Missing Data and Interpretation of Cancer Surgery Outcomes at the American College of Surgeons National Surgical Quality Improvement Program

Helen M Parsons, MPH, William G Henderson, PhD, MPH, Jeanette Y Ziegenfuss, PhD, Michael Davern, PhD, MA, Waddah B Al-Refaie, MD, FACS

BACKGROUND: The American College of Surgeons National Surgical Quality Improvement Program (ACS NSQIP) has become an important surgical quality program in the United States, yet few studies describe their methods for handling missing data. Our study examines the impact of missing data on predictive models for short-term operative outcomes after cancer surgery in the ACS NSQIP database.

STUDY DESIGN: We identified 97,230 patients who underwent oncologic resections for neoplasms in the 2005–2009 ACS NSQIP. We used multivariable logistic regression to assess the impact of pre-, intra-, and postoperative factors on short-term operative outcomes by type of procedure where missing values were included as a variable category, excluded, and imputed.

RESULTS: A large proportion (72.8%) of patients had one or more missing pre-, intra-, or postoperative characteristics, particularly preoperative laboratory values. Missing data were more frequent in healthier patients and those undergoing lower-risk procedures. Although data were not missing at random, the impact of preoperative risk factors on adverse operative outcomes after cancer surgery was similar across methods for handling missing data. However, analytic approaches using only patients with complete or imputed information risk basing the analysis on a potentially nonrepresentative sample.

CONCLUSIONS: Missing data present challenges to interpreting predictors of short-term operative outcomes after cancer surgery at ACS NSQIP hospitals. Similar to best practices for other data sets, this study highlights the importance of using missing values carefully when using ACS NSQIP. Given its potential to introduce bias, the approach to handling missing values should be detailed in future ACS NSQIP studies. (J Am Coll Surg 2011;213:379–391. © 2011 by the American College of Surgeons)

The American College of Surgeon's National Surgical Quality Improvement Program (ACS NSQIP) is considered an important surgical quality program for general and vascular surgery in the United States.[1-5] With ongoing efforts to improve the quality of cancer surgery outcomes, ACS NSQIP provides an opportunity for service lines and their surgeons to identify those factors predictive of operative outcomes after oncologic resections. Compared with single-center data sets, ACS NSQIP provides a wide range of pre-, intra-, and postoperative patient characteristics to aid surgeons in evaluating quality and examining predictors of operative outcomes in a large, multihospital setting.[6]

Although few published studies using ACS NSQIP fully describe their methods of handling missing data, the presence of missing data has the potential to substantially influence conclusions if not properly addressed.[7-9] Although present in almost all databases, missing data can present challenges to conducting high-quality research for 2 main reasons: the potential for biased results due to differences in patients with observed and unobserved characteristics and the variety of methodological choices for handling missing data.[10-13] Considering that the amount of missing data can vary by both the extent of cancer surgery or a patient's

Disclosure Information: Nothing to disclose.

Supported by The 2008 Veterans of Foreign Wars and Ladies Cancer Research Center Endowment Fund.

Received February 1, 2011; Revised March 10, 2011; Accepted April 10, 2011.

From the University of Minnesota and Minneapolis Veteran Affairs Medical Center, Minneapolis, MN (Parsons, Al-Refaie), The Colorado Health Outcomes Program, The University of Colorado, Aurora, CO (Henderson), Mayo Clinic, Rochester, MN (Ziegenfuss), and the National Opinion Research Center, The University of Chicago, Chicago, IL (Davern).

Correspondence address: Waddah B Al-Refaie, MD, FACS, Division of Surgical Oncology, Department of Surgery, Mayo Mail Code 195, University of Minnesota and Minneapolis Veterans Affairs Medical Center, 420 Delaware St SE, Minneapolis, MN 55455. email: alref003@umn.edu

ISSN 1072-7515/11/$36.00
doi:10.1016/j.jamcollsurg.2011.05.010

Abbreviations and Acronyms

ACS NSQIP = American College of Surgeon's National Surgical Quality Improvement Program
OR = odds ratio
PTT = partial thromboplastin time

comorbidities, identifying how different assumptions about missing data affect study results should be an important consideration in future ACS NSQIP studies. As a result of these concerns, several statistical methods have been proposed to appropriately use patients with missing data in other settings.[10-12,14-18] However, to date, no study has fully examined the impact of missing data found in predictors of adverse outcomes after cancer surgery within ACS NSQIP.

In our study, we examine the impact of different methods for handling missing data on the importance of predictors of short-term operative outcomes after cancer surgery in the ACS NSQIP database. We hypothesize that the impact of pre-, intra-, and postoperative patient characteristics on short-term operative outcomes will vary according to the method for handling missing data.

METHODS

Data source

For our study, we used the 2005–2009 ACS NSQIP Participant Use File, a large, prospective, multihospital database.[6] Currently, ACS NSQIP collects detailed patient data for those undergoing both inpatient and outpatient surgical procedures in >250 participating university and private-sector medical centers. ACS NSQIP captures data on preoperative risk factors and laboratory values, intraoperative variables, and 30-day postoperative mortality and morbidity outcomes. Approval to conduct our study was obtained from both ACS NSQIP and the University of Minnesota Institutional Review Board. ACS NSQIP is not responsible for the statistical validity of the data analysis or the conclusions derived by users of the data.

Patients

Between 2005 and 2009, we identified 97,230 patients older than 18 years of age who were surgically treated for thoracic, abdominal, or pelvic neoplasms according to ICD-9 diagnoses codes within ACS NSQIP and who underwent elective and nonelective resection for their cancer according to prespecified Current Procedure Terminology codes. Patients were categorized into low-, intermediate-, or high-risk procedures to examine the impact of different methods for handling missing data on predictors of short-term operative outcomes across different types of surgical procedures. Low-risk procedures included mastectomy, thyroid lobectomy, and thyroidectomy. Intermediate-risk procedures included colectomy and proctectomy. High-risk procedures included pneumonectomy and lobectomy, espophagectomy, total and distal gastrectomy, pancreaticoduodenectomy or total pancreatectomy, open distal pancreatectomy, hepatic and bile duct resection, abdominoperineal resection and pelvic exenteration, radical nephrectomy, and urinary cystectomy.

