.97	0.58	<.0001	25.92	0.62	<.0001	5.17	4.95	.299
Insurance															
Private	0.65	0.55	.240	4.40	0.05	<.0001	116.11	0.05	<.0001	3.21	0.06	<.0001	0.66	0.55	.231
Medicaid	1.57	1.10	.158	8.21	0.11	<.0001	252.38	0.07	<.0001	4.26	0.14	<.0001	1.54	1.10	.166
Uninsured	10.06	8.31	.229	4.23	1.08	<.001	610.09	0.70	<.0001	47.93	1.23	<.0001	10.24	8.30	.221
Small metropolitan	0.57	0.71	.426	24.45	0.11	<.0001	80.68	0.20	<.0001	14.12	0.12	<.0001	0.56	0.71	.433
Micropolitan	4.46	3.12	.156	133.55	0.53	<.0001	158.65	0.38	<.0001	62.27	0.72	<.0001	4.41	3.13	.162
Nonmetropolitan	9.80	16.97	.565	630.19	2.73	<.0001	231.92	2.11	<.0001	328.23	3.07	<.0001	9.82	16.97	.564
Congestive heart failure	0.02	0.13	.889	0.58	0.01	<.0001	105.80	0.03	<.0001	0.16	0.02	<.0001	0.03	0.13	.801
Cardiac valve disease	0.18	0.85	.834	4.27	0.07	<.0001	286.19	0.05	<.0001	2.01	0.09	<.0001	0.18	0.85	.835
Complicated diabetes	0.23	0.17	.181	0.82	0.02	<.0001	65.68	0.03	<.0001	0.69	0.02	<.0001	0.23	0.17	.174
Hypertension	0.15	0.43	.720	5.66	0.04	<.0001	383.93	0.05	<.0001	3.35	0.04	<.0001	0.16	0.42	.710
Electrolyte disorders	0.35	0.25	.170	0.14	0.02	<.0001	152.16	0.02	<.0001	0.06	0.03	.049	0.35	0.26	.172
Neurologic disorder	0.21	0.28	.450	1.71	0.03	<.0001	32.04	0.02	<.0001	1.57	0.03	<.0001	0.20	0.28	.469
Paralysis	0.03	0.27	.905	1.47	0.03	<.0001	21.46	0.02	<.0001	1.06	0.04	<.0001	0.03	0.27	.922
Vascular disease	0.05	0.12	.675	1.07	0.01	<.0001	21.48	0.04	<.0001	1.02	0.01	<.0001	0.04	0.12	.748
Renal failure	0.93	0.64	.147	14.85	0.08	<.0001	533.67	0.05	<.0001	6.23	0.09	<.0001	0.93	0.63	.147
Weight loss	0.00	0.25	.997	0.77	0.02	<.0001	77.46	0.01	<.0001	0.49	0.03	<.0001	0.01	0.25	.974
Obesity	0.43	0.82	.600	5.50	0.09	<.0001	474.78	0.09	<.0001	2.16	0.09	<.0001	0.44	0.82	.592
Deficiency anemia	0.35	0.25	.153	2.68	0.03	<.0001	154.79	0.04	<.0001	1.64	0.03	<.0001	0.36	0.25	.141
Diagnostic angiogram	0.05	0.07	.496	0.19	0.01	<.0001	124.25	0.01	<.0001	0.06	0.01	<.0001	0.05	0.07	.471
Elective	0.09	0.14	.511	1.18	0.01	<.0001	373.33	0.02	<.0001	0.84	0.01	<.0001	0.09	0.14	.503
LER volume/y															
Q1: 0-11	0.25	0.17	.131	0.51	0.01	<.0001	71.72	0.00	<.0001	0.04	0.01	.004	0.25	0.17	.141
Q2: 12-71	0.08	0.20	.694	0.94	0.02	<.0001	169.11	0.03	<.0001	0.35	0.02	<.0001	0.08	0.20	.687
Q3: 72-248	0.25	0.21	.227	2.05	0.02	<.0001	90.09	0.04	<.0001	0.99	0.03	<.0001	0.25	0.21	.233
Small hospital	0.88	0.50	.084	0.65	0.05	<.0001	282.13	0.04	<.0001	0.54	0.06	<.0001	0.89	0.51	.081
Medium hospital	1.42	1.04	.173	0.55	0.10	<.0001	154.33	0.12	<.0001	1.82	0.11	<.0001	1.44	1.03	.167
Midwest	0.05	21.10	.998	111.21	2.41	<.0001	1553.02	1.62	<.0001	101.16	2.38	<.0001	0.54	21.13	.980
South	0.62	0.63	.333	10.92	0.08	<.0001	108.66	0.10	<.0001	9.04	0.09	<.0001	0.62	0.63	.331
West	1.59	3.10	.610	8.09	0.49	<.0001	289.80	0.30	<.0001	66.48	0.57	<.0001	1.58	3.10	.611

CCA, Complete case analysis; *LER,* lower extremity revascularization; *MI,* multiple imputation; *MIV,* missing indicator variable; *PI,* Pacific Islander; *Q,* quartile; *RB,* relative bias; *SE,* standard error; *RF,* replacement with observed frequencies; *RWEE,* reweighted estimating equations.

