

Results of Database Studies in Spine Surgery Can Be Influenced by Missing Data

Bryce A. Basques MD, Ryan P. McLynn BS, Michael P. Fice BS, Andre M. Samuel MD, Adam M. Lukasiewicz MD, MSc, Daniel D. Bohl MD, MPH, Junyoung Ahn MD, Kern Singh MD, Jonathan N. Grauer MD

Published online: 28 November 2016
© The Association of Bone and Joint Surgeons® 2016

Abstract

Background National databases are increasingly being used for research in spine surgery; however, one limitation of such databases that has received sparse mention is the frequency of missing data. Studies using these databases often do not emphasize the percentage of missing data for each variable used and do not specify how patients with missing data are incorporated into analyses. This study uses the American College of Surgeons National Surgical Quality Improvement Program (ACS-NSQIP) database to examine whether different treatments of missing data can influence the results of spine studies.

One of the authors certifies that he (KS), or a member of his immediate family, has or may receive payments or benefits, during the study period: an amount less than USD 10,000 from Zimmer (Warsaw, IN, USA), an amount less than USD 10,000 from Stryker (Kalamazoo, MI, USA), an amount less than USD 10,000 from Pioneer Surgical Technology (Marquette, MI, USA), an amount less than USD 10,000 from Thieme (St Charles, IL, USA), an amount less than USD 10,000 from Japryee Publishing (London, UK), an amount less than USD 10,000 from Slack Inc (Thorofare, NJ, USA), an amount less than USD 10,000 from DePuy Synthes (Raynham, MA, USA), an amount less than USD 10,000 from Vital 5 LLC (Logan, UT, USA), an amount less than USD 10,000 from TruVue Surgical (Chicago, IL, USA), an amount less than USD 10,000 from Avaz Surgical (Chicago, IL, USA), an amount less than USD 10,000 from Bijali Surgical (Chicago, IL, USA), an amount less than USD 10,000 from the Cervical Spine Research Society, and an amount less than USD 10,000 from Lippincott Williams & Wilkins (Philadelphia, PA, USA), all outside of the submitted work. One of the authors certifies that he (JNG), or a member of his immediate family, has or may receive payments or benefits, during the study period: an amount less than USD 10,000 from Bioventus (Durham, NC, USA), an amount less than USD 10,000 from ISTO Technologies (St Louis, MO, USA), an amount less than USD 10,000 from Medtronic (Minneapolis, MN, USA), an amount less than USD 10,000 from Stryker (Kalamazoo, MI, USA), an amount less than USD 10,000 from Novella Clinical (Morrisville, NC, USA), an amount less than USD 10,000 from

Questions/purposes (1) What is the frequency of missing data fields for demographics, medical comorbidities, pre-operative laboratory values, operating room times, and length of stay recorded in ACS-NSQIP? (2) Using three common approaches to handling missing data, how frequently do those approaches agree in terms of finding particular variables to be associated with adverse events? (3) Do different approaches to handling missing data influence the outcomes and effect sizes of an analysis testing for an association with these variables with occurrence of adverse events?

Andante Medical Devices (White Plains, NY, USA), and an amount less than USD 10,000 from Vertex Pharmaceuticals (Boston, MA, USA); ongoing expert testimony with legal case reviews; and an amount less than USD 10,000 from the Orthopaedic Trauma Association (Rosemont, IL, USA), all outside the submitted work. All ICMJE Conflict of Interest Forms for authors and *Clinical Orthopaedics and Related Research*® editors and board members are on file with the publication and can be viewed on request. Each author certifies that all investigations were conducted in conformity with ethical principles of research. This work was performed at Yale School of Medicine, New Haven, CT, USA.

B. A. Basques, M. P. Fice, D. D. Bohl, J. Ahn, K. Singh
Department of Orthopaedic Surgery, Rush University Medical Center, Chicago, IL, USA

R. P. McLynn, A. M. Lukasiewicz, J. N. Grauer (✉)
Department of Orthopaedics and Rehabilitation, Yale School of Medicine, 800 Howard Avenue, New Haven, CT 06510, USA
e-mail: jonathan.grauer@yale.edu

A. M. Samuel
Department of Orthopedic Surgery, Hospital for Special Surgery, New York, NY, USA

Methods Patients who underwent spine surgery between 2005 and 2013 were identified from the ACS-NSQIP database. A total of 88,471 patients undergoing spine surgery were identified. The most common procedures were anterior cervical discectomy and fusion, lumbar decompression, and lumbar fusion. Demographics, comorbidities, and perioperative laboratory values were tabulated for each patient, and the percent of missing data was noted for each variable. These variables were tested for an association with “any adverse event” using three separate multivariate regressions that used the most common treatments for missing data. In the first regression, patients with any missing data were excluded. In the second regression, missing data were treated as a negative or “reference” value; for continuous variables, the mean of each variable’s reference range was computed and imputed. In the third regression, any variables with > 10% rate of missing data were removed from the regression; among variables with ≤ 10% missing data, individual cases with missing values were excluded. The results of these regressions were compared to determine how the different treatments of missing data could affect the results of spine studies using the ACS-NSQIP database.

Results Of the 88,471 patients, as many as 4441 (5%) had missing elements among demographic data, 69,184 (72%) among comorbidities, 70,892 (80%) among preoperative laboratory values, and 56,551 (64%) among operating room times. Considering the three different treatments of missing data, we found different risk factors for adverse events. Of 44 risk factors found to be associated with adverse events in any analysis, only 15 (34%) of these risk factors were common among the three regressions. The second treatment of missing data (assuming “normal” value) found the most risk factors (40) to be associated with any adverse event, whereas the first treatment (deleting patients with missing data) found the fewest associations at 20. Among the risk factors associated with any adverse event, the 10 with the greatest effect size (odds ratio) by each regression were ranked. Of the 15 variables in the top 10 for any regression, six of these were common among all three lists.

Conclusions Differing treatments of missing data can influence the results of spine studies using the ACS-NSQIP. The current study highlights the importance of considering how such missing data are handled.

Clinical Relevance Until there are better guidelines on the best approaches to handle missing data, investigators should report how missing data were handled to increase the quality and transparency of orthopaedic database research. Readers of large database studies should note whether handling of missing data was addressed and consider potential bias with high rates or unspecified or weak methods for handling missing data.

Introduction

The evolution of large national databases has led to a dramatic increase in the number of studies about spine surgery that derive from these databases [9, 12, 22, 23, 28, 36, 37]. Using databases such as the American College of Surgeons National Surgical Quality Improvement Program (ACS-NSQIP), these studies aim to draw conclusions about the demographics of those undergoing specific procedures, surgical characteristics, and perioperative outcomes. Although these studies are strengthened by large sample sizes, critics of these studies highlight issues with data quality [18], patient selection [11], and database structure as potentially influencing conclusions [11, 13].

One area of concern regarding data quality is the handling of missing data. Studies using these databases often do not mention the percent of missing data for each variable used, and they generally fail to note how missing data are handled in the subsequent analyses [38]. Common approaches to missing data include excluding those patients with any missing data elements [4, 34], assuming that a missing element implies a negative or “reference” value [15, 31], or only including variables with low amounts of missing data. Potentially furthering the effects of missing data, ACS-NSQIP has changed several elements of its data collection across different years. Many variables concerning comorbidities, laboratory values, and operative times that were required in early editions of ACS-NSQIP are no longer collected, whereas others have been added, so an analysis that includes many database years may have missing values for those affected variables. Additionally, the number of participating centers has increased from 121 in 2005 to 2006 to 435 in 2013 [40, 41]. As ACS-NSQIP grew, it now allowed some of the new, smaller participating institutions to collect a reduced subset of variables for several years until these variables were phased out entirely in 2013 [1]. This also can create missing data in the variables that many centers were excused from collecting. This variation in the number of variables collected by each participating institution is not explicitly outlined in the ACS-NSQIP user guide [40]. Many statistical packages default to dropping patients with missing data in a regression analysis. Therefore, if investigators are not aware that variables may have large proportions of missing data, their analysis may include only a small portion of intended patients, which could adversely affect the study’s generalizability and power.

In considering how to handle missing data, one should also consider the amount of missing data and the mechanism of missing data. The missing data mechanism can be described as missing completely at random, missing at random, or not missing at random [26]. Missing completely

at random signifies that the distribution of missing data is not related to any variables collected or missing. In data missing at random, the missingness may be related to variables with known values but not any missing data. For data not missing at random, the distribution of missing data is related to other missing values. In terms of the amount of missing data, data sets with small proportions of missing data (ie, < 5%–10%) are less susceptible to bias when the data set is manipulated, whereas larger proportions of missing data can lead to greater bias, resulting in different results when different approaches to handling the missing data are used [16]. With small fractions of missing data and/or data missing completely at random, it may be acceptable to drop cases with any missing data, known as complete-case analysis [26]. Otherwise, complete-case analysis risks an analysis with decreased power and limited generalizability to the study population. In these cases, there are many approaches such as weighting, imputation, or maximum likelihood estimation to replace missing values with the optimal selection varying with sample size, proportion of missing data, and the type and pattern of missing data.

Although the limitations of database studies have received increasing attention [11, 13, 17, 18], the issue of how to treat database patients with missing data remains incompletely explored in orthopaedic research. To our knowledge, there are no orthopaedic studies on the effects of missing data in large databases and few in the literature overall. In a study comparing the effects of different statistical analyses on a clinical trial examining operative fixation of clavicle fractures, Neuhaus and Ring [31] found that different methods of handling missing data did not change the overall trend of findings but changed odds ratios and numbers needed to treat by up to 17%, which were statistically significant in many cases. Similarly, Biau and colleagues [10] analyzed the effect of missing data among their primary and secondary outcomes in a meta-analysis of patellar tendon versus hamstring tendon autograft in anterior cruciate ligament repairs. They found that there was no change in the overall direction of findings as a result of missing data, but computed odds ratios varied by 60% for primary outcome and 13% for secondary outcome. Of note, both studies experienced low rates of missing data (12% for Neuhaus and Ring, 5% for Biau et al.), so studies with larger shares of missing data may be more markedly affected by different handling methods.

With changes in data collection practices across different years and no clear explanation of which centers collect each data element, how investigators decide to include or exclude patients with missing data could influence study results. Because ACS-NSQIP does not explain whether certain patients or centers have high rates of missing data, one cannot reliably determine the mechanism of missing

data. Therefore, attempts to impute missing data may cause studies to be poorly applicable to those populations. Missing data may increase the likelihood of making type I errors, because a metaepidemiologic study of pain relief in patients with osteoarthritis found that studies with more excluded cases and missing data found greater effect sizes and lower p values than studies without missing cases and with higher quality data [32]. Additionally, if investigators do not expound on their handling of missing data, then attempts to reproduce, evaluate, or compare with these studies may be of little value if a different handling method is unknowingly chosen.

Therefore, we asked: (1) What is the frequency of missing data fields for demographics, medical comorbidities, preoperative laboratory values, operating room times, and length of stay recorded in ACS-NSQIP? (2) Using three common approaches to handling missing data, how frequently do those approaches agree in terms of finding particular variables to be associated with adverse events? (3) Do different approaches to handling missing data influence the outcomes and effect sizes of an analysis testing for an association with these variables with occurrence of adverse events?

Materials and Methods

A retrospective study was conducted using the ACS-NSQIP from 2005 to 2013. Patients who underwent spine surgery between 2005 and 2013 were identified from the ACS-NSQIP database. Spine procedures were selected based on the following primary Current Procedural Terminology codes: anterior cervical discectomy and fusion (22551, 22554, or 63075), anterior cervical corpectomy (63081 or 63085), posterior cervical fusion (22600), cervical laminotomy (63020 or 63040), cervical laminectomy (63015, 63045, or 63265), posterior thoracic fusion (22610), thoracic laminectomy (63046 or 63266), lumbar laminotomy (63030 or 63042), lumbar laminectomy (63047, 63005, 63012, or 63267), anterior lumbar fusion (22558), and posterior lumbar fusion (22612, 22630, or 22633).

Demographics and comorbidities of ACS-NSQIP patients that were assessed in this study included: age, sex, body mass index (calculated from weight and height), race (Native American, Asian, black, Pacific Islander, or white), ethnicity (Hispanic or non-Hispanic), functional status, diabetes, current smoking, smoking history (in pack-years), alcohol use, current pneumonia, ascites, esophageal varices, congestive heart failure, history of myocardial infarction, previous percutaneous coronary intervention, previous cardiac surgery, angina, hypertension, peripheral vascular disease, rest pain, renal failure, dialysis, impaired

sensorium, coma > 24 hours, hemiplegia, history of transient ischemic attack, stroke with neurologic deficit, stroke without neurologic deficit, central nervous system tumor, paraplegia, quadriplegia, disseminated cancer, wound infection, steroid use, weight loss, bleeding disorder, preoperative blood transfusion, chemotherapy, radiotherapy, pregnancy, and a prior operation within 30 days.

Preoperative laboratory values that were assessed included: sodium, blood urea nitrogen, creatinine, albumin, bilirubin, aspartate transaminase, alkaline phosphatase, white blood cell count, hematocrit, platelets, partial thromboplastin time, international normalized ratio, and prothrombin time.

Operating room times were defined as follows: preoperative room time (time from the patient arriving in the room until the opening incision), operative time (time from the opening incision to the end of wound closure), and postoperative room time (time from the end of wound closure until the patient leaves the room). Postoperative length of stay was also available in the database.

The ACS-NSQIP records the occurrence of postoperative adverse events up to 30 days. For this study, “any adverse event” was defined as the occurrence of any of the following individual adverse events: death, coma > 24 hours, on a ventilator > 48 hours, unplanned intubation, stroke/cerebrovascular accident, thromboembolic event (deep venous thrombosis or pulmonary embolism), surgical site infection (superficial surgical site infection [SSI], deep SSI, organ/space infection), sepsis/septic shock, cardiac arrest, myocardial infarction, renal failure, return to the operating room, wound dehiscence, graft/prosthesis/flap failure, blood transfusion, urinary tract infection, pneumonia, renal insufficiency, or a peripheral nerve injury.

In the first research question, the primary outcome was the proportion of patients who had missing data for each of the variables of interest. For the second question, the primary outcome was the occurrence of “any adverse event,” a common outcome in large database studies [6–8, 14, 25, 27, 35], and the agreement among the three regressions in terms of which variables are associated with any adverse event. For the third question, primary outcomes were the odds ratios found in each regression, the difference among the odds ratios for each variable, and the associated risk factors with the greatest effect size for each regression.

Statistical Analysis

Patient demographics, comorbidities, preoperative laboratory values, and the type of procedure were tested for association with “any adverse event” using three separate

multivariate logistic regressions that used the three most common treatments for missing data. The models included all independent variables, including demographics, medical comorbidities, preoperative laboratory values, and operative times. Variables with p values < 0.05 were considered to be statistically significant.

In the first regression, patients with any missing data were simply excluded and complete-case analysis was performed. In the second regression, missing data were treated as a negative or “reference” value. Age, body mass index, and American Society of Anesthesiologists class were converted into categorical variables, and the lowest category was imputed. Operative time and total time in the operating room were entered as the means among patients with known values. For laboratory values, the mean of each variable’s reference range was computed and entered [43]. In the third regression, only variables with missing data rates of $\leq 10\%$ were incorporated into the regression, which included 41 variables. For variables with missing data in $\leq 10\%$ of patients, patients with missing values were excluded. The proportion of variables for which all regression methods produced the same outcome was calculated, and odds ratios were compared in terms of percent difference and 95% confidence intervals for each variable to evaluate how the different treatments of missing data could affect the results of spine surgery studies using the ACS-NSQIP database.

A total of 88,471 patients undergoing spine surgery were identified. The average age was 57 ± 14 years (mean \pm SD). The most common procedure types were lumbar laminotomy (28%), anterior cervical discectomy and fusion (21%), and lumbar laminectomy (20%), which together accounted for more than two-thirds of cases (Table 1).

Table 1. Percentage of patients undergoing each procedure type

Procedure	Number of patients	Percent of patients
Anterior cervical discectomy and fusion	18,464	21
Anterior lumbar fusion	3324	4
Cervical laminectomy	2782	3
Cervical laminotomy	1701	2
Cervical corpectomy	1828	2
Lumbar laminectomy	17,545	20
Lumbar laminotomy	24,359	28
Posterior lumbar fusion	15,552	18
Posterior cervical fusion	1535	2
Thoracic fusion	712	1
Thoracic laminectomy	669	1

Total number of patients = 88,471.

Results

Of the 88,471 patients, as many as 4441 (5%) had missing elements from demographic data, 69,184 (72%) from comorbidities, 70,892 (80%) from preoperative laboratory values, and 56,551 (64%) from operating room times. In particular, for demographics (Table 2), there was no missing data for age, but it was highest for race (4441 [5%]) and ethnicity (3194 [4%]). For patient comorbidity, there were no missing data for diabetes, current smoker, ascites, congestive heart failure, hypertension, renal failure, dialysis, disseminated cancer, wound infection, steroid use, weight loss, bleeding disorder, or preoperative blood transfusion, whereas status of former smoker had the highest rate (63,540 [72%]) (Table 3). It is notable that the rate of missing data was 66% ($n = 57,984$) for each of the following comorbidities: alcohol use, pneumonia, esophageal varices, history of myocardial infarction, previous percutaneous coronary intervention, previous cardiac surgery, angina, peripheral vascular disease, rest pain, impaired sensorium, coma, hemiplegia, history of transient ischemic attack, stroke with/without neurologic deficit, central nervous system tumor, quadriplegia, chemotherapy, and radiotherapy. This was attributed to certain institutions collecting only select variables, and this issue is covered further in the discussion. For preoperative laboratory values (Table 4), the variable with the lowest rate of missing data was hematocrit (7830 [9%]), and the highest rate was for prothrombin time (70,892 [80%]). For operating room times and length of stay (Table 5), operative time had the fewest missing data points (18 [0%]), whereas postoperative room time had the most (56,551 [64%]).

Adverse events occurred in 10,183 patients (12%) with blood transfusion, return to the operating room, and urinary tract infection being the most common events (Table 6). Among the three regressions performed, 44 risk factors were found to be associated with adverse events in at least one analysis, yet only 15 (34%) of these risk factors were common among all three regressions (Table 7). In the first regression, patients with any missing data were excluded,

Table 2. Missing demographics data

Demographic variable	Number of missing values	Percent missing
Age	0	0
Sex	62	0
Body mass index	690	1
Race	4441	5
Ethnicity	3194	4
Functional status	540	1

Total number of patients = 88,471.

leaving 5788 (out of an initial 88,471 patients). Twenty risk factors (29% of possible variables) were found to be associated with “any adverse event.” In the second regression, missing data were treated as the negative or “reference” value. All patients were included in this analysis, and 40 risk factors were identified (59% of possible variables). In the third regression, only variables with

Table 3. Missing comorbidity data

Comorbidity variable	Number of missing values	Percent missing
Diabetes	0	0
Current smoker	0	0
Former smoker	63,540	72
Alcohol	57,984	66
Current pneumonia	57,984	66
Ascites	0	0
Esophageal varices	57,984	66
CHF	0	0
History of MI	57,984	66
Previous PCI	57,984	66
Previous cardiac surgery	57,984	66
Angina	57,984	66
Hypertension	0	0
Peripheral vascular disease	57,984	66
Rest pain	57,984	66
Renal failure	0	0
Dialysis	0	0
Impaired sensorium	57,984	66
Coma > 24 hours	57,984	66
Hemiplegia	57,984	66
History of TIA	57,984	66
Stroke with neurologic deficit	57,984	66
Stroke without neurologic deficit	57,984	66
CNS tumor	57,984	66
Paraplegia	57,984	66
Quadriplegia	57,984	66
Disseminated cancer	0	0
Wound infection	0	0
Steroid use	0	0
Weight loss	0	0
Bleeding disorder	0	0
Preoperative blood transfusion	0	0
Chemotherapy	57,984	66
Radiotherapy	57,984	66
Pregnancy	58,214	66
Prior operation within 30 days	57,984	66

Total number of patients = 88,471; CHF = congestive heart failure; MI = myocardial infarction; PCI = percutaneous coronary intervention; TIA = transient ischemic attack; CNS = central nervous system.

Table 4. Missing preoperative laboratory values

Laboratory variable	Number of missing values	Percent missing
Na	12,262	14
BUN	14,438	16
Cr	12,607	14
Albumin	54,870	62
Bilirubin	56,020	63
AST	55,153	62
AlkPhos	56,268	64
WBC	8652	10
HCT	7830	9
Plt	8697	10
PTT	39,175	44
INR	33,575	38
PT	70,892	80

Total number of values = 88,471; Na = sodium; BUN = blood urea nitrogen; Cr = creatinine; AST = aspartate transaminase; AlkPhos = alkaline phosphatase; WBC = white blood cell count; HCT = hematocrit; Plt = platelets; PTT = partial thromboplastin time; INR = international normalized ratio; PT = and prothrombin time.

Table 5. Missing operating room times and length of stay

Variable	Number of missing values	Percent missing
Preoperative room time	55,808	63
Operative time	18	0
Postoperative room time	56,551	64
Total room time	53,321	60
Length of stay	26	0

< 10% missing data were incorporated into the regression with the exception of the demographic variables, for which patients with missing data were excluded (leaving 71,049 patients for analysis); 32 risk factors (78% of possible variables) were identified out of 41 included variables.

For each variable, the percent difference among the three odds ratios ranged from 0% (many variables) to 183% (dialysis). The percent difference in odds ratios was greatest for procedure types (mean difference = 35%) and least for laboratory results (mean difference = 4%). Twenty-five percent of variables (17 of 68) had an odds ratio difference of $\geq 30\%$. The three odds ratios were compared in terms of their 95% confidence intervals, and the following variables were found to have differences: white race (mean 1.2, 95% confidence interval [CI], 1.1–1.3 in regression 2, 1.7 [1.3–2.1] in regression 3), black race (1.3 [1.1–1.4] in regression 2, 1.8 [1.4–2.3] in

Table 6. Incidence of postoperative complications

Complication	Number of patients with complications	Percent with complications
Death	235	0
Coma lasting > 24 hours	7	0
Mechanical ventilation \geq 48 hours	394	0
Unplanned intubation	408	0
Stroke/cerebrovascular accident	99	0
Deep venous thrombosis	517	1
Pulmonary embolism	324	0
Superficial surgical site infection	771	1
Deep surgical site infection	487	1
Organ space surgical site infection	188	0
Sepsis/septic shock	729	1
Cardiac arrest	148	0
Myocardial infarction	170	0
Renal failure	62	0
Renal insufficiency	86	0
Return to operating room	2438	3
Wound dehiscence	190	0
Graft/prosthesis/flap failure	10	0
Blood transfusion	5292	6
Urinary tract infection	1138	1
Pneumonia	558	1
Peripheral nerve injury	62	0
Any adverse event	10,183	12

Total number of patients = 88,471.

regression 3), corpectomy (2.0 [1.3–2.1] in regression 1, 1.2 [1.0–1.3] in regression 2), operative time \geq 75th percentile (1.9 [1.4–2.6] in regression 1, 3.4 [3.2–3.5] in regression 2, 3.4 [3.2–3.6] in regression 3), total time in operating room \geq 75th percentile (1.5 [1.1–2.1] in regression 1, 1.0 [1.0–1.1] in regression 2), and hematocrit (0.93 [0.92–0.93] in regression 3 versus 0.95 [0.93–0.97] in regression 1 and 0.94 [0.93–0.94] in regression 2). Among variables found to be associated with any adverse event, the variables with the greatest effect were thoracic fusion, anterior lumbar fusion, and quadriplegia in regression 1; preoperative pneumonia, thoracic fusion, and operative time \geq 75th percentile in regression 2; and thoracic fusion, operative time \geq 75th percentile, and preoperative blood transfusion in regression 3 (Table 8). Of the 15 variables found to be in the top 10 for odds ratio in at least one regression, six variables were common to all three analyses (Table 8).

These discrepancies, in terms of both the agreement of which variables were associated with any adverse events and the magnitude of the effect, highlight the impact of different management strategies for missing data.

Table 7. Multivariate logistic regressions for any adverse event with differing treatments of missing data

Risk factor	Patients with missing data excluded (n = 5788)		Missing data treated as negative (n = 88,471)		Variables with missing data excluded* (n = 71,049)	
	OR	p value	OR	p value	OR	p value
Paraplegia	1.4	0.015	1.2	0.014	†	
Quadriplegia	2.2	0.008	2.0	< 0.001	†	
Preoperative pneumonia	2.4	0.245	3.5	0.017	†	
CNS tumor	1.2	0.632	1.4	0.164	†	
Previous cardiac surgery	0.8	0.292	0.9	0.117	†	
Angina	0.9	0.901	0.6	0.120	†	
Alcohol use	1.3	0.223	1.2	0.146	†	
Current smoker	1.1	0.514	1.0	0.555	1.0	0.185
Dependent functional status	1.4	0.013	1.4	< 0.001	1.6	< 0.001
Weight loss	1.6	0.202	1.6	0.001	1.9	< 0.001
Steroid use	1.2	0.272	1.2	< 0.001	1.2	< 0.001
Dialysis	0.6	0.532	1.7	0.013	1.7	0.001
Bleeding disorder	1.1	0.808	1.3	< 0.001	1.4	< 0.001
Preoperative wound infection	1.5	0.179	1.5	< 0.001	1.8	< 0.001
Preoperative transfusion	2.3	0.091	2.0	< 0.001	2.3	< 0.001
Congestive heart failure	2.2	0.202	1.2	0.223	1.5	0.025
Disseminated cancer	1.8	0.066	1.8	< 0.001	2.1	< 0.001
Age 50-59 years vs. age < 40 years	1.0	0.842	1.2	< 0.001	1.3	< 0.001
Age 60-69 years vs. age < 40 years	1.3	0.193	1.6	< 0.001	1.6	< 0.001
Age 70-79 years vs. age < 40 years	1.5	0.048	1.8	< 0.001	1.9	< 0.001
Age 80+ years vs. age < 40 years	2.0	0.003	1.9	< 0.001	2.0	< 0.001
BMI 30-35 kg/m ² vs. BMI < 25	1.1	0.377	1.1	0.108	1.1	0.061
BMI 35+ kg/m ² vs. BMI < 25	1.1	0.415	1.1	0.029	1.1	0.007
Male sex	0.9	0.202	0.9	0.001	0.9	0.039
White race	1.5	0.048	1.2	0.001	1.7	< 0.001
Black/African American race	1.8	0.007	1.3	< 0.001	1.8	< 0.001
Asian race	1.6	0.303	1.2	0.042	1.7	< 0.001
Other race	0.6	0.434	1.0	0.972	1.4	0.107
Hispanic ethnicity	0.9	0.458	0.9	0.044	0.9	0.105
ASA 3 versus ASA 1-2	1.3	0.027	1.4	< 0.001	1.3	< 0.001
ASA 4+ versus ASA 1-2	2.1	< 0.001	2.1	< 0.001	2.1	< 0.001
Procedure type						
Anterior cervical discectomy and fusion	0.7	0.079	0.6	< 0.001	0.6	< 0.001
Anterior lumbar fusion	3.2	< 0.001	2.1	< 0.001	2.1	< 0.001
Cervical laminectomy	1.8	0.014	1.3	< 0.001	1.3	< 0.001
Corpectomy	2.0	0.001	1.1	0.156	1.2	0.023
Lumbar laminectomy	1.8	< 0.001	1.3	< 0.001	1.3	< 0.001
Posterior lumbar fusion	2.0	< 0.001	2.0	< 0.001	2.1	< 0.001
Posterior cervical fusion	1.6	0.136	1.4	< 0.001	1.5	< 0.001
Thoracic fusion	3.4	< 0.001	3.4	< 0.001	3.7	< 0.001
Thoracic laminectomy	1.6	0.133	2.1	< 0.001	2.3	< 0.001
Operative time ≥ 75 percentile	1.9	< 0.001	3.4	< 0.001	3.4	< 0.001
Total time in OR ≥ 75 percentile	1.5	0.009	1.0	0.534	†	
Laboratory values [‡]						
Albumin	0.84	0.065	0.75	< 0.001	†	
AST	1.00	0.271	1.00	0.003	†	
AlkPhos	1.00	< 0.001	1.00	0.005	†	
Sodium	0.98	0.153	0.99	0.003	†	
BUN	1.00	0.965	1.01	0.002	†	
WBC	1.03	0.092	1.03	< 0.001	1.04	< 0.001
HCT	0.95	< 0.001	0.94	< 0.001	0.93	< 0.001
Platelets	1.00	0.167	1.00	0.236	1.00	0.219
INR	1.12	0.706	1.26	< 0.001	†	
PT	1.05	0.031	1.02	0.072	†	

Table includes all variables for which p < 0.20 in at least one regression; the following variables were included in the regressions but not included in the table: hemiplegia[†], stroke with neuro deficit[†], stroke without neuro deficit[†], transient ischemic attack[†], impaired sensorium[†], past myocardial infarction[†], previous percutaneous coronary intervention[†], previous cardiac surgery[†], peripheral vascular disease[†], rest pain[†], diabetes, age 40–49 years versus age < 40 years, body mass index 25–30 versus BMI < 25 kg/m², cervical laminotomy, bilirubin[†], creatinine[†], and partial thromboplastin time; shaded fields are statistically significant, p < 0.05; †variable excluded from multivariate regression; ‡odds ratios per one-unit increase in each laboratory value; OR = odds ratio; BMI = body mass index; ASA = American Society of Anesthesiologists classification; AST = aspartate transaminase; AlkPhos = alkaline phosphatase; BUN = blood urea nitrogen; WBC = white blood cell count; HCT = hematocrit; PTT = partial thromboplastin time; INR = international normalized ratio; PT = prothrombin time.

Table 8. Risk factors with greatest effect size (odds ratio) in each regression

Regression 1		Regression 2		Regression 3	
Variable	OR	Variable	OR	Variable	OR
1. Thoracic fusion	3.4	1. Preoperative pneumonia	3.5	1. Thoracic fusion	3.7
2. ALF	3.2	2. Thoracic fusion	3.4	2. Operative time \geq 75th percentile	3.4
3. Quadriplegia	2.2	2. Operative time \geq 75th percentile	3.4	3. Preoperative transfusion	2.3
4. ASA 4+ versus ASA 1–2	2.1	4. ASA4+ versus ASA 1–2	2.1	3. Thoracic laminectomy	2.3
5. Age 80+ years versus age < 40 years	2.0	4. ALF	2.1	5. Disseminated cancer	2.1
5. Corpectomy	2.0	4. Thoracic laminectomy	2.1	5. ASA 4+ versus ASA 1–2	2.1
5. PLF	2.0	7. Quadriplegia	2.0	5. ALF	2.1
8. Operative time \geq 75th percentile	1.9	7. Preoperative transfusion	2.0	5. PLF	2.1
9. Black/African American race	1.8	7. PLF	2.0	9. Age 80+ years versus age < 40 years	2.0
9. Cervical laminectomy	1.8	10. Age 80+ years versus age < 40 years	1.9	10. Weight loss	1.9
				10. Age 70–79 years versus age < 40 years	1.9

$p < 0.05$ for all values; bolding indicates that a risk factor was found to be in the top 10 for all three regressions;

OR = odds ratio; ALF = anterior lumbar fusion; PLF = posterior lumbar fusion.

Discussion

The quantity of spine research conducted using national databases has recently increased; however, there remain concerns about the quality of this research [11, 13, 17, 18]. Although missing data can occur even in high-quality prospective studies, the high prevalence of missing data in one of the most commonly used databases for clinical spine research, and the lack of acknowledgement of this limitation in many studies that derive from it, is concerning. The present study demonstrated that three common methods of handling missing data identify different variables as risk factors for adverse events with poor agreement (34%) and varying effect sizes. Therefore, different approaches to missing data could produce notably different answers to a particular question. In the scenario posed here, a study utilizing the second regression would find that experiencing a postoperative adverse event is most strongly associated with preoperative pneumonia, whereas the first regression would not find an association and the third method would drop the variable entirely. Varying outcomes like this one are a potential problem in many past database studies that failed to carefully consider the quantity of and approach to missing data.

These results are fairly consistent with past work on the effects of missing data handling. This study found less agreement among its analyses than others reported by Neuhaus and Ring [31] and by Biau and colleagues [10], although the magnitude of differences in odds ratios was similar to that of the latter. Of note, the proportion of variables with missing data and the percentage missing

were much greater in this study than those reported in the other studies, so the larger proportion of missing data in the current study may have led to more noticeable differences in outcomes. This parallels previous findings that missing data rates of $\geq 30\%$ created widely varying results compared with a rate of $\leq 10\%$ [16].

Another potential implication of missing data in database studies is in the calculation of various measures of overall medical comorbidity. One common measure of overall medical condition, the American Society of Anesthesiologists (ASA) physical status classification system, does not appear to be affected by missing data values. ASA classification is a single score determined preoperatively by an anesthesiologist to describe overall medical comorbidity [3], and it is collected by all NSQIP centers for all cases [40]. Missing ASA values were very uncommon in this study with only 145 missing values (0%). With such a low proportion, missing data handling techniques are unlikely to affect it [16]. However, there are other aggregate scores of medical condition that can be calculated from NSQIP variables such as the Charlson Comorbidity Index, the Modified Frailty Index, and the Elixhauser Comorbidity Index and have been used in prior database studies [5, 24, 29]. At least one variable included in each of these is frequently missing in NSQIP. Past database studies have used these measures to control for overall level of comorbidity, but it is not known whether missing values for individual comorbidities or laboratory values could affect these scores and possibly the outcomes of the studies. Further investigation is needed on this subject.

The present study had several limitations. As mentioned, the amount of missing data in the ACS-NSQIP has increased each year as more institutions and variables are included. Importantly, studies using the very early versions of the data set are less likely to be affected by the amount of missing data, whereas the percentage of missing data increases in studies using a greater number of years or more recent editions. With differing prevalence of missing data, it may be difficult to compare the results of studies that solely use early database years versus recent database years versus all editions of NSQIP. A study exclusively using early database years would likely be less susceptible to bias and variability from missing data than more recent studies. An additional limitation of this study is that only three methods for treating missing data were discussed. Although many methods of treating missing data have been described [26, 33], the aim of the present study was to illustrate how three commonly used methods in orthopaedic research could influence study results. Future research should explore the optimal method of treating missing data.

It is also important to note that differences in the number of observations and number of variables considered in each regression limit the ability to make direct comparisons. The sample size varied in each of the regressions with the sample size in the first regression (5788) considerably smaller than those in the second (88,471) and third regressions (71,049). The first regression found the fewest number of variables to be associated with any adverse event, but it is possible that additional variables would have been associated with a larger sample size and greater power. This is a well-recognized limitation of complete-case analysis and a notable reason it is not recommended in data sets with large proportions of missing data [26]. Similarly, the third regression considered 27 fewer variables than the other two regressions, because these variables were dropped as a result of large amounts of missing data, which may limit direct comparisons between the findings of regression 3 and the other two regressions.

Notably, there was a consistent rate of missing data (66%) in 20 commonly used comorbidity variables, including alcohol use, current pneumonia, history of stroke, and history of myocardial infarction. This recurring number suggested a systematic cause of this high percentage of missing data. Only after inquiry with our institution's surgical clinical reviewers did it become clear that this was because a fixed proportion of institutions was not collecting these data elements for 2011 to 2012, and these variables have since been discontinued altogether. For all these variables, the rate of missing data was 64% in 2011, 73% in 2012, and 100% in 2013 after discontinuation. As mentioned in the introduction, the ACS-NSQIP allowed certain institutions to collect a smaller subset of data for several years [1]. This is not revealed in the basic participant user

manual distributed with the data set, and researchers must be diligent when using data from more recent years.

The pattern of missing data for other variables does not have such a clear pattern but can still provide insights. Among the laboratory values, missingness was correlated with younger age, lower ASA class, and less medical comorbidities, suggesting that physicians selectively did not obtain laboratory values on seemingly healthier patients. This would be most consistent with a missing at random mechanism. For demographic variables, there was a correlation in missingness among race, Hispanic ethnicity, and body mass index, although never a perfect association (ie, all patients with missing body mass index were also missing race). There was a trend toward patients with missing values being younger, having lower ASA class, and Hispanic ethnicity. However, the correlation of other variables with missingness was less robust for demographic data, so the mechanism remains unclear.

In their proposal of how to optimally handle missing data from medical records, Haneuse and Daniels [20] note that the determining the missing data mechanism and imputing values is particularly complicated because there are many steps in which data can be recorded or omitted, which can vary at the level of the patient, physician, other provider, or health system. In many cases, different mechanisms may exist for different patients or variables in a single data set. To optimally understand the missing data mechanism and adjust for it, they propose that it is necessary to consider each step in which data may be missing and account for as many covariates and confounders as possible to explain the missing data [20]. In a database study, there are several barriers between the initial encounter and what is visible on the data sheet. One cannot always say whether a data point is missing because the physician did not ask about it, the patient declined it, it was not documented in the electronic medical record, a coding error by NSQIP, or it was the result of systemic differences in NSQIP collecting practices. Database research will never completely clarify these factors, but the database should explain its data collection practices and limitations to offer researchers as much information as possible.

Although missing data handling never has a perfect method and requires assumptions to be made, it is useful to examine how other investigators have addressed missing data in studies. In a review of randomized clinical trials in the orthopaedic literature, Herman and colleagues [21] found that leading methods, when specified, were carrying forward the last known observation (40%), imputation of mean or median values (9%), imputation of worst possible outcome (7%), and imputation based on more complex regression (4%). In a similar review of the otolaryngology literature, only 27% (85 of 316) explicitly mentioned handling of missing data. Among these, 86% (73 of 85)

used complete-case analysis or pairwise deletion. Of the remaining 12 studies to actively manipulate missing values, eight used mean substitution (67%), two concluded that data were missing completely at random, one used maximum likelihood estimation, and another used multiple imputations to replace missing values [30].

To study the amount of bias generated by several approaches to missing data, Hallgren and colleagues [19] artificially created missing data in two dichotomous variables, pulling from a completed clinical trial on alcohol dependence. The investigators compared complete-case analysis, last observation carried forward, worst-case scenario, and multiple imputation. They concluded that multiple imputations created less bias than other methods of handling missing data. To further account for missing data in many variables or in outcome variables, more sophisticated methods of multiple imputation have been developed such as multiple imputation using chained equations and multiple imputation and then deletion [39, 42].

As shown in this sample, the treatment of missing data can influence the results of spine studies performed with this data set. Many studies have used members of this cohort of spine patients in the ACS-NSQIP, and the majority of these studies fail to comment on the amount of missing data or how it was treated in analyses. This raises serious concerns—including increased risk of type I error, limited generalizability of results, and poor reproducibility—about drawing clinical conclusions from these types of database studies.

For researchers to better understand the pattern and mechanism of missing data, NSQIP and other databases should clearly explain any practices that could create missing data such as different data collection practices at different sites as much as possible without undermining patient or hospital privacy. General information about systemic sources of missing data, possibly including summary stats about hospitals that collected full versus limited data, could help investigators to determine the missing data mechanism and choose the best method to handle missing values.

It remains unclear which method of treating missing data is superior at this time. Although using complete cases is statistically simplest, there is the potential for selection bias when patients with missing data elements are excluded. Indeed, in the ACS-NSQIP, this may bias toward the inclusion of patients from large institutions that have the resources and staff to collect all variables. This method is generally not recommended, although it may be acceptable with small proportions of missing data and a suspected missing completely at random mechanism. Another solution is to include all patients but to exclude variables with large proportions of missing data in analyses, which

minimizes the risk of bias when trying to impute missing values. However, this limits the ability to control for important comorbidities in analyses. Newer model-based methods of imputing missing values such as maximum likelihood estimation or multiple imputations generate minimal bias and avoid many of the limitations of ad hoc analyses. However, some features of the missing data in NSQIP such as missing categorical data and ambiguous missing data mechanism may undermine the validity of the replaced values [33]. There are more nuanced versions of these techniques that may better account for a complex pattern of missing data [39, 42], but describing the optimal scenario for each of these is beyond the scope of this study.

Studies using NSQIP or other large databases require several important considerations in handling of missing data. One important consideration is the prevalence of dichotomous or categorical variables in the data set. A large proportion of medical comorbidities and postoperative complications is recorded as categorical variables, often simply “yes/no” [40]. Statistical tests using continuous, parametric variables are generally regarded as more robust than those using categorical variables, and methods of imputing missing data are considered superior in continuous variables as well [2, 26]. Additionally, patients in these databases tend to have a single measure of each variable, so there is no option to review past records to carry forward the most recent known value or otherwise impute the value based on patient-specific factors. Because of these limitations, handling of missing data from NSQIP should be considered with particular importance, and a clear explanation of handling methods is essential.

The gold standard for handling missing data in large databases like NSQIP remains unclear. General recommendations include (1) designing the study to minimize variables with missing data—in particular, studies that utilize later editions (ie, 2010 to the present) should not include the variables shown to have frequently missing values and irregular collection; (2) considering excluding variables with large amounts of missing data; (3) explicitly stating the rates of missing data for each variable and characterizing how missing values are treated; (4) favoring modern techniques like multiple imputations or maximum likelihood estimation over ad hoc methods; (5) considering the specific pattern of missing data, the missing data mechanism, and the types of variables studied to choose the optimal data handling method; and (6) performing several methods of data handling and comparing the varied outcomes that occur with different techniques and evaluate potential sources of bias.

Readers of studies that utilize NSQIP or databases should specifically seek whether the investigators declare the proportion and handling of missing data. If missing data methods are not mentioned or not robust, then the reader

should consider which variables are known to have large percentages of missing data and be most skeptical of findings based on these variables. Additionally, it may be worthwhile for other investigators to repeat prior studies using different database years and/or methods of handling missing values to determine whether these affect the outcomes of previously published work.

The ACS-NSQIP is the largest and highest quality surgical database developed to date; however, it does have its limitations. Differing treatments of missing data have the potential to influence the results from such database studies. It is important for researchers to be aware of the limitations of databases when designing, performing, and evaluating such investigations. Until there are better guidelines on the best ways to handle missing data, investigators should consistently report the amount of missing data and the corresponding treatment to increase the quality and transparency of orthopaedic database research. Readers of clinical research derived from these databases should specifically examine handling of missing data, if described, and question the outcomes of studies with no mention of missing data, key variables with high rates of missing values, or weak methods of handling missing data.

References

1. ACS NSQIP: Options and Pricing. 2015. Available at: <https://www.facs.org/~media/files/quality%20programs/nsqip/nsqipoptionspricingsheet1012.aspx>. Accessed November 1, 2015.
2. Altman DG, Bland JM. Parametric v non-parametric methods for data analysis. *BMJ*. 2009;338:a3167.
3. American Society of Anesthesiologists. New classification of physical status. *Anesthesiology*. 1963;24:111.
4. Basques BA, Bohl DD, Golinvaux NS, Leslie MP, Baumgaertner MR, Grauer JN. Postoperative length of stay and 30-day readmission after geriatric hip fracture: an analysis of 8434 patients. *J Orthop Trauma*. 2015;29:e115–120.
5. Basques BA, Golinvaux NS, Bohl DD, Yacob A, Toy JO, Varthi AG, Grauer JN. Use of an operating microscope during spine surgery is associated with minor increases in operating room times and no increased risk of infection. *Spine (Phila Pa 1976)*. 2014;39:1910–1916.
6. Basques BA, Ibe I, Samuel AM, Lukasiewicz AM, Webb ML, Bohl DD, Grauer JN. Predicting postoperative morbidity and readmission for revision posterior lumbar fusion. *Clin Spine Surg*. 2016 Jun 8. [Epub ahead of print]
7. Basques BA, Lukasiewicz AM, Samuel AM, Webb ML, Bohl DD, Smith BG, Grauer JN. Which pediatric orthopaedic procedures have the greatest risk of adverse outcomes? *J Pediatr Orthop*. 2015 Nov 10. [Epub ahead of print]
8. Bastiampillai R, Lavallée LT, Nossen S, Witiuk K, Mallick R, Fergusson D, Schramm D, Morash C, Cagiannos I, Breau RH. Laparoscopic nephroureterectomy is associated with higher risk of adverse events compared to laparoscopic radical nephrectomy. *Can Urol Assoc J*. 2016;10:126–131.
9. Bekelis K, Desai A, Bakhom SF, Missios S. A predictive model of complications after spine surgery: the National Surgical Quality Improvement Program (NSQIP) 2005–2010. *Spine J*. 2014;14:1247–1255.
10. Biau DJ, Katsahian S, Kartus J, Harilainen A, Feller JA, Sajovic M, Ejhered L, Zaffagnini S, Röpke M, Nizard R. Patellar tendon versus hamstring tendon autografts for reconstructing the anterior cruciate ligament: a meta-analysis based on individual patient data. *Am J Sports Med*. 2009;37:2470–2478.
11. Bohl DD, Basques BA, Golinvaux NS, Baumgaertner MR, Grauer JN. Nationwide Inpatient Sample and National Surgical Quality Improvement Program give different results in hip fracture studies. *Clin Orthop Relat Res*. 2014;472:1672–1680.
12. Bohl DD, Grauer JN, Leopold SS. Editor's Spotlight/Take 5: Nationwide inpatient sample and national surgical quality improvement program give different results in hip fracture studies. *Clin Orthop Relat Res*. 2014;472:1667–1671.
13. Bohl DD, Russo GS, Basques BA, Golinvaux NS, Fu MC, Long WD 3rd, Grauer JN. Variations in data collection methods between national databases affect study results: a comparison of the nationwide inpatient sample and national surgical quality improvement program databases for lumbar spine fusion procedures. *J Bone Joint Surg Am*. 2014;96:e193.
14. Dasenbrock HH, Rudy RF, Gormley WB, Smith TR, Frerichs KU, Aziz-Sultan MA, Du R. Predictors of complications after clipping of unruptured intracranial aneurysms: a National Surgical Quality Improvement Program analysis. *Neurosurgery*. 2016;63(Suppl 1):147.
15. Duchman KR, Gao Y, Pugely AJ, Martin CT, Noiseux NO, Callaghan JJ. The effect of smoking on short-term complications following total hip and knee arthroplasty. *J Bone Joint Surg Am*. 2015;97:1049–1058.
16. Fairclough DL, Peterson HF, Chang V. Why are missing quality of life data a problem in clinical trials of cancer therapy? *Stat Med*. 1998;17:667–677.
17. Golinvaux NS, Bohl DD, Basques BA, Fu MC, Gardner EC, Grauer JN. Limitations of administrative databases in spine research: a study in obesity. *Spine J*. 2014;14:2923–2928.
18. Golinvaux NS, Bohl DD, Basques BA, Grauer JN. Administrative database concerns: accuracy of international classification of diseases, ninth revision coding is poor for preoperative anemia in patients undergoing spinal fusion. *Spine (Phila Pa 1976)*. 2014;39:2019–2023.
19. Hallgren KA, Witkiewitz K, Kranzler HR, Falk DE, Litten RZ, O'Malley SS, Anton RF; in conjunction with the Alcohol Clinical Trials Initiative (ACTIVE) Workgroup. Missing data in alcohol clinical trials with binary outcomes. *Alcohol Clin Exp Res*. 2016;40:1548–1557.
20. Haneuse S, Daniels M. A general framework for considering selection bias in EHR-based studies: what data are observed and why? *EGEMs (Wash DC)*. 2016;4:1203.
21. Herman A, Botser IB, Tenenbaum S, Chechick A. Intention-to-treat analysis and accounting for missing data in orthopaedic randomized clinical trials. *J Bone Joint Surg Am*. 2009;91:2137–2143.
22. Kim BD, Edelstein AI, Patel AA, Lovecchio F, Kim JY. Preoperative anemia does not predict complications after single-level lumbar fusion: a propensity score-matched multicenter study. *Spine (Phila Pa 1976)*. 2014;39:1981–1989.
23. Kim BD, Smith TR, Lim S, Cybulski GR, Kim JY. Predictors of unplanned readmission in patients undergoing lumbar decompression: multi-institutional analysis of 7016 patients. *J Neurosurg Spine*. 2014;20:606–616.
24. Leven DM, Lee NJ, Kothari P, Steinberger J, Guzman J, Skovrlj B, Shin JL, Caridi JM, Cho SK. Frailty Index is a significant predictor of complications and mortality following surgery for adult spinal deformity. *Spine (Phila Pa 1976)*. 2016 Aug 31. [Epub ahead of print]