Outcomes definitions

We examined 30-day mortality, development of major complications, and total number of postoperative complications. However, because those who underwent low-risk procedures experienced extremely low mortality and few complications, we only analyzed the development of major complications for this group.

Major complications included deep incisional surgical site infection, organ/space surgical site infection, wound disruptions, pneumonia, need for reintubation, pulmonary embolism, progressive renal insufficiency, acute renal failure, cerebrovascular accidents, coma, peripheral nerve injury, cardiac arrest, myocardial infarction, graft/flap failure, sepsis, septic shock, the need for return to the operating room, or longer than 48 hours on a ventilator. Minor complications included superficial surgical site infections, urinary tract infections, and deep vein thrombosis/thrombophlebitis. The total number of complications consisted of the sum of the major and minor complications occurring up to 30 days after the operation, with the potential for multiple complications within each category. Thirty-day mortality was death from any cause measured from the date of procedure to 30 days after the procedure.

Statistical analyses

For our bivariable analysis, we compared demographics, preoperative factors, preoperative laboratory values, intraoperative data, 30-day operative mortality, and postoperative complications by surgical risk category using chi-square tests for categorical data and the ANOVA test for continuous outcomes.

To evaluate the impact of different methods for handling missing values on the predictors of short-term operative outcomes after cancer surgery, we used 3 common approaches for accounting for missing data in the statistical analysis: complete case analysis, coding a missing value within each predictor variable, and hotdeck imputation for all missing values.[11,14,18] Complete case analysis involves analysis of only those patients for whom data exist for all predictors of interest, and all other cases are excluded from evaluation. The second option involves creating a category for missing values within each predictor (ie, "missing

value"), allowing for the use of data for all patients. Imputation involves creating a set of plausible values for each missing variable of interest, selecting that value based on data from other similar patients in the database.

For each surgical risk group, we ran 3 analyses to evaluate the impact of each of these 3 methods for handling missing data on predictors of short-term operative outcomes. First, we ran a logistic model predicting major complications using complete case analyses (deleting all patients with 1 or more missing values for a pre-, intra-, or postoperative predictor). Second, we ran the same logistic model, including any patient with 1 or more missing values for a predictor categorized as missing for the variable of interest. Finally, we ran a logistic model using imputed values for all missing variables using the hotdeck method of imputation.

Hotdeck imputation is a 3-step process.[18] First, a set of plausible values for each missing variable is created based on values of that variable for similar individuals in the data set. In our study, we used age, sex, and American Society of Anesthesiologists score to predict missing values for each of the missing variables of interest. Next, this set of plausible values was used to fill in the missing values in the data set and create a complete data set with no missing information. However, the hotdeck procedure does retain the relative frequency of values for each variable, so the distribution of patient characteristics remains similar before and after imputation.[18] A set of complete data with imputed values is then analyzed using multivariable techniques.

When using ACS NSQIP, researchers should ideally use a multiple imputation framework, where patients with missing data take on multiple potential values for that missing variable.[19] When evaluated using multivariable techniques, the model incorporates the uncertainty of these values by producing larger standard errors relative to a nonimputed model. The purpose of our analysis in this article attempts to show the overall impact of various techniques for handling missing data on point estimates, so we did not use multiple imputation for simplification. As a result, our hotdeck analysis will have underestimated standard errors.

After conducting multivariable analyses using each of our 3 methods for handling missing data, we compared variation in odds ratio estimates, C-indices, and Hosmer-Lemeshow statistics across models. Model fit was first evaluated using the C-index, a measure of a model's ability to discriminate between events and nonevents. A C-index of 0.5 indicates no predictive ability for a model. In other words, the predictive ability of the model is no better than would be obtained by chance. A C-index of 1 indicates perfect predictive ability, where all outcomes of interest would be correctly identified according to the model. C-indices derived from multivariate models fall between 1 of these 2 extreme values. Finally, we evaluated model calibration using the Hosmer-Lemeshow test, which examines how well the percentage of observed outcomes matches the percentage of predicted outcomes over deciles of risk. A significant test implies that the model is not well-calibrated. To examine the extent of agreement between the 3 methods of handling missing values in the multivariable models, we also computed Spearman rank correlations of the patients' predicted probabilities of major complications (in quintiles) between the 3 models. As rank is known to be overly sensitive to small changes in performance estimates, we present correlations of complication probabilities by risk quintile rather than absolute risk changes between methods for handling missing data.

All analyses were performed using Stata IC version 10.1 (StataCorp).

RESULTS

Of the 97,230 patients identified in our database search who underwent resections for thoracic, abdominal, and pelvic neoplasms, 52.7% underwent low-risk, 27.1% underwent intermediate-risk, and 20.2% underwent high-risk procedures. Patients undergoing lower-risk procedures were more likely to be younger, female, to never have been smokers, with a lower American Society of Anesthesiologists score and fewer overall comorbidities ($p < 0.0001$ for all) (Table 1).

Preoperative laboratory values

Distribution of preoperative laboratory values varied considerably by level of surgical risk (Table 1). Most notably, the proportion of missing values for patients undergoing low-risk procedures was considerably higher than patients undergoing high-risk operations. Within low-risk operations, the proportion of missing data varied from 17.5% of patients missing values for hematocrit to 79.3% missing values for activated partial thromboplastin time (PTT). Within high-risk procedures, the variation in missing values was considerably less, from 2.8% of patients missing values of hematocrit to 27.8% missing values for PTT.