[a]RB: 0%-5% = negligible, 5%-10% = minimal, 10%-20% = moderate, 20%-30% = heavy, >30% = severe.

Supplementary Table IV (online only). Missing at random: 5% missing

	CCA			*RF*			*MIV*			*MI*			*RWEE*		
Variable	*RB*[a]	*SE*	P	*RB*[a]	*SE*	P	*RB*[a]	*SE*	P	*RB*[a]	*SE*	P	*RB*[a]	*SE*	P
Age, year															
Q2: 62-70	0.25	0.20	.220	0.30	0.02	<.0001	0.26	0.04	<.0001	0.44	0.02	<.0001	0.41	0.20	.047
Q3: 71-78	0.55	0.22	.016	1.39	0.03	<.0001	1.37	0.06	<.0001	1.46	0.03	<.0001	0.22	0.23	.339
Q4: >78	0.20	0.52	.704	3.77	0.07	<.0001	3.72	0.12	<.0001	2.91	0.07	<.0001	0.54	0.53	.314
Race															
Black	0.36	0.06	<.0001	2.47	0.06	<.0001	1.79	0.06	<.0001	3.68	0.07	<.0001	0.19	0.06	.002
Hispanic	0.04	0.11	.737	2.30	0.12	<.0001	2.43	0.12	<.0001	0.76	0.12	<.0001	0.13	0.11	.261
Asian/PI	0.41	0.43	.344	1.26	0.44	.005	1.38	0.44	.003	8.84	0.45	<.0001	0.35	0.43	.428
Native American	0.30	0.36	.408	2.15	0.36	<.0001	1.45	0.36	.0001	1.48	0.36	<.0001	0.26	0.37	.481
Other	1.67	0.79	.037	2.59	0.78	.001	3.44	0.79	<.0001	16.49	0.80	<.0001	0.28	0.82	.736
Female sex	0.00	0.21	.990	1.80	0.03	<.0001	464.78	1.57	<.0001	1.66	0.03	<.0001	0.17	0.22	.442
Income															
Q1	0.10	0.31	.759	9.62	0.10	<.0001	198.42	0.05	<.0001	5.71	0.10	<.0001	0.79	0.32	.016
Q2	0.28	0.46	.553	4.14	0.08	<.0001	61.49	0.15	<.0001	2.92	0.08	<.0001	0.69	0.47	.147
Q3	2.85	1.64	.085	5.91	0.23	<.0001	307.24	0.43	<.0001	4.69	0.26	<.0001	1.66	1.69	.329
Insurance															
Private	0.54	0.18	.004	0.90	0.02	<.0001	111.00	0.04	<.0001	0.68	0.02	<.0001	0.24	0.19	.206
Medicaid	1.29	0.38	.001	1.52	0.07	<.0001	247.32	0.06	<.0001	0.80	0.07	<.0001	0.32	0.39	.414
Uninsured	0.98	3.07	.750	0.43	0.42	.308	565.80	0.56	<.0001	10.55	0.45	<.0001	1.49	3.15	.637
Small metropolitan	11.27	0.27	<.0001	4.43	0.06	<.0001	81.28	0.13	<.0001	3.17	0.06	<.0001	0.39	0.28	.163
Micropolitan	36.72	1.47	<.0001	26.92	0.31	<.0001	189.91	0.32	<.0001	13.31	0.31	<.0001	1.66	1.45	.256
Nonmetropolitan	111.04	6.37	<.0001	138.46	1.50	<.0001	38.81	2.11	<.0001	83.01	1.52	<.0001	10.11	6.53	.124
Congestive heart failure	0.39	0.05	<.0001	0.11	0.01	<.0001	101.68	0.02	<.0001	0.03	0.01	<.0001	0.07	0.05	.182
Cardiac valve disease	0.68	0.28	.017	0.93	0.03	<.0001	287.98	0.04	<.0001	0.60	0.03	<.0001	0.24	0.28	.390
