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

Daniel D. Bohl, MPH, Glenn S. Russo, MD, Bryce A. Basques, BS, Nicholas S. Golinvaux, BA, Michael C. Fu, MD, William D. Long III, MD, and Jonathan N. Grauer, MD

Investigation performed at the Department of Orthopaedics and Rehabilitation, Yale School of Medicine, New Haven, Connecticut

Background: There has been an increasing use of national databases to conduct orthopaedic research. Questions regarding the validity and consistency of these studies have not been fully addressed. The purpose of this study was to test for similarity in reported measures between two national databases commonly used for orthopaedic research.

Methods: A retrospective cohort study of patients undergoing lumbar spinal fusion procedures during 2009 to 2011 was performed in two national databases: the Nationwide Inpatient Sample and the National Surgical Quality Improvement Program. Demographic characteristics, comorbidities, and inpatient adverse events were directly compared between databases.

Results: The total numbers of patients included were 144,098 from the Nationwide Inpatient Sample and 8434 from the National Surgical Quality Improvement Program. There were only small differences in demographic characteristics between the two databases. There were large differences between databases in the rates at which specific comorbidities were documented. Non-morbid obesity was documented at rates of 9.33% in the Nationwide Inpatient Sample and 36.93% in the National Surgical Quality Improvement Program (relative risk, 0.25; p < 0.05). Peripheral vascular disease was documented at rates of 2.35% in the Nationwide Inpatient Sample and 0.60% in the National Surgical Quality Improvement Program (relative risk, 3.89; p < 0.05). Similarly, there were large differences between databases in the rates at which specific inpatient adverse events were documented. Sepsis was documented at rates of 0.38% in the Nationwide Inpatient Sample and 0.81% in the National Surgical Quality Improvement Program (relative risk, 0.47; p < 0.05). Acute kidney injury was documented at rates of 1.79% in the Nationwide Inpatient Sample and 0.21% in the National Surgical Quality Improvement Program (relative risk, 8.54; p < 0.05).

Peer Review: This article was reviewed by the Editor-in-Chief and one Deputy Editor, and it underwent blinded review by two or more outside experts. It was also reviewed by an expert in methodology and statistics. The Deputy Editor reviewed each revision of the article, and it underwent a final review by the Editor-in-Chief prior to publication. Final corrections and clarifications occurred during one or more exchanges between the author(s) and copyeditors.

Disclosure: None of the authors received payments or services, either directly or indirectly (i.e., via his or her institution), from a third party in support of any aspect of this work. One or more of the authors, or his or her institution, has had a financial relationship, in the thirty-six months prior to submission of this work, with an entity in the biomedical arena that could be perceived to influence or have the potential to influence what is written in this work. No author has had any other relationships, or has engaged in any other activities, that could be perceived to influence or have the potential to influence what is written in this work. The complete Disclosures of Potential Conflicts of Interest submitted by authors are always provided with the online version of the article.

A commentary by Paul E. Levin, MD, is linked to the online version of this article at jbjs.org.
Conclusions: As database studies become more prevalent in orthopaedic surgery, authors, reviewers, and readers should view these studies with caution. This study shows that two commonly used databases can identify demographically similar patients undergoing a common orthopaedic procedure; however, the databases document markedly different rates of comorbidities and inpatient adverse events. The differences are likely the result of the very different mechanisms through which the databases collect their comorbidity and adverse event data. Findings highlight concerns regarding the validity of orthopaedic database research.

Orthopaedic surgery has recently seen a marked increase in the use of national databases to conduct research. By achieving sample sizes not possible through traditional single or multi-institution approaches, such databases enable the investigation of rare and catastrophic outcomes. By using intricate population sampling techniques, national databases enable the generation of national estimates and investigation of trends over time. The Nationwide Inpatient Sample (NIS) and the American College of Surgeons National Surgical Quality Improvement Program (ACS NSQIP) are two of the most frequently used databases in orthopaedic surgery research now.

Spine surgery is an area in which database studies have become particularly prevalent. A review of the spine literature reveals a striking increase in the numbers of database studies published over the past several years (Fig. 1). As examples, recent studies with use of NIS have investigated the complications associated with the use of bone morphogenetic proteins in spine fusion procedures7 and the incidence and mortality of cardiac events in lumbar spine surgery8. Similarly, recent studies with use of NSQIP have investigated the risk factors for morbidity and mortality immediately following spine surgery9 and pre-operative anemia and its impact on perioperative outcomes10.