Short-term operative outcomes

Overall 30-day mortality rate within the cohort was relatively low (Table 1). However, there was significant variation in mortality rates between levels of surgical risk (0.1% in low-risk procedures vs 2.8% in high-risk procedures; $p < 0.0001$). In addition, the disparity in complication rates was significant between levels of operative risk. In the cohort, major complications developed in 8.6% of patients undergoing low-risk surgery and in 25.0% of patients undergoing high-risk procedures ($p < 0.0001$). More than 1

Table 1. Patient Preoperative Characteristics and Short-Term Operative Outcomes by Complexity of Surgery

Preoperative patient characteristics	Low risk (n = 51,262)	Intermediate risk (n = 26,331)	High risk (n = 19,637)	p Value
Age (y), %				<0.0001
< 50	33.70	12.50	17.20	
51–60	25.40	19.50	22.60	
61–70	21.40	23.90	29.00	
≥70	19.50	44.10	31.20	
White race, %	73.80	74.10	75.40	<0.0001
Female, %	94.20	49.50	47.30	<0.0001
Never smoker, %	74.40	67.40	20.70	<0.0001
ASA, %				<0.0001
1+2	72.20	45.30	31.70	
3+	27.80	54.70	68.30	
Cardiac conditions, % yes	3.80	13.90	11.40	<0.001
Hypertension, % yes	40.10	56.20	51.90	<0.0001
Diabetes, % yes	10.30	17.70	18.80	<0.001
Pulmonary condition, % yes	8.70	17.00	14.80	<0.0001
Preoperative laboratory values, %				
Sodium, mmol/L				<0.0001
≤135	3.80	10.10	10.40	
136–144	66.90	80.40	82.90	
≥145	1.70	2.80	2.40	
Missing	27.60	6.70	4.30	
BUN				<0.0001
≤39	69.80	89.30	92.60	
Missing	29.50	9.10	6.20	
Creatinine, mg/dL				<0.0001
≤1.5	73.60	5.00	3.90	
Missing	24.90	5.80	3.50	
Albumin, g/dL				<0.0001
>2.6	48.60	65.90	78.80	
Missing	51.20	30.50	18.00	
Total bilirubin, mg/dL				<0.0001
<1.0	47.10	62.70	63.30	
Missing	49.00	28.80	17.00	
SGOT				<0.0001
≤39	49.40	65.70	62.70	
Missing	47.30	28.60	16.50	
Alkaline phosphatase, U/L				<0.0001
≤124	50.10	65.40	60.20	
Missing	47.00	27.80	16.10	
White blood cell count, K/uL				<0.0001
≤4.4	7.00	9.00	9.60	
4.5–10.9	68.70	77.50	80.20	
≥11.0	4.00	9.10	7.30	
Missing	20.30	4.40	2.90	
Hematocrit				<0.001
≤37	26.40	54.30	43.80	
38–44	51.30	36.00	46.10	

(continued)

Table 1. Continued

Preoperative patient characteristics	Low risk (n = 51,262)	Intermediate risk (n = 26,331)	High risk (n = 19,637)	p Value
≥45	4.80	6.20	7.30	
Missing	17.50	3.50	2.80	
Platelet count, K/uL				<0.0001
≤149	2.30	4.90	8.10	
150–399	74.00	79.50	80.80	
≥400	3.40	11.20	8.10	
Missing	20.30	4.40	3.00	
PTT				<0.0001
≤34	19.20	44.50	64.90	
Missing	79.30	50.30	27.80	
Short-term operative outcomes, %				
30-d mortality	0.10	2.60	2.80	<0.0001
Total no. postoperative complications (major or minor)				<0.0001
0	89.80	78.60	71.80	
1	8.90	13.40	14.30	
≥2	1.3	8.00	13.90	
Major complications				<0.0001
Yes	8.60	16.60	25.00	
Total operative time, min				
Mean ± SD	124.4 ± 92.1	163.7 ± 85.7	282.1 ± 138.3	<0.0001
Median	101	146	261	<0.0001
Total anesthesia time, min				
Mean ± SD	173.8 ± 102.2	224.8 ± 94.7	362.2 ± 149.2	<0.0001
Median	149	208	340	<0.0001

The following variables were excluded from analyses because <5% of patients fell into the following categories: nonindependent functional status; admission from acute/chronic care facility; alcohol use >2 drinks/day; DNR order; presence of disseminated cancer; contaminated/dirty wound class; presence of bleeding disorders; >4 U RBCs transfused before surgery; preoperative sepsis; emergency surgery; presence of airway trauma; presence of renal conditions; earlier operation within 30 days.
ASA, American Society of Anesthesiologists; OR, odds ratio; PTT, partial thromboplastin time; SGOT, serum glutamic oxaloacetic transaminase.

complication developed in <0.5% of patients undergoing a low-risk procedure, vs 13.9% in the high-risk procedure group (Table 1).

Evaluating predictors of short-term operative outcomes

High-risk procedures

In patients undergoing a high-risk procedure, 44.5% had 1 or more missing values for a pre-, intra-, or postoperative characteristic (Table 2). Analyzing predictors of major complications after excluding those with missing values resulted in use of only 55.5% of the patient population (n = 10,892 of 19,637). Comparing the logistic regression models that included (n = 19,637) or excluded (n = 10,982) patients with missing values, we found both models identified similar pre-, intra-, and postoperative predictors of short-term operative outcomes. Although each analysis produced estimated odds ratios that differed up to 0.33 for the same predictor (BUN ≥40 vs <39; odds ratio [OR] = 1.06; 95% CI, 0.78–1.43, for patients with missing values included, vs OR = 1.39; 95% CI, 0.95–2.05, for complete case analysis), the 95% CIs overlapped for most predictors when comparing models. As a result, interpretations of important risk factors for short-term operative outcomes would remain similar between the 2 analyses. However, the same comparison did identify differences in the interpretation of a few predictors. For example, when all patients were included in the analysis, with missing values as a category, creatinine was a significant predictor of major complications (OR = 1.22; 95% CI, 1.03–1.45). Using complete case analysis (ie, excluding those with unknown creatinine levels or other predictors), the significance of creatinine as a predictor of major complications was diminished (OR = 1.06; 95% CI, 0.84–1.35). Again, however, the 95% CIs for creatinine overlapped in both models. Although results remained consistent between methods, complete case analysis resulted in a substantially