Complicated diabetes	0.21	0.06	<.0001	0.11	0.01	<.0001	63.77	0.02	<.0001	0.16	0.01	<.0001	0.06	0.05	.281
Hypertension	0.57	0.15	<.0001	1.26	0.02	<.0001	383.12	0.04	<.0001	0.83	0.02	<.0001	0.09	0.15	.529
Electrolyte disorders	0.02	0.09	.811	0.04	0.01	.004	154.60	0.02	<.0001	0.00	0.01	.927	0.00	0.09	.988
Neurologic disorder	0.58	0.11	<.0001	0.41	0.02	<.0001	32.09	0.01	<.0001	0.38	0.02	<.0001	0.16	0.12	.172
Paralysis	0.37	0.10	<.0001	0.27	0.01	<.0001	22.54	0.02	<.0001	0.20	0.01	<.0001	0.15	0.10	.140
Vascular disease	0.70	0.05	<.0001	0.23	0.01	<.0001	220.02	0.03	<.0001	0.23	0.01	<.0001	0.05	0.05	.254
Renal failure	1.59	0.23	<.0001	3.19	0.04	<.0001	529.98	0.05	<.0001	1.58	0.05	<.0001	0.20	0.23	.405
Weight loss	0.48	0.09	<.0001	0.22	0.01	<.0001	79.74	0.01	<.0001	0.15	0.01	<.0001	0.14	0.10	.160
Obesity	3.45	0.30	<.0001	1.23	0.04	<.0001	472.51	0.08	<.0001	0.60	0.04	<.0001	0.21	0.30	.484
Deficiency anemia	0.58	0.09	<.0001	0.50	0.01	<.0001	151.00	0.04	<.0001	0.38	0.01	<.0001	0.11	0.09	.195
Diagnostic angiogram	0.17	0.03	<.0001	0.02	0.00	<.0001	123.94	0.01	<.0001	0.00	0.00	.791	0.01	0.03	.746
Elective	0.08	0.05	.074	0.21	0.01	<.0001	373.61	0.01	<.0001	0.17	0.01	<.0001	0.02	0.05	.714
LER volume/y															
Q1: 0-11	0.40	0.05	<.0001	0.06	0.01	<.0001	72.01	0.00	<.0001	0.02	0.01	.0041	0.01	0.05	.807
Q2: 12-71	0.17	0.07	.017	0.17	0.01	<.0001	167.99	0.04	<.0001	0.06	0.01	<.0001	0.04	0.07	.535
3: 72-248	0.07	0.08	.414	0.30	0.01	<.0001	88.40	0.03	<.0001	0.14	0.01	<.0001	0.03	0.08	.722
Small hospital	4.59	0.18	<.0001	0.25	0.02	<.0001	278.82	0.03	<.0001	0.20	0.02	<.0001	0.12	0.18	.518
Medium hospital	12.19	0.35	<.0001	1.00	0.04	<.0001	141.08	0.12	<.0001	0.35	0.04	<.0001	0.01	0.33	.982
Midwest	313.54	10.04	<.0001	41.98	1.26	<.0001	1358.75	1.94	<.0001	28.90	1.20	<.0001	7.54	9.70	.439
South	5.55	0.23	<.0001	3.12	0.04	<.0001	120.92	0.14	<.0001	2.51	0.04	<.0001	0.32	0.23	.155
West	2.24	0.85	.010	0.48	0.23	.043	251.48	0.23	<.0001	11.88	0.24	<.0001	4.82	0.86	<.0001

CCA, Complete case analysis; *LER,* lower extremity revascularization; *MI,* multiple imputation; *MIV,* missing indicator variable; *PI,* Pacific Islander; *Q,* quartile; *RB,* relative bias; *SE,* standard error; *RF,* replacement with observed frequencies; *RWEE,* reweighted estimating equations.