Any study is limited by the biases of the data from which its results are drawn. In particular, studies based on administrative, reimbursement-derived databases like NIS are exposed to an array of biases associated with the pressures and laxities of the reimbursement system. Given the increasing use of national databases for orthopaedic research, it is imperative to understand to what extent the methods used to collect data may be impacting study results.

This study explores the inter-database reliability of NIS and NSQIP with respect to lumbar spine fusion. We hypothesized that we would be able to identify demographically similar populations of patients who underwent lumbar spine fusion from the two databases, but that there would be substantial differences in documented rates for selected comorbidities and inpatient adverse events.

Materials and Methods

A retrospective cohort study of patients undergoing lumbar fusion procedures during 2009 to 2011 was performed in two national databases: NIS and NSQIP. From NIS, we selected all inpatient stays associated with an International Classification of Diseases, Ninth Revision (ICD-9) procedure code of 81.06, 81.07, or 81.08. From NSQIP, we selected all surgical cases associated with a Current Procedural Terminology (CPT) code of 22533, 22558, 22612, 22630, or 22633. Detailed definitions of these codes are provided in the Appendix.

We did not attempt to further select patients on the basis of diagnosis (degenerative disease, trauma, or others) or type of lumbar fusion procedure (specific ICD-9 procedure code or CPT code).

Eight commonly studied comorbidities for which the presence could be identified in both NIS and NSQIP were selected (see Appendix). For NIS, the presence of each of the eight comorbidities was determined for each patient with use of either of the ICD-9 codes7 or directly reported comorbidity variables (which are summaries of ICD-9 codes). For NSQIP, the presence of each of the eight comorbidities was determined for each patient using directly reported comorbidity variables, anthropometric values, or laboratory values.

Avereen commonly studied adverse events for which the occurrence could be identified in both NIS and NSQIP were selected (see Appendix). For NIS, the occurrence of each of the fourteen adverse events was determined with use of ICD-9 codes7. For NSQIP, the occurrence of each of the fourteen adverse events was determined with use of directly reported adverse event variables. NIS is an inpatient-only database that only captures adverse events on patients before discharge. However, NSQIP collects data both before discharge and after discharge up to the thirtieth postoperative day; hence, adverse events in NSQIP were categorized as having occurred while the patient was still an inpatient (including any time on the day of discharge) or after the patient was discharged.

First, demographic characteristics were compared between databases with use of the Pearson chi-square test.
Second, comorbidity and inpatient adverse event rates were compared between databases with use of Poisson regression with robust error variance\(^8\),\(^9\) with an indicator variable for the database (NIS compared with NSQIP). When Poisson regression is applied to binomial data, the error for the estimated relative risk will be overestimated; hence, Poisson regression with robust error variance (which in this case refers to a technique called "sandwich estimation," which corrects the error estimates)\(^8\) was used to directly estimate the error for the relative risk. These calculations were performed with use of Stata version 13.1 (StataCorp, College Station, Texas) using the following code:

```
glm [depvar] [indepvar(s)], fam(poisson) link(log) nolog vce(robust) eform
```

Third, in an effort to exclude confounding as a source of differences in comorbidity and adverse event rates, Poisson regressions were repeated with adjustment for age (eighteen to forty-nine years, fifty to fifty-nine years, sixty to sixty-nine years, or seventy years or more), sex (male or female), admission type (non-elective or elective), location before admission (home, a different acute care hospital, or another type of health facility), and number of levels fused (one to two, three to seven, or eight or more).

Finally, in NSQIP, documented rates of inpatient (before-discharge) adverse events were compared with those of after-discharge adverse events. NSQIP collects data on patients both before and after discharge up to the thirtieth postoperative day, but NIS only collects data on patients before discharge; the purpose of this final analysis was to characterize the extent of limitation that the inpatient-only nature of NIS imposes on studies that use its data.

All statistical tests were two-tailed, and the level of significance was set at \(\alpha = 0.05\).

**Source of Funding**

There was no source of external funding for this study.