Table 2. Predictors of Major Complications in High-Risk Surgery

	Missing values included (n = 19,637)		Missing values excluded (n = 10,892)		Missing values imputed (n = 19,637)	
Predictor	**OR**	**95% CI**	**OR**	**95% CI**	**OR**	**95% CI**
Age, y						
51–60 vs <50	1.02	0.91–1.14	1.02	0.88–1.18	1.01	0.90–1.13
61-70 vs <50	1.03	0.92–1.15	0.99	0.86–1.15	1.01	0.90–1.13
>70 vs <50*	1.16	1.03–1.3	1.16	1.01–1.35	1.16	1.04–1.30
Race						
African American vs white	0.95	0.84–1.08	1.04	0.89–1.21	1.02	0.91–1.15
Hispanic vs white	0.95	0.81–1.12	0.99	0.81–1.22	0.99	0.85–1.15
Other vs white	1.05	0.9–1.23	1.15	0.95–1.4	1.08	0.93–1.25
Missing vs white	0.96	0.84–1.1				
Sex						
Male vs female*	1.3	1.22–1.4	1.27	1.16–1.4	1.31	1.22–1.40
Smoking status						
Current vs never*	1.17	1.07–1.28	1.18	1.05–1.32	1.17	1.07–1.27
Previous vs never	1.06	0.98–1.15	1.1	0.99–1.23	1.05	0.97–1.15
ASA						
3+ vs 1-2*	1.38	1.27–1.49	1.46	1.32–1.63	1.42	1.31–1.54
Cardiac conditions						
Yes vs no	1.16*	1.04*–1.28*	1.13	0.99–1.29	1.17*	1.06*–1.30*
Hypertension						
Yes vs no	1.05	0.98–1.13	1.04	0.94–1.14	1.06	0.98–1.14
Diabetes						
Yes vs no	1.07	0.98–1.16	1.03	0.92–1.15	1.08	0.99–1.18
Pulmonary conditions*						
Yes vs no	1.4	1.28–1.53	1.32	1.17–1.48	1.41	1.29–1.55
Sodium, mmol/L						
≤135 vs 136–144	1.15*	1.04*–1.28*	1.13	0.98–1.29	1.09	0.98–1.21
≥145 vs 136–144	1.3*	1.05*–1.6*	1.27	0.96–1.68	1.13	0.91–1.40
Missing vs ≤136–144	1.11	0.83–1.5				
BUN						
≥40 vs ≤39	1.06	0.78–1.43	1.39	0.95–2.05	0.95	0.79–1.14
Missing vs ≤39	0.65*	0.51*–0.83*				
Creatinine, mg/dL						
>1.6 vs ≤1.5	1.22*	1.03*–1.45*	1.06	0.84–1.35	1.29*	1.10*–1.51*
Missing vs ≤1.5	1.26	0.9–1.76				
Albumin, g/dL						
≤2.5 vs >2.6*	1.42	1.2–1.69	1.58	1.29–1.94	1.52	1.05–2.19
Missing vs >2.6	0.87	0.73–1.03				
Total bilirubin, mg/dL						
≥1.0 vs <1.0	1.09	0.99–1.2	1.11	0.99–1.25	1.13*	1.03*–1.23*
Missing vs <1.0	1.09	0.88–1.35				
SGOT						
≥40 vs ≤39	0.97	0.88–1.07	1.02	0.9–1.15	1.01	0.92–1.10
Missing vs ≤39	1.02	0.81–1.28				
Alkaline phosphatase, U/L						
≥125 vs ≤124	1.18*	1.08*–1.3*	1.17*	1.04*–1.31*	1.08	0.99–1.18
Missing vs ≤124	1.01	0.79–1.27				

(continued)

Table 2. Continued

	Missing values included (n = 19,637)		Missing values excluded (n = 10,892)		Missing values imputed (n = 19,637)	
Predictor	**OR**	**95% CI**	**OR**	**95% CI**	**OR**	**95% CI**
White blood cell count, K/uL						
≤4.4 vs 4.5–10.9	0.97	0.86–1.1	1.01	0.85–1.17	1.06	0.94–1.19
≥11.0 vs 4.5–10.9	1.24*	1.1*–1.4*	1.18*	1.01*–1.39*	1.06	0.93–1.19
Missing vs 4.5–0.9	1.18	0.6–2.09				
Hematocrit, %						
≤37 vs 38–44	1.16*	1.08*–1.25*	1.1	0.99–1.21	1.03	0.96–1.10
≥45 vs 38–44	1.07	0.94–1.22	1.04	0.87–1.24	0.95	0.83–1.09
Missing vs 38–44	1.45	0.92–2.26				
Platelet count, K/uL						
≤149 vs 150–399	1.1	0.98–1.24	1.04	0.98–1.22	1.09	0.97–1.23
≥400 vs 150–399	1.13*	1.01*–1.28*	1.15	0.89–1.34	0.99	0.8–1.12
Missing vs 150-399	0.61	0.35–1.06				
PTT						
≥35 vs ≤34	1.13*	1.01*–1.28*	1.04	0.91–1.19	0.99	0.89–1.10
Missing vs ≤34	0.99	0.92–1.08				
>10% Weight loss						
Yes vs no	1.13*	1.03*–1.25*	1.13*	1.01*–1.29*	0.89	0.78–1.01
Preoperative chemo/radiation						
Yes vs no	1.13	0.99–1.28	1.08	0.92–1.28	0.82	0.75–0.91
C-index	0.62		0.61		0.6	
HL chi-square	4.12		6.2		6.68	
p Value	0.85		0.62		0.57	

*p Value <0.05.
ASA, American Society of Anesthesiologists; OR, odds ratio; PTT, partial thromboplastin time; SGOT, serum glutamic oxaloacetic transaminase.

decreased sample size and changes to the underlying study population.

Analyzing predictors of major complications after imputing values for those with missing data produced estimates that were similar to both complete case analysis and analysis with missing values as a category. Predominantly, imputation produced larger standard errors for estimators due to variation in the set of plausible values selected for imputation. Of note, patients without missing operative and demographic characteristics, whose data constituted the set from which selections were drawn, tended to be sicker, indicating that the data might not be missing at random (Table 1). As a result, these imputations might be selecting plausible values for patient characteristics from a nonrepresentative cohort, potentially biasing the underlying population used to create the model. However, 17 of 22 predictor variables across regression models did give the same or similar findings, and in only 5 of 22 predictor variables did regressions show dissimilar results (for creatinine, white blood count, hematocrit, PTT, and >10% weight loss). Although these results were based on a single imputation framework, multiple imputations would most likely increase the standard errors for the model. As a result, the similarities between methods for handling missing data could increase. Overall, each model was well-calibrated and had similar C-indices, regardless of the approach for handling missing data.