[a]RB: 0%-5% = negligible, 5%-10% = minimal, 10%-20% = moderate, 20%-30% = heavy, >30% = severe.

Supplementary Table V (online only). Missing at random: 15% missing

	CCA			RF			MIV			MI			RWEE		
Variable	*RB*[a]	*SE*	P	*RB*[a]	*SE*	P	*RB*[a]	*SE*	P	*RB*[a]	*SE*	P	*RB*[a]	*SE*	P
Age, years															
Q2: 62-70	0.69	0.36	.057	0.95	0.04	<.0001	0.86	0.05	<.0001	1.34	0.04	<.0001	0.11	0.38	.765
Q3: 71-78	2.91	0.46	<.0001	4.58	0.05	<.0001	4.28	0.08	<.0001	4.52	0.05	<.0001	0.30	0.49	.543
Q4: >78	4.01	0.83	<.0001	12.63	0.10	<.0001	11.54	0.19	<.0001	9.20	0.11	<.0001	0.95	0.91	.296
Race															
Black	0.83	0.09	<.0001	8.59	0.10	<.0001	5.69	0.10	<.0001	11.93	0.10	<.0001	0.03	0.10	.774
Hispanic	0.43	0.25	.092	10.17	0.27	<.0001	8.68	0.27	<.0001	1.66	0.25	<.0001	0.04	0.27	.867
Asian/PI	0.38	0.82	.644	0.59	0.81	.470	5.09	0.83	<.0001	27.21	0.80	<.0001	0.60	0.84	.480
Native American	0.77	0.64	.229	7.21	0.62	<.0001	4.44	0.64	<.0001	2.78	0.58	<.0001	1.47	0.69	.035
Other	3.94	1.48	.009	10.23	1.41	<.0001	10.73	1.49	<.0001	59.12	135	<.0001	2.06	1.62	.209
Female sex	1.05	0.38	.007	6.01	0.05	<.0001	465.52	1.02	<.0001	5.09	0.05	<.0001	0.52	0.41	.213
Income															
Q1	3.06	0.52	<.0001	32.87	0.14	<.0001	194.45	0.08	<.0001	18.46	0.16	<.0001	0.56	0.55	.313
Q2	1.27	0.91	.166	15.02	0.13	<.0001	93.96	0.25	<.0001	9.42	0.14	<.0001	0.13	0.99	.897
Q3	15.25	3.23	<.0001	23.18	0.35	<.0001	346.26	0.81	<.0001	15.53	0.41	<.0001	0.54	3.40	.875
Insurance															
Private	0.87	0.37	.020	3.06	0.04	<.0001	112.54	0.08	<.0001	2.23	0.04	<.0001	0.38	0.40	.349
Medicaid	5.77	0.71	<.0001	5.47	0.10	<.0001	250.23	0.10	<.0001	2.59	0.10	<.0001	0.96	0.77	.214
Uninsured	1.40	5.91	.813	1.29	0.70	.067	592.04	0.82	<.0001	33.52	0.76	<.0001	0.52	6.30	.935
Small metropolitan	32.57	0.55	<.0001	15.29	0.09	<.0001	80.40	0.19	<.0001	10.14	0.10	<.0001	0.29	0.56	.603
Micropolitan	96.29	3.01	<.0001	91.92	0.55	<.0001	81.66	0.61	<.0001	43.51	0.53	<.0001	2.77	3.12	.377
Nonmetropolitan	275.35	13.80	<.0001	465.06	2.26	<.0001	508.08	3.57	<.0001	265.48	2.39	<.0001	20.36	15.08	.180
Congestive heart failure	1.58	0.08	<.0001	0.38	0.01	<.0001	107.74	0.04	<.0001	0.09	0.01	<.0001	0.03	0.09	.752
Cardiac valve disease	2.93	0.58	<.0001	3.07	0.05	<.0001	287.08	0.06	<.0001	1.76	0.05	<.0001	0.62	0.61	.312
