TRAUMA

Do we really know our patient population in database research?

A COMPARISON OF THE FEMORAL SHAFT FRACTURE PATIENT POPULATIONS IN THREE COMMONLY USED NATIONAL DATABASES

Aims
While use of large national clinical databases for orthopaedic trauma research has increased dramatically, there has been little study of the differences in populations contained therein. In this study we aimed to compare populations of patients with femoral shaft fractures across three commonly used national databases, specifically with regard to age and comorbidities.

Patients and Methods
Patients were identified in the Nationwide Inpatient Sample (NIS), National Surgical Quality Improvement Program (NSQIP) and National Trauma Data Bank (NTDB).

Results
The distributions of age and Charleston comorbidity index (CCI) reflected a predominantly older population with more comorbidities in NSQIP (mean age 71.5; SD 15.6), mean CCI 4.9; SD 1.9) than in the NTDB (mean age 45.2; SD 21.4), mean CCI = 2.1; SD 2.0). Bimodal distributions in the NIS population showed a more mixed population (mean age 56.9; SD 24.9), mean CCI 3.2; SD 2.3). Differences in age and CCI were all statistically significant (p < 0.001).

Conclusion
While these databases have been commonly used for orthopaedic trauma research, differences in the populations they represent are not always readily apparent. Care must be taken to understand fully these differences before performing or evaluating database research, as the outcomes they detail can only be analysed in context.

Take home message: Researchers and those evaluating research should be aware that orthopaedic trauma populations contained in commonly studied national databases may differ substantially based on sampling methods and inclusion criteria.

Cite this article: Bone Joint J 2016;98-B:425–32.

The use of national databases for orthopaedic clinical research has increased dramatically over the past five years (Fig. 1). Examples of such databases include the Nationwide Inpatient Sample (NIS), the American College of Surgeons National Surgical Quality Improvement Program database (NSQIP) and the American College of Surgeons National Trauma Data Bank Research Data Set (NTDB RDS). Their use can expedite clinical research and allow the study of rare injury patterns by making large cohorts available, something not possible for the individual researcher or institution. With increased use of these databases, however, it is crucial to understand how their populations differ and how patients are included and excluded from each before conclusions can be made and generalised to external populations.

Previous comparisons of databases have focused on differences in the quality of data, namely that pertaining to comorbidities and adverse events. These studies of patients undergoing lumbar vertebral fusions, hip fracture fixation, pancreaticoduodenectomy and oesophageal resection identified considerable differences in the recording of this data between administratively-coded databases, such as NIS, and clinical registry databases such as NSQIP. These studies focused, however, on patient populations that were largely demographically similar.

Regarding high-energy orthopaedic trauma, various databases may differ in which patients are included because of varying inclusion criteria. The NIS, NSQIP and NTDB RDS have all been used extensively for orthopaedic trauma...
research, but there has been little analysis of the extent to which these samples can be generalised to the wider population of orthopaedic trauma patients. In this retrospective cohort study we compared patients with femoral shaft fractures across three commonly-used databases, looking specifically at age and comorbidities.

**Patients and Methods**

**Study design and setting.** A retrospective cohort study was performed using the NTDB RDS and NSQIP from 2011 and 2012 and NIS data from 2011 only. The study was approved by the local institutional review board.

**Patients.** International Classification of Diseases (ICD-9 79.05, 79.15, 79.25, 79.35) and Current Procedural Terminology codes (current procedural terminology T 27500, 27502, 27506, 27507) were used to identify patients undergoing surgical treatment of femoral shaft fractures.

**Methods.** Patient age and comorbidities with equivalent definitions across the three databases were queried, namely: alcoholism, coagulopathy, diabetes mellitus, cancer, hypertension, obesity, and smoking. The supplementary material available online lists specific data elements used to identify comorbidities in each database.

A modified Charlson comorbidity index (CCI) was derived from this. The modified score has been shown to have comparable predictive value to the original and has been used previously with national databases such as NSQIP.

A total of nine adverse events with equivalent definitions across databases were queried: acute kidney injury (AKI), cardiac arrest, cerebrovascular accident (CVA), death, deep vein thrombosis or pulmonary embolism (DVT/PE), myocardial infarction (MI), pneumonia, surgical site infection (SSI), and urinary tract infection (UTI). As the databases differ in the periods of time for which this is recorded, only in-patient events were included. From these, rates of mortality, serious adverse event (SAE) occurrence (death, cardiac arrest, MI, CVA, DVT/PE, and SSI), and all adverse event (AAE) incidence were calculated.

**Statistical analysis.** Mean age and CCI were compared by one-way analysis of variance (ANOVA) and visual impact of data was presented in histograms. The rates of incidence of each of the seven comorbidities were compared between databases using Pearson’s chi-squared statistic and the relative risk (RR) with 95% confidence intervals (CI) estimated for both NSQIP and NTDB. The NSQIP and NTDB data were compared with NIS as a reference, and then NSQIP compared with NTDB using Poisson regression with robust error variance. NIS was chosen as the denominator for both comparisons as this is thought to be the most nationally representative patient population. Pearson’s chi-squared statistic was used to compare rates of mortality and adverse events between populations. Multivariate ANOVA was then used to compare the rates after adjusting for age and CCI.