Examining each analysis technique for predicting mortality and an increased number of complications showed results similar to those identified from the major complication analysis. Estimates varied slightly in both the predicted effect and standard errors for each predictor of interest, with the greatest variation seen between the predicted effect of BUN ≥40 vs <39 in complete case analysis and categorized missing values (OR = 1.447 vs 1.05) (results not shown).

Intermediate-risk procedures

More than 65% of patients undergoing an intermediate-risk procedure had 1 or more pre-, intra-, or postoperative characteristics missing (Table 3). Excluding any patient with a missing value from analysis resulted in a loss of 17,191 patients for 9,140 usable cases (35% of total). Again, analyses using complete case analyses compared with categorization of missing values resulted in similarities in the estimated effect of most predictors, as well as their

Table 3. Predictors of Major Complications in Intermediate Risk Surgery

	Missing values included (n = 26,331)		Missing values excluded (n = 9,140)		Missing values imputed (n = 26,331)	
Predictor	**OR**	**95% CI**	**OR**	**95% CI**	**OR**	**95% CI**
Age, y						
51–60 vs <50	0.85	0.75–0.97	0.82	0.67–1.01	0.84*	0.74*–0.96*
61–70 vs <50	0.89	0.78–1.01	0.79*	0.65*–0.97*	0.86*	0.76*–0.97*
>70 vs <50	0.97	0.86–1.1	0.89	0.74–1.09	0.95	0.85–1.08
Race						
African American vs white	1.18*	1.06*–1.32*	1.18*	1.01*–1.39*	1.00	0.91–1.10
Hispanic vs white	1.02	0.86–1.21	0.99	0.77–1.27	0.99	0.85–1.16
Other vs white	0.82*	0.68*–0.99*	0.97	0.73–1.3	0.96	0.81–1.14
Missing vs white	0.92	0.81–1.05				
Sex, male vs female	1.35*	1.26*–1.45*	1.37*	1.23*–1.54*	1.31*	1.22*–1.41*
Smoking status						
Current vs never	1.26*	1.14*–1.38*	1.24*	1.07*–1.45*	1.26*	1.14*–1.38*
Previous vs never	1.04	0.95–1.14	0.98	0.85–1.13	0.99	0.91–1.09
ASA 3+ vs 1–2	1.47*	1.36*–1.59*	1.48*	1.29*–1.7*	1.55*	1.43*–1.68*
Cardiac conditions, yes vs no	1.14*	1.03*–1.25*	1.15*	1.01*–1.33*	1.16*	1.06*–1.27*
Hypertension, yes vs no	1.08*	1.01*–1.17*	1.15*	1.01*–1.3*	1.09*	1.01*–1.17*
Diabetes, yes vs no	1.04	0.96–1.14	1.06	0.93–1.22	1.07	0.98–1.17
Pulmonary conditions, yes vs no	1.46*	1.34*–1.59*	1.47*	1.29*–1.67*	1.53*	1.41*–1.67*
Sodium, mmol/L						
≤135 vs 136–144	1.25*	1.12*–1.38*	1.3*	1.12*–1.52*	1.22*	1.10*–1.35*
≥145 vs 136–144	1.24*	1.02*–1.5*	1.37*	1.03*–1.83*	1.14	0.95–1.38
Missing vs ≤136–144	0.93	0.73–1.19				
BUN						
≥40 vs ≤39	1.56*	1.23*–1.96*	1.46*	1.05*–2.03*	1.04	0.83–1.29
Missing vs ≤39	0.86	0.69–1.07				
Creatinine, mg/dL						
>1.6 vs ≤1.5	1.34*	1.16*–1.55*	1.41*	1.14*–1.74*	1.32*	1.15*–1.52*
Missing vs ≤1.5	1.42*	1.08*–1.87*				
Albumin, g/dL						
≤2.5 vs >2.6	1.68*	1.44*–1.95*	1.56*	1.29*–1.89*	1.20*	1.05*–1.36*
Missing vs >2.6	0.9	0.78–1.05				
Total bilirubin, mg/dL						
≥1.0 vs <1.0	1.12	0.99–1.25	1.19*	1.01*–1.39*	1.09	0.99–1.21
Missing vs <1.0	1	0.78–1.29				
SGOT						
≥40 vs ≤39	1.25*	1.09*–1.44*	1.33*	1.11*–1.6*	1.16*	1.03*–1.30*
Missing vs ≤39	1.28*	1.04*–1.58*				
Alkaline phosphatase, U/L						
≥125 vs ≤124	1.12	0.98–1.26	1.04	0.85–1.23	0.99	0.89–1.12
Missing vs ≤124	0.85	0.64–1.11				
White blood cell count, K/uL						
≤4.4 vs 4.5–10.9	0.99	0.88–1.12	0.86	0.71–1.06	0.91	0.80–1.02
≥11.0 vs 4.5–10.9	1.51*	1.36*–1.68*	1.73*	1.48*–2.02*	1.25*	1.13*–1.38*
Missing vs 4.5–0.9	0.82	0.51–1.33				

(continued)

Table 3. Continued

	Missing values included (n = 26,331)		Missing values excluded (n = 9,140)		Missing values imputed (n = 26,331)	
Predictor	**OR**	**95% CI**	**OR**	**95% CI**	**OR**	**95% CI**
Hematocrit, %						
≤37 vs 38–44	1.15*	1.07*–1.25*	1.2*	1.06*–1.37*	1.01	0.94–1.10
≥45 vs 38–44	0.99	0.86–1.17	0.8	0.59–1.09	0.92	0.78–1.09
Missing vs 38–44	1.55*	1.07*–2.25*				
Platelet count, K/uL						
≤149 vs 150–399	1.13	0.98–1.31	1.28*	1.04*–1.58*	1.29*	1.13*–1.47*
≥400 vs 150–399	1.09	0.98–1.21	1.05	0.89–1.24	1.02	0.92–1.13
Missing vs 150–399	0.96	0.6–1.53				
PTT						
≥35 vs ≤34	1.27*	1.11*–1.46*	1.21*	1.04*–1.42*	1.13*	1.02*–1.25*
Missing vs ≤34	0.89*	0.82*–0.95*				
>10% Weight loss, yes vs no	1.23*	1.08*–1.39*	1.16	0.97–1.38	0.69*	0.61*–0.78*
Preoperative chemo/radiation, yes vs no	1.3*	1.15*–1.47*	1.16	0.95–1.43	0.82*	0.61*–0.78*
C-index	0.66		0.67		0.64	
HL chi-square	7.65		13.58		8.52	
p Value	0.46		0.093		0.38	