Complicated diabetes	1.01	0.11	<.0001	0.39	0.01	<.0001	63.00	0.04	<.0001	0.50	0.01	<.0001	0.09	0.12	.458
Hypertension	1.65	0.26	<.0001	4.28	0.03	<.0001	386.35	0.06	<.0001	2.64	0.04	<.0001	0.04	0.27	.872
Electrolyte disorders	0.03	0.16	.871	0.12	0.02	<.0001	153.06	0.03	<.0001	0.01	0.02	.620	0.01	0.18	.963
Neurologic disorder	1.49	0.17	<.0001	1.36	0.02	<.0001	32.04	0.03	<.0001	1.22	0.03	<.0001	0.09	0.19	.653
Paralysis	1.76	0.18	<.0001	0.98	0.02	<.0001	21.82	0.03	<.0001	0.67	0.03	<.0001	0.00	0.19	.981
Vascular disease	2.44	0.09	<.0001	0.79	0.01	<.0001	220.76	0.04	<.0001	0.71	0.01	<.0001	0.05	0.11	.668
Renal failure	7.36	0.48	<.0001	10.86	0.06	<.0001	532.35	0.08	<.0001	5.00	0.07	<.0001	0.94	0.52	.076
Weight loss	1.36	0.19	<.0001	0.75	0.02	<.0001	78.32	0.02	<.0001	0.50	0.02	<.0001	0.06	0.20	.768
Obesity	10.84	0.60	<.0001	3.97	0.06	<.0001	474.04	0.12	<.0001	1.58	0.06	<.0001	0.07	0.63	.911
Deficiency anemia	1.70	0.19	<.0001	1.72	0.02	<.0001	149.04	0.07	<.0001	1.21	0.02	<.0001	0.18	0.20	.385
Diagnostic angiogram	0.58	0.06	<.0001	0.07	0.01	<.0001	124.61	0.01	<.0001	0.02	0.01	<.0001	0.03	0.06	.649
Elective	0.05	0.10	.623	0.70	0.01	<.0001	373.20	0.02	<.0001	0.51	0.01	<.0001	0.17	0.10	.100
LER volume/y															
Q1: 0-11	1.52	0.10	<.0001	0.27	0.01	<.0001	71.90	0.01	<.0001	0.02	0.01	.150	0.07	0.11	.516
Q2: 12-71	0.76	0.12	<.0001	0.67	0.02	<.0001	168.72	0.06	<.0001	0.06	0.02	.000	0.23	0.14	.095
Q3: 72-248	0.22	0.17	.196	1.19	0.02	<.0001	89.15	0.06	<.0001	0.63	0.02	<.0001	0.14	0.18	.443
Small hospital	13.91	0.37	<.0001	1.02	0.04	<.0001	280.27	0.05	<.0001	0.61	0.05	<.0001	0.05	0.38	.894
Medium hospital	35.69	0.72	<.0001	3.47	0.07	<.0001	119.10	0.23	<.0001	1.08	0.07	<.0001	0.52	0.71	.470
Midwest	789.50	18.28	<.0001	121.80	1.88	<.0001	956.38	3.43	<.0001	74.12	2.15	<.0001	52.77	18.81	.006
South	15.56	0.49	<.0001	10.23	0.06	<.0001	149.62	0.22	<.0001	8.30	0.05	<.0001	3.31	0.52	.0001
West	20.44	1.70	<.0001	3.021	0.45	<.0001	253.20	0.40	<.0001	35.79	0.42	<.0001	28.43	1.77	<.0001

CCA, Complete case analysis; *LER,* lower extremity revascularization; *MI,* multiple imputation; *MIV,* missing indicator variable; *PI,* Pacific Islander; *Q,* quartile; *RB,* relative bias; *SE,* standard error; *RF,* replacement with observed frequencies; *RWEE,* reweighted estimating equations.