**Results**

The study population was constituted by 144,098 patients from NIS and 8434 patients from NSQIP who were identified as undergoing lumbar fusion procedures. A comparison of the demographic characteristics of these patients between databases showed relatively small (but significant) differences in age (\(p < 0.001\)), location before admission (\(p < 0.001\)), and number of levels fused (\(p < 0.001\)) (Table 1). There were no differences in sex or admission type. Detailed age distributions are shown in Figure 2.

Documented rates of comorbidities are compared between databases in Figure 3. The figure is sorted by relative risk.

![Distribution of Patient Age in NIS](image)

![Distribution of Patient Age in NSQIP](image)

**Fig. 2**

The distributions of patient age. Patient age was within two years of each other for NIS (Fig. 2-A) and NSQIP (Fig. 2-B) with regard to the means (57.4 and 58.9 years) and the 5th percentile (thirty-two and thirty-four years), 25th percentile (forty-seven and forty-nine years), 50th percentile (forty-eight and sixty years), 75th percentile (sixty-eight and sixty-nine years), and 95th percentile (seventy-nine and eighty years). There was a relatively small, but significant, difference in age distributions between the two databases (\(p < 0.001\) from the Pearson chi-square test as shown in Table 1). Note that the regularly spaced spikes in the distributions are the result of some states’ requirement that age be rounded to make data less identifiable.
A comparison of comorbidity rates between NIS and NSQIP. The left graph gives the rates at which each comorbidity was documented in each of the databases. The middle graph gives the relative risk comparing the rate in NIS with that in NSQIP (simply the rate in NIS divided by the rate in NSQIP). The right graph gives these same relative risks adjusted for demographic differences between databases. The error bars indicate the 95% confidence intervals of the relative risks. The asterisks indicate that the relative risks were significantly different from 1.00 (p < 0.05). Non-morbid obesity represents a body mass index of 30.0 to 39.9 kg/m². Morbid obesity represents a body mass index of ≥40.0 kg/m². COPD = chronic obstructive pulmonary disease and PVD = peripheral vascular disease.

(middle graph) comparing the documented rate in NIS with that in NSQIP. These relative risks ranged from 0.25 to 3.89. Non-morbid obesity (a body mass index of 30 to 39.9 kg/m²) was documented at rates of 9.33% in NIS and 36.93% in NSQIP. Peripheral vascular disease was documented at rates of 2.35% in NIS and 0.60% in NSQIP. Most differences persisted after adjustment for demographic characteristics.

Documented rates of inpatient adverse events are compared between databases in Figure 4. The figure is sorted by relative risk (middle graph) comparing the documented rate in NIS with that in NSQIP. These relative risks ranged from 0.47 to 8.54. Sepsis was documented at rates of 0.38% in NIS and 0.81% in NSQIP. Acute kidney injury was documented at rates of 0.15% in NIS and 0.21% in NSQIP. Most differences persisted after adjustment for demographic characteristics.

Finally, in NSQIP, documented rates of before-discharge adverse events are compared with those of after-discharge adverse events in Figure 5. The figure is sorted by ascending ratio of before-discharge rates of adverse events to after-discharge rates of adverse events (right graph), so those adverse events listed at the top of the figure are those that have the highest proportion occurring after discharge. NIS and other inpatient-only databases do not capture those adverse events occurring after discharge, which for those adverse events listed toward the top of the figure is a large proportion. Events with the highest ratio of after-discharge rates to before-discharge rates of adverse events included surgical site infection, wound disruption, and urinary tract infection.

Discussion

This study demonstrates that two cohorts undergoing a commonly performed orthopaedic procedure, each selected from national databases that use very different methods for data acquisition, have reasonably similar demographic characteristics. However, despite the demographic similarities between these patients, the two databases document the occurrence of specific comorbidities and inpatient adverse events at markedly different rates. These findings raise concerns about the external validity of database studies in orthopaedics—in particular, those studies drawing on administrative reimbursement databases like NIS.

The differences in comorbidity and inpatient adverse event rates identified in this study are most likely the results of the very different mechanisms through which NIS and NSQIP generate their comorbidity and adverse event data. NIS is an administrative database based on ICD-9 codes. Hospital-employed coders review charts to identify “reimbursable” comorbidities and adverse events. These coders then list these identified comorbidities and adverse events on reimbursement claims submitted to payers.
The disadvantage of this system is that the data are subject to a wide array of potential biases imposed by the pressures and laxities of the reimbursement system. For example, comorbidities or adverse events that will not increase payment may be less likely to be reported; similarly, those that do increase payment may be more likely to be reported.