*p Value <0.05.
ASA, American Society of Anesthesiologists; OR, odds ratio; PTT, partial thromboplastin time; SGOT, serum glutamic oxaloacetic transaminase.

associated standard errors. The estimated effects of BUN presented the largest variation across all 3 analytic methods, with an OR of 1.56 ($p < 0.05$) for categorized missing values, OR of 1.46 ($p < 0.05$) for complete case analysis, and OR of 1.04 ($p > 0.05$) for imputed values. OR estimates were again similar across all 3 models for 20 of the 22 predictor variables and different for only 2 variables (>10% weight loss and preoperative chemo/radiation), with 95% CIs overlapping for most predictors.

The C-indices of the 3 models were similar. Calibration for all models was acceptable, although the calibration for the model based on complete cases tended to be somewhat inferior to the other 2 models (not significant). Analysis techniques predicting mortality and an increased number of complications again presented similar estimated effects for each predictor of interest, as well as model fit across methods for handling missing data.

Low-risk procedures

Low-risk procedures presented the most substantial challenges for estimation of predicted effects because of the very high proportion of missing values (Table 4). More than 44,845 of 51,262 patients (87%) had at least one pre-, intra-, or postoperative characteristic missing. As a result, complete case analysis would produce the estimated effects of patient characteristics based on only 13% of the population. Estimated effects were similar for 19 of 22 predictor variables, and were different for only 3 predictor variables (creatinine, alkaline phosphatase, and PTT). The C-index was slightly higher for the model based on complete cases compared with the other 2 models, but the calibration was somewhat worse.

Spearman rank correlations of the predicted probabilities of major complications

Table 5 presents the Spearman rank correlations of the predicted probabilities of major complications from the 3 models. For the high-risk procedures, the correlations ranged from 0.84 to 0.95, indicating that all 3 methods for handling missing values tended to produce patient probabilities for major complications that ranked the patients very similarly. However, the correlations between the models were smaller for the intermediate- and low-risk procedures, indicating that the choice of method for handling missing values might be more important for these procedures compared with high-risk procedures.

We performed several sensitivity analyses for all multivariate models to ensure that our results were not due to our modeling decisions. Placing a large number of potential risk factors into a multivariate model presents a risk for overadjusting for these factors and creating models with few to no significant results. To evaluate this potential concern, we performed our multivariate analyses with groups of 2 to 3 laboratory values (which have high levels of missing data) to see if their predictive ability changed in the absence of a broader array of laboratory data. Our results remained unchanged.

Table 4. Predictors of Major Complications in Low-Risk Surgery

	Missing values included (n = 51,262)		Missing values excluded (n = 6,417)		Missing values imputed (n = 51,262)	
Predictor	**OR**	**95% CI**	**OR**	**95% CI**	**OR**	**95% CI**
Age, y						
51–60 vs <50	1.071	0.985–1.163	0.93	0.76–1.214	1.09*	1.01*–1.19*
61–70 vs <50	1.006	0.917–1.103	0.838	0.644–1.09	1.04	0.95–1.13
>70 vs <50	0.799*	0.719*–0.888*	0.802	0.605–1.062	0.82*	0.74*–0.91*
Race						
African American vs white	1.005	0.902–1.119	0.927	0.721–1.192	0.97	0.88–1.06
Hispanic vs white	0.962	0.823–1.12	0.894	0.658–1.215	1.01	0.91–1.13
Other vs white	0.865	0.732–1.023	0.765	0.476–1.229	0.94	0.80–1.09
Missing vs white	0.999	0.889–1.123				
Sex, male vs female	0.609*	0.515*–0.721*	0.593*	0.373*–0.942*	0.57*	0.48*–0.67*
Smoking status						
Current vs never	1.13*	1.034*–1.235*	1.437*	1.138*–1.814*	1.80*	1.52*–2.13*
Previous vs never	1.035	0.939–1.141	1.023	0.787–1.33	1.11	0.85–1.45
ASA, 3+ vs 1–2	1.135*	1.05*–1.227*	1.198	0.981–1.463	1.15*	1.06*–1.25*
Cardiac conditions, yes vs no	1.122	0.956–1.317	0.902	0.608–1.338	1.13	0.97–1.33
Hypertension, yes vs no	1.069	0.992–1.152	1.144	0.933–1.404	1.07	0.99–1.15
Diabetes, yes vs no	0.958	0.86–1.067	0.913	0.697–1.196	0.96	0.87–1.07
Pulmonary conditions, yes vs no	1.035	0.926–1.156	1.092	0.833–1.43	1.03	0.93–1.16
Sodium, mmol/L						
≤135 vs 136–144	1.053	0.898–1.234	1.389	0.976–1.976	1.08	0.94–1.24
≥145 vs 136–144	1.017	0.805–1.285	1.271	0.756–2.137	1.20	0.99–1.44
Missing vs ≤136–144	1.063	0.895–1.285				
BUN						
≥40 vs ≤39	1.305	0.898–1.898	0.537	0.211–1.365	1.05	0.74–1.49
Missing vs ≤39	0.72*	0.594*–0.872*				
Creatinine, mg/dL						
>1.6 vs ≤1.5	1.104	0.841–1.448	1.953*	1.162*–3.28*	1.14	0.89–1.46
Missing vs ≤1.5	1.351*	1.126*–1.623*				
Albumin, g/dL						
≤2.5 vs >2.6	1.607	0.966–2.671	1.408	0.716–2.768	1.48	1.07–2.05
Missing vs >2.6	1.052	0.895–1.237				
Total bilirubin, mg/dL						
≥1.0 vs <1.0	1.092	0.935–1.275	1.325	0.986–1.78	1.03	0.92–1.16
Missing vs <1.0	1.112	0.887–1.393				
SGOT						
≥40 vs ≤39	1.057	0.894–1.251	1.312	0.96–1.793	1.07	0.94–1.21
Missing vs ≤39	0.772*	0.622*–0.958*				
Alkaline phosphatase, U/L						
≥125 vs ≤124	1.268*	1.074*–1.498*	1.136	0.82–1.573	0.95	0.83–1.08
Missing vs ≤124	0.974	0.756–1.256				
White blood cell count, K/uL						
≤4.4 vs 4.5–10.9	0.961	0.846–1.092	1.016	0.755–1.368	0.98	0.88–1.08
≥11.0 vs 4.5–10.9	0.98	0.835–1.15	1.049	0.726–1.514	0.89	0.77–1.04
Missing vs 4.5–10.9	1.342	0.915–1.968				

(continued)

Table 4. Continued

Predictor	Missing values included (n = 51,262) OR	95% CI	Missing values excluded (n = 6,417) OR	95% CI	Missing values imputed (n = 51,262) OR	95% CI
Hematocrit, %						
≤37 vs 38–44	0.977	0.904–1.057	1.04	0.859–1.26	0.98	0.91–1.05
≥45 vs 38–44	1.149	0.991–1.332	1.116	0.736–1.69	1.03	0.88–1.19
Missing vs 38–44	0.867	0.716–1.049				
Platelet count, K/uL						
≤149 vs 150–399	1.05	0.855–1.29	0.615	0.373–1.013	1.08	0.94–1.26
≥400 vs 150–399	0.977	0.822–1.162	1.361	0.954–1.942	1.18*	1.00*–1.38*
Missing vs 150-399	0.936	0.639–1.369				
PTT						
≥35 vs ≤34	1.718*	1.386*–2.129*	1.695*	1.286*–2.235*	1.09	0.97–1.22
Missing vs ≤34	1.039	0.956–1.129				
>10% Weight loss, yes vs no	0.995	0.66–1.5	1.514	0.733–3.124	1.12	0.95–1.33
Preoperative chemo/radiation, yes vs no	0.867	0.728–1.033	1.139	0.791–1.638	0.95	0.63–1.43
C-index	0.57		0.61		0.55	
HL chi-square	7.34		14.8		7.01	
p Value	0.5		0.064		0.53	

*p Value <0.05.
ASA, American Society of Anesthesiologists; OR, odds ratio; PTT, partial thromboplastin time; SGOT, serum glutamic oxaloacetic transaminase.

DISCUSSION

In our study, we demonstrate that although predictors of short-term operative outcomes after cancer surgery in the ACS NSQIP database remain similar across methods for handling missing data, certain approaches should be favored due to the ability to evaluate all eligible patients. The ACS NSQIP is an important surgical outcomes database, providing a wide variety of pre-, intra-, and postoperative patient characteristics for use in studying a diverse set of surgical procedures and risk-adjusted short-term operative outcomes. However, investigators should be aware that results for coefficients on certain variables can change in magnitude depending on the approach to handling missing data, particularly for models evaluating the impact of preoperative laboratory values in low-risk procedures. Additionally, model discrimination can vary by method, as highlighted in lower-risk procedures (Table 4). To our knowledge, the present analysis is one of the first studies to systematically evaluate the impact of missing data on multiple predictors of operative outcomes after cancer surgery in the ACS NSQIP database.

Table 5. Spearman Rank Correlation: Predicted Probability of Developing Major Complications by Method for Handling Missing Data (in Quintiles)

Procedures	Missing values as a category	Complete case	Imputed
Low-risk			
Missing values as a category	1.00	0.62	0.64
Complete case	0.62	1.00	0.51
Imputed	0.64	0.51	1.00
Intermediate risk			
Missing values as a category	1.00	0.93	0.81
Complete case	0.93	1.00	0.90
Imputed	0.81	0.90	1.00
High-risk			
Missing values as a category	1.00	0.95	0.84
Complete case	0.95	1.00	0.95
Imputed	0.84	0.93	1.00

Our results are consistent with previous studies identifying the potential for missing data to influence underlying study populations and potentially interpretations of risk factors in ACS NSQIP. Hamilton and colleagues have examined surgically treated patients captured within ACS NSQIP during the 2006 calendar year to evaluate interpretations of albumin under different assumptions about the missing data.[7] Their analysis demonstrated that among ACS NSQIP patients, >45% had missing values for albumin levels. In addition, the authors found that albumin's effect on operative morbidity and mortality is altered depending on the method for addressing missing values. They conclude that a bias exists in patients who actually have their albumin levels reported and that "missingness" was significantly associated with morbidity and mortality outcomes. In effect, excluding or imputing values for these individuals can inappropriately bias albumin levels for

these individuals, as they were considerably healthier than their counterparts with the laboratory value recorded.

We demonstrate that across the spectrum of preoperative laboratory values, the proportion of missing data was higher among those undergoing low-risk procedures or those who also tended to be healthier. In addition, we also found that multivariable regressions of low-risk procedures are subject to exclusion of a large number of patients if only individuals with complete case information were included in these multivariate models. For these reasons, we do not recommend using complete case analysis for developing prediction models using NSQIP data. In addition, this study identifies the importance of understanding the assumptions behind the methods investigators choose to handle missing data. Specifically, the observation that a set of predictors are no longer significant after hotdeck imputation, particularly among low-risk procedures, might bring into question the validity of the missing at random assumption required for imputation of data. In other words, patients with missing values for laboratory or other information can be qualitatively different than those who have complete information. As a result, investigators should cautiously use imputation methods when analyzing ACS NSQIP because of the potential for inappropriately extrapolating results based on a sicker overall population.

Missing data can present challenges to interpretations of both the effect of laboratory values and comorbidities on short-term operative outcomes. These missing values can result from a wide variety of factors, including patient refusal to disclose information (eg, race), physicians requesting specific laboratory tests only, incomplete data captured by the trained ACS NSQIP data collector, and the potential lack of coordination between laboratory reporting and inclusion in a patient's medical record. As previously noted, high-risk procedures tend to have more complete laboratory values and fewer overall missing values for the procedure of interest. On the other hand, low-risk procedures present challenges for health outcomes investigators interested in the predictive ability of pre-, intra-, and postoperative variables because of the potential for losing the majority of cases in a complete case analysis. Although estimated effects between most predictors remained stable among all 3 methods for handling missing data, the ability to generate meaningful and practical conclusions based on imputed values or complete case analysis might still present challenges to the validity of such results due to the large proportion of missing data.