[a]RB: 0%-5% = negligible, 5%-10% = minimal, 10%-20% = moderate, 20%-30% = heavy, >30% = severe.

Supplementary Table VI (online only). Missing at random: 30% missing

	CCA			*RF*			*MIV*			*MI*			*RWEE*		
Variable	*RB*[a]	*SE*	P	*RB*[a]	*SE*	P	*RB*[a]	*SE*	P	*RB*[a]	*SE*	P	*RB*[a]	*SE*	P
Age, years															
Q2: 62-70	2.12	0.52	<.0001	1.52	0.04	<.0001	1.41	0.08	<.0001	2.11	0.05	<.0001	1.28	0.57	.028
Q3: 71-78	5.07	0.60	<.0001	7.51	0.05	<.0001	7.21	0.09	<.0001	7.16	0.07	<.0001	0.77	0.68	.256
Q4: >78	9.24	1.33	<.0001	20.81	0.12	<.0001	19.59	0.21	<.0001	14.55	0.14	<.0001	4.31	1.54	.006
Race															
Black	1.56	0.16	<.0001	14.87	0.17	<.0001	9.33	0.17	<.0001	19.88	0.15	<.0001	0.14	0.18	.440
Hispanic	0.99	0.32	.003	18.86	0.33	<.0001	13.82	0.34	<.0001	2.16	0.31	<.0001	0.15	0.36	.685
Asian/PI	0.72	1.23	.557	6.18	1.18	<.0001	7.31	1.24	<.0001	40.59	1.09	<.0001	1.70	1.26	.180
Native American	2.99	0.99	.003	13.76	0.96	<.0001	8.99	0.99	<.0001	5.01	0.76	<.0001	0.53	1.09	.624
Other	11.56	2.00	<.0001	24.81	1.90	<.0001	22.73	2.05	<.0001	92.72	1.63	<.0001	3.12	2.26	.172
Female sex	0.47	0.57	.413	9.72	0.06	<.0001	468.24	0.83	<.0001	7.89	0.07	<.0001	0.41	0.64	.525
Income															
Q1	4.89	0.77	<.0001	54.06	0.18	<.0001	191.06	0.09	<.0001	29.69	0.21	<.0001	0.11	0.83	.893
Q2	2.02	1.25	.109	26.30	0.14	<.0001	122.96	0.34	<.0001	15.53	0.16	<.0001	0.43	1.39	.759
Q3	19.74	4.54	<.0001	41.92	0.48	<.0001	387.23	0.95	<.0001	25.27	0.54	<.0001	6.72	5.10	.191
Insurance															
Private	1.95	0.51	<.001	5.06	0.04	<.0001	114.37	0.10	<.0001	3.59	0.05	<.0001	0.14	0.57	.805
Medicaid	7.98	0.95	<.0001	9.03	0.11	<.0001	253.09	0.13	<.0001	4.08	0.14	<.0001	0.07	1.10	.952
Uninsured	14.98	8.08	.067	1.91	0.80	.018	610.37	1.01	<.0001	52.01	1.01	<.0001	5.47	9.13	.551
Small metropolitan	46.33	0.73	<.0001	25.84	0.11	<.0001	80.38	0.25	<.0001	16.60	0.13	<.0001	0.39	0.76	.613
Micropolitan	136.88	4.24	<.0001	151.58	0.56	<.0001	1.76	0.73	.018	71.04	0.73	<.0001	2.88	4.34	.508
Nonmetropolitan	378.96	17.86	<.0001	752.32	2.82	<.0001	915.44	3.34	<.0001	422.77	3.59	<.0001	44.32	19.27	.024
Congestive heart failure	2.63	0.12	<.0001	0.65	0.01	<.0001	112.40	0.05	<.0001	0.20	0.01	<.0001	0.01	0.13	.936
Cardiac valve disease	2.51	0.78	.002	5.01	0.06	<.0001	286.19	0.08	<.0001	2.66	0.07	<.0001	0.35	0.83	.674