Finally, a simulated theoretical analysis of risk factors associated with inpatient adverse events was conducted for each database population. Multivariate logistic regression was used to identify any predictive relationship between age and comorbidities and the rate of adverse events.
All statistical analyses were performed using Stata version 13.0 (StataCorp, LP, College Station, Texas). All statistical tests were two-tailed, and the level of significance was set at $\alpha = 0.05$.

Results
A total of 25,121 patients undergoing surgery for a femoral shaft fracture were included in this study. This total comprised 3,943 patients from NIS, 663 from NSQIP and 20,515 from NTDB. The mean age of patients in NIS was 56.9 years (standard deviation (SD) 24.9) in NSQIP 71.5 years (SD 15.6) and in NTDB 45.2 years (SD 21.4). The differences were statistically significant (ANOVA; $p < 0.001$). Both NSQIP and NTDB round ages greater than 90 years to a value of 90 to maintain anonymity. The distribution of ages is markedly different between NSQIP (primarily older patients) and NTDB (primarily younger patients), with the bimodal distribution in NIS seemingly representing a more mixed population (Fig. 2).

The mean CCI of patients in NIS was 3.2 (SD 2.3), in NSQIP 4.9 (SD 1.9) and in NTDB 2.1 (SD 2.0). These differences were also statistically significant (ANOVA; $p < 0.001$). Similar to age, the distribution of CCI is markedly different between patients in NSQIP (more comorbidities) and NTDB (primarily fewer comorbidities), with NIS again representing a more mixed population with a bimodal distribution (Fig. 3).

The NSQIP population had the highest incidence of all comorbidities except alcoholism and smoking, which were greatest in the NTDB population (Fig. 4). The incidence of each in the NIS population lay between those in the NTDB and NSQIP respectively. The difference in RR between NSQIP and NIS populations was statistically significant (Poisson regression: disseminated cancer ($p < 0.001$), obesity ($p < 0.001$), bleeding disorder ($p < 0.001$), hypertension ($p < 0.001$), alcoholism ($p < 0.001$)), except for diabetes (Poisson regression; $p = 0.173$) and smoking status (Poisson regression: $p = 0.184$). The relative risks of all comorbidities in the NTDB population were significantly different to those in the NIS (Poisson regression: disseminated cancer ($p = 0.040$), obesity ($p < 0.001$), bleeding disorder ($p < 0.001$), hypertension ($p < 0.001$), diabetes ($p < 0.001$), smoker ($p = 0.011$), alcoholism ($p < 0.001$)).

The differences in mortality rate between the different databases were not statistically significant before risk adjustment (Pearson’s chi–squared; $p = 0.059$) but, after adjusting for age and CCI, the differences became significant (Multivariate ANOVA; $p < 0.001$). The differences in rate of significant adverse events were statistically significant before and after risk adjustment (Pearson’s chi–squared; $p < 0.001$).
Multivariate ANOVA; \( p < 0.001 \) as were those of all adverse events (Pearson’s chi-squared; \( p < 0.001 \), Multivariate ANOVA; \( p < 0.001 \)).

In the simulated theoretical analysis there were several differences identified in the risk factors for adverse events occurring while an in-patient (Table I). In the NIS population, increasing age was associated with more AAEs, a trend also seen in the NSQIP population. In the NTDB population, however, the converse was true. Similarly, diabetes was associated with more AAEs in the NIS and NTDB populations but not in the NSQIP population.

**Discussion**

Over the past three years there has been a more than two-fold increase in peer-reviewed publications using NIS, NSQIP and NTDB data in orthopaedic journals alone (Fig. 1). The commonly used databases treat trauma patients very differently, however, and the resulting differences in database

---

**Fig. 3a**

Graphs showing the distribution of Charlson comorbidity index (CCI) of patients with a fracture of the femoral shaft by database: a) Nationwide Inpatient Sample, b) the American College of Surgeons National Surgical Quality Improvement Program database and c) the American College of Surgeons National Trauma Data Bank Research Data Set.

\[ n = 3951 \]

\[ n = 663 \]

\[ n = 20,501 \]
populations lead to significant differences in outcomes reported as a result of the nature of the enquiries made to each database.

Our study found that the NSQIP and NTDB populations are strikingly different, when considering age and comorbidities. The NTDB population is younger and has fewer comorbidities, while the NSQIP population is older with more. The NIS population appears to be bimodally distributed in terms of age and comorbidities. While previous database comparisons have primarily focused on differences in specific data elements among demographically similar populations, this study is the first to highlight the

The incidence of individual comorbidities by database. Error bars represent 95% confidence interval of relative risks. Nationwide Inpatient Sample (NIS), the American College of Surgeons National Surgical Quality Improvement Program database (NSQIP) or the American College of Surgeons National Trauma Data Bank Research Data Set (NTDB).

Table I. Theoretical analysