Missing Data May Invalidate Hip Fracture Database Studies

Bryce A Basques, MD, MHS[1]*; Adam Lukasiewicz, MSc*[2]*; Andre Samuel, BBA*[2]*; Matthew Webb, BA*[2]*; Daniel D Bohl, MD, MPH*[1]*; Jonathan Grauer, MD*[2]*;*
[1]*Rush University Medical Center, Chicago, Illinois, USA;*
[2]*Yale School of Medicine, New Haven, Connecticut, USA*

Purpose/Background: National databases are increasingly being used for research in orthopaedics, as they offer significant power for analyses. However, these databases have significant limitations. One limitation that has received sparse mention in the literature is the prevalence of missing data. Studies using these databases often do not mention the percent of missing data for each variable used, and do not make note of how patients with missing data are incorporated into analyses. This study uses the American College of Surgeons National Quality Improvement Program (ACS-NSQIP) database to illustrate how different treatments of missing data can significantly skew results.

Methods: Patients who underwent hip fracture surgery between 2005 and 2013 were identified from the ACS-NSQIP database using Current Procedural Terminology (CPT) codes. Demographics, comorbidities, and type of procedure were tabulated for each patient and the percent of missing data was noted for each variable. These variables were tested for association with "any adverse event" using two separate multivariate regressions that used the two most common treatments for missing data. In the first regression, patients with any missing data were simply excluded. In the second regression, missing data were treated as a negative, or "reference" value. The results of these regressions were compared in order to determine how the different treatments of missing data can affect the results of hip fracture studies using the ACS-NSQIP database.

Results: A total of 26,066 hip fracture patients were identified. The average age was 80.1 ± 10.9 years (mean ± standard deviation). The following rates of missing data were found for each demographic category: 0.00% for age, 0.05% for sex, 12.25% for body mass index (BMI), and 18.19% for race. The rate of missing data was 70.94% 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. Multivariate logistic regressions for the association of demographics, comorbidities, and procedure characteristics with any adverse event within 30 days of surgery were performed with the two most common techniques for handling missing data: excluding patients with missing data, and treating missing data as the negative, or "reference" value. As seen in Table 1, these different techniques lead to finding vastly different significant risk factors for adverse events on multivariate analysis. Out of 17 risk factors found to be significantly associated with adverse events in either analysis, only six of these risk factors were common between the two regressions.

Conclusion: This study illustrates that a significant amount of missing data can be found in a hip fracture sample drawn from the ACS-NSQIP and extreme caution needs to be taken when selecting variables for inclusion in analyses. Specifically, 19 comorbidity variables have

See pages 47 - 108 for financial disclosure information.

70.94% missing data, as they are now only collected at certain ACS-NSQIP participating sites. This is not made clear in the basic participant user manual distributed with the data set and researchers must be diligent when using data from more recent years. In addition, as shown in this sample, the treatment of missing data can significantly affect the results of hip fracture studies performed with this data set. There are multiple studies in the literature that have used this cohort of hip fracture 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 study raises significant questions about the validity of these studies and it is important for researchers to be aware of the limitations of databases when designing, performing, and evaluating such investigations. It is critical that studies using these data sources report how missing data are handled.

Table 1. Results of multivariate analysis for any adverse event with differing treatments of missing data.

	After excluding patients with missing data (n=5,760)		Missing data treated as negative (n=26,066)	
Risk Factor	OR	P-value	OR	P-value
Cerebrovascular accident	1.4	0.021	1.4	0.003
Preoperative pneumonia			1.9	0.010
Angina			2.0	0.001
Impaired Sensorium			1.4	0.020
Age 70-79 vs age < 50			1.7	0.003
Age 80-89 vs age < 50			2.0	<0.001
Age 90+ vs age < 50			2.4	<0.001
BMI 25-30 vs BMI <25			0.9	0.023
BMI 35+ vs BMI <25			1.2	0.030
Native American or Pacific Islander vs White race	2.6	0.042		
Male sex	1.2	0.038	1.2	<0.001
ASA 3 vs ASA 1-2	2.2	<0.001	1.8	<0.001
ASA 4+ vs ASA 1-2	3.6	<0.001	3.0	<0.001
Procedure type (vs percutaneous pinning)				
Hemiarthroplasty	1.5	0.005	1.5	<0.001
Primary total hip arthroplasty			1.6	<0.001
Plate/screw			1.3	0.009
Intramedullary nail	1.4	0.020	1.4	0.001

OR = odds ratio; BMI = body mass index; ASA = American Society of Anesthesiologists.