Complicated diabetes	2.25	0.15	<.0001	0.70	0.02	<.0001	62.63	0.05	<.0001	0.76	0.02	<.0001	0.26	0.16	.105
Hypertension	3.26	0.40	<.0001	6.80	0.04	<.0001	388.54	0.08	<.0001	4.08	0.05	<.0001	0.65	0.44	.144
Electrolyte disorders	0.11	0.24	.639	0.17	0.02	<.0001	151.81	0.04	<.0001	0.03	0.03	.233	0.00	0.27	.986
Neurologic disorder	2.55	0.27	<.0001	2.15	0.03	<.0001	32.04	0.02	<.0001	1.94	0.03	<.0001	0.44	0.29	.135
Paralysis	2.49	0.28	<.0001	1.65	0.02	<.0001	21.26	0.04	<.0001	1.10	0.03	<.0001	0.38	0.30	.202
Vascular disease	4.00	0.11	<.0001	1.28	0.01	<.0001	221.46	0.06	<.0001	1.08	0.01	<.0001	0.02	0.13	.887
Renal failure	11.64	0.71	<.0001	17.59	0.08	<.0001	534.45	0.09	<.0001	7.90	0.11	<.0001	0.79	0.79	.323
Weight loss	2.19	0.24	<.0001	1.13	0.02	<.0001	77.02	0.02	<.0001	0.73	0.03	<.0001	0.06	0.27	.811
Obesity	14.75	0.91	<.0001	6.71	0.06	<.0001	475.23	0.14	<.0001	2.59	0.08	<.0001	1.41	1.00	.164
Deficiency anemia	2.55	0.23	<.0001	2.86	0.03	<.0001	148.43	0.07	<.0001	1.96	0.03	<.0001	0.15	0.25	.550
Diagnostic angiogram	0.89	0.09	<.0001	0.16	0.01	<.0001	125.15	0.01	<.0001	0.01	0.01	.061	0.13	0.10	.166
Elective	0.31	0.15	.046	1.19	0.01	<.0001	372.74	0.02	<.0001	0.81	0.01	<.0001	0.03	0.17	.846
LER volume/y															
Q1: 0-11	2.43	0.13	<.0001	0.50	0.01	<.0001	71.77	0.01	<.0001	0.01	0.01	.666	0.14	0.14	.325
Q2: 12-71	1.41	0.18	<.0001	1.08	0.02	<.0001	169.39	0.07	<.0001	0.14	0.02	<.0001	0.42	0.20	.042
Q3: 72-248	0.84	0.19	<.0001	2.05	0.02	<.0001	90.03	0.06	<.0001	1.12	0.03	<.0001	0.52	0.21	.014
Small hospital	20.35	0.48	<.0001	1.63	0.05	<.0001	281.88	0.08	<.0001	0.79	0.05	<.0001	0.15	0.53	.770
Medium hospital	52.97	0.79	<.0001	4.94	0.09	<.0001	104.88	0.26	<.0001	1.08	0.10	<.0001	0.34	0.87	.697
Midwest	1103.13	29.82	<.0001	165.59	2.02	<.0001	676.05	3.98	<.0001	118.39	2.57	<.0001	17.50	32.38	.590
South	25.81	0.69	<.0001	15.78	0.06	<.0001	164.26	0.24	<.0001	12.77	0.08	<.0001	3.02	0.76	<.001
West	25.70	2.51	<.0001	6.54	0.53	<.0001	258.14	0.44	<.0001	54.16	0.55	<.0001	37.14	2.73	<.0001

CCA, Complete case analysis; *LER,* lower extremity revascularization; *MI,* multiple imputation; *MIV,* missing indicator variable; *PI,* Pacific Islander; *Q,* quartile; *RB,* relative bias; *SE,* standard error; *RF,* replacement with observed frequencies; *RWEE,* reweighted estimating equations.
[a]RB: 0%-5% = negligible, 5%-10% = minimal, 10%-20% = moderate, 20%-30% = heavy, >30% = severe.