The procedures used to generate data in NSQIP are very different. For NSQIP, risk-assessment nurses undergo rigorous uniform training to become dedicated surgical clinical reviewers. These surgical clinical reviewers, each located at a data-contributing institution, conduct a directed, thorough, retrospective review of each prospectively identified patient’s medical records. This review is conducted for the sole purpose of collecting clinical data for NSQIP for quality-improvement and research purposes. Items reviewed include the inpatient chart, outpatient notes, and laboratory values. For a patient to be considered to have each specific comorbidity or adverse event, a very specifically defined set of criteria, involving physician documentation and/or laboratory values, must be met. Inter-rater reliability is calculated for all variables and has very low rates of disagreement.

Because NSQIP involves such a highly protocolled medical record review that is designed specifically to capture clinical data, it can be regarded as a sort of gold-standard retrospective review with which NIS can be compared. In this context, by highlighting the major differences between the databases in documented rates of comorbidities and inpatient adverse events, this study underlines the potential pitfalls of studies drawing conclusions regarding comorbidities and adverse events from administrative data. Administrative databases may be acceptable to track trends or to make comparisons between competing procedures or patient factors, as the databases may be internally valid; however, the differences reported here suggest that readers should be cautious in interpreting studies claiming to establish any solid estimate of the incidence or rate of an adverse event using administrative data. Given NSQIP’s more recent development, much more orthopaedic research has been conducted with use of NIS and other administrative databases than with use of NSQIP (Fig. 1). However, we expect the rate of publication with use of NSQIP data to increase fairly rapidly as additional data become available each year.

A final lesson that can be drawn from these data involves the inpatient-only nature of NIS and other inpatient-only databases. NIS only includes data on adverse events occurring before discharge. In contrast, NSQIP collects data on adverse events occurring both during the inpatient stay and after discharge up to the thirtieth postoperative day. Figure 5 ranks adverse events by the ratio of adverse events occurring before or after discharge. Adverse events occurring mostly after discharge (those listed toward the top of the figure) are unlikely to be captured by inpatient-only databases like NIS. For example, with surgical site infection, the before-discharge adverse event rate was only 0.28%, but the after-discharge adverse event rate...
was 2.12%. The reader should not take for granted the major limitation that the abridged postoperative follow-up of an inpatient-only database incurs.

This study was limited most by the lack of patient-identifiable data with which our results might have been directly compared on a patient-by-patient basis. To protect patient privacy, NIS and NSQIP are de-identified at an early stage in the data-collection process. There was also the potential for bias resulting from a comparison of nonequivalent parameters by matching ICD-9 codes to NSQIP criteria. However, we made every effort to select codes that both were aligned with NSQIP criteria and were representative of the methods used in current literature.

Compared with registry-based systems like NSQIP, administrative databases like NIS may provide superior sampling of U.S. hospitalizations. However, by showing that the two databases identify patients with reasonably similar demographic characteristics but produce dramatically different estimates of the rates of many comorbidities and inpatient adverse events (even after adjustment for the minor differences in demographic characteristics), this study raises concerns regarding the validity of conclusions drawn from administrative databases. As database studies are published with increasing frequency, researchers, readers, and reviewers alike must pay particular attention to the strengths and weaknesses of the data sources to ensure proper interpretation of study results.

Appendix

Tables showing codes used to identify lumbar fusion procedures in NIS and NSQIP, data elements from NIS and NSQIP used to identify cases with comorbidities, and data elements from NIS and NSQIP used to identify cases with adverse events are available with the online version of this article as a data supplement at jbjs.org.

Daniel D. Bohl, MPH
Glenn S. Russo, MD
Bryce A. Basques, BS
Nicholas S. Golinvaux, BA
William D. Long III, MD
Jonathan N. Grauer, MD
Department of Orthopaedics and Rehabilitation, Yale School of Medicine, 800 Howard Avenue, New Haven, CT 06510.
E-mail address for J.N. Grauer: jonathan.grauer@yale.edu
References