25. Lim S, Carabini LM, Kim RD, Khanna R, Dahdaleh NS, Smith ZA. Evaluation of ASA classification as thirty-day morbidity predictor after single-level elective anterior cervical discectomy and fusion. *Spine J*. 2016 Sep 23. [Epub ahead of print]
26. Little RJA, Rubin DB. *Statistical Analysis With Missing Data*, 2nd ed. Hoboken, NJ, USA: Wiley-Interscience; 2002.
27. Lovy AJ, Keswani A, Koehler SM, Kim J, Hausman M. Short-term complications of distal humerus fractures in elderly patients: open reduction internal fixation versus total elbow arthroplasty. *Geriatric Orthop Surg Rehabil*. 2016;7:39–44.
28. McCutcheon BA, Ciacci JD, Marcus LP, Noorbakhsh A, Gonda DD, McCafferty R, Taylor W, Chen CC, Carter BS, Chang DC. Thirty-day perioperative outcomes in spinal fusion by specialty within the NSQIP database. *Spine (Phila Pa 1976)*. 2015;40:1122–1131.
29. Menendez ME, Neuhaus V, Niek van Kijk C, Ring D. The Elixhauser comorbidity method outperforms the Charlson Index in predicting inpatient death after orthopaedic surgery. *Clin Orthop Relat Res*. 2014;472:2878–2886.
30. Netten AP, Dekker FW, Rieffe C, Soede W, Briaire JJ, Frijns JH. Missing data in the field of otorhinolaryngology and head & neck surgery: need for improvement. *Ear Hear*. 2016 Aug 22. [Epub ahead of print]
31. Neuhaus V, Ring D. Effect of different statistical methods on union or time to union in a published study about clavicular fractures. *J Shoulder Elbow Surg*. 2013;22:471–477.
32. Nüesch E, Trelle S, Reichenbach S, Rutjes AWS, Bürgi E, Scherer M, Altman DG, Jüni P. The effects of excluding patients from the analysis in randomized controlled trials: meta-epidemiological study. *BMJ*. 2009;339:b3244.
33. Pigott TD. A review of methods for missing data. *Educational Research and Evaluation*. 2001;7:353–383.
34. Pugely AJ, Martin CT, Gao Y, Klocke NF, Callaghan JJ, Marsh JL. A risk calculator for short-term morbidity and mortality after hip fracture surgery. *J Orthop Trauma*. 2014;28:63–69.
35. Roxbury CR, Jatana KR, Shah RK, Boss EF. Safety and post-operative adverse events in pediatric airway reconstruction: analysis of ACS-NSQIP-P 30-day outcomes. *Laryngoscope*. 2016 Jul 14. [Epub ahead of print]
36. Seicean A, Alan N, Seicean S, Worwag M, Neuhauser D, Benz EC, Weil RJ. Impact of increased body mass index on outcomes of elective spinal surgery. *Spine (Phila Pa 1976)*. 2014;39:1520–1530.
37. Seicean A, Seicean S, Alan N, Schiltz NK, Rosenbaum BP, Jones PK, Kattan MW, Neuhauser D, Weil RJ. Preoperative anemia and perioperative outcomes in patients who undergo elective spine surgery. *Spine (Phila Pa 1976)*. 2013;38:1331–1341.
38. Sorensen AA, Wojahn RD, Manske MC, Calfee RP. Using the Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement to assess reporting of observational trials in hand surgery. *J Hand Surg Am*. 2013;38:1584–1589.
39. Sullivan TR, Salter AB, Ryan P, Lee KJ. Bias and precision of the ‘multiple imputation, then deletion’ method for dealing with missing outcome data. *Am J Epidemiol*. 2015;182:528–534.
40. *User Guide for the 2013 ACS NSQIP Participant Use Data File*. Chicago, IL, USA: American College of Surgeons; 2014.
41. *User Guide for the Participant Use Data File*. Chicago, IL, USA: American College of Surgeons; 2007.
42. White IR, Royston P, Wood AM. Multiple imputation using chained equations: Issues and guidance for practice. *Stat Med*. 2011;30:377–399.
43. Wiens FH. *Merck Manual of Diagnosis and Therapy (Professional Version)*. Appendix: Normal Laboratory Values. Available at: <https://www.merckmanuals.com/professional/appendixes/normal-laboratory-values/blood-tests-normal-values#v8508814>. Accessed December 3, 2015.