Although our study provides an addition to current knowledge of appropriate analysis in the presence of missing data within ACS NSQIP, we acknowledge several limitations. First, our results did not assess procedure-specific complications. We have previously demonstrated that ACS NSQIP has low predictive ability for postoperative complications for specific procedures compared with operative mortality after major cancer surgery.[20] ACS NSQIP is considering adding disease- and operation-specific factors in its future format, such as anastomic leaks after colectomies. Second, given our choice to only include people with cancer, the current findings might not be applicable to the remainder of NSQIP cohort. Third, the present study did not adjust for surgeon or hospital volume, which might correlate with mortality and morbidity, because ACS NSQIP does not release data on the treating facility to encourage participation and reporting of adverse events. As a result, it was not possible to compare the different models on how they rank performance of providers or institutions. Rather, we relied on the model's comparison of predicting individual complications between methods for handling missing data. Our findings are based on an important tool that assesses surgical quality from >250 participating hospitals, which can be superior to studies based on single-center databases.

Findings from the current analysis present several important implications for future studies using ACS NSQIP to identify risk factors for complications after cancer surgery at its participating hospitals. First, surgical outcomes investigators should examine patterns of missingness in their data and fully consider the strengths and benefits of, at a minimum, the 3 techniques outlined in this article. Weighing the pros and cons of each technique with respect to usable sample, interpretability of results (ie, what does a "missing" category mean?), and impact on variance estimation, will allow researchers to better interpret findings. Second, future studies should conduct and report a complete sensitivity analysis using multiple multivariable techniques to ensure that results are not an artifact of sample selection or exclusion of patients with only complete values for all predictors of interest. Investigators should be aware of the potential changes to analyses due to the presence of missing data within the ACS NSQIP. Importantly, future studies should carefully consider the potential for bias within analyses under different modeling techniques and approaches to handling missing data and be transparent about the chosen technique in their methods.

CONCLUSIONS

Missing data presents challenges to interpreting predictors of short-term operative outcomes after cancer surgery at ACS NSQIP hospitals. Similar to best practices for other data sets, this study highlights the importance of using missing values carefully when using ACS NSQIP. Given its potential to introduce bias, the approach to handling missing values should be detailed in future ACS NSQIP studies.

Author Contributions

Study conception and design: Parsons, Davern, Al-Refaie
Acquisition of data: Parsons, Al-Refaie
Analysis and interpretation of data: Parsons, Henderson, Ziegenfuss, Davern, Al-Refaie
Drafting of manuscript: Parsons, Henderson, Ziegenfuss, Davern, Al-Refaie
Critical revision: Parsons, Henderson, Ziegenfuss, Davern, Al-Refaie

REFERENCES

1. Henderson WG, Daley J. Design and statistical methodology of the National Surgical Quality Improvement Program: why is it what it is? Am J Surg 2009;198[Suppl]:S19–S27.
2. Hall BL, Hamilton BH, Richards K, et al. Does surgical quality improve in the American College of Surgeons National Surgical Quality Improvement Program: an evaluation of all participating hospitals. Ann Surg 2009;250:363–376.
3. Birkmeyer JD, Shahian DM, Dimick JB, et al. Blueprint for a new American College of Surgeons National Surgical Quality Improvement Program. J Am Coll Surg 2008;207:777–782.
4. Itani KM. Fifteen years of the National Surgical Quality Improvement Program in review. Am J Surg 2009;198[Suppl]:S9–S18.
5. Khuri SF, Henderson WG, Daley J, et al. Successful implementation of the department of Veterans Affairs' National Surgical Quality Improvement Program in the private sector: the Patient Safety in Surgery Study. Ann Surg 2008;248:329–336.
6. American College of Surgeon's National Quality Improvement Program. Available at: http://www.acsnsqip.org/. Cited July 10, 2010. Accessed July 8, 2010.
7. Hamilton BH, Ko CY, Richards K, Hall BL. Missing data in the American College of Surgeons National Surgical Quality Improvement Program are not missing at random: implications and potential impact on quality assessments. J Am Coll Surg 2010;210:125–139.
8. Al-Refaie WB, Parsons HM, Henderson WG, et al. Major cancer surgery in the elderly: results from the American College of Surgeons National Surgical Quality Improvement Program. Ann Surg 2010;251:311–318.
9. Al-Refaie WB, Parsons HM, Henderson WG, et al. Body mass index and major cancer surgery: lack of association or need for alternative measures of obesity? Ann Surg Oncol 2010;17:2264–2273.
10. Rubin DB. Inference and missing data. Biometrika 1976;63:581–592.
11. Rubin DB. Formalizing subjective notions about the effect of non-respondents in sample surveys. J Am Stat Assoc 1977;72:538–543.
12. Little RJ, Rubin DB. The analysis of social science data with missing values. Sociol Methods Res 1989;18:292–326.
13. Graham JW, Schafer JL. Missing data: our view of the state of the art. Psychol Methods 2002;7:147–177.
14. Laird NM. Missing data in longitudinal studies. Stat Med 1988;7:305–315.
15. Ibrahim JG. Incomplete data in generalized linear models. J Am Stat Assoc 1990;85:765–769.
16. Robbins JM, Rotnitzky A, Zhao LP. Estimation of regression coefficients when some regressors are not always observed. J Am Stat Assoc 1994;89:846–866.
17. Little RJ. Missing-data adjustments in large surveys. J Bus Econ Stat 1988;6:287–296.
18. Horton NJ, Lipsitz SR. Multiple imputation in practice: comparison of software packages for regression models with missing variables. Am Stat 2001;55:244–254.
19. Rubin DR. Multiple imputation after 18+ years. J Am Stat Assoc 1996;91(434):473–489.
20. Borja-Cacho D, Parsons HM, Habermann EB, et al. Assessment of ACS NSQIP's predictive ability for adverse events after major cancer surgery. Ann Surg Oncol 2010;17:2274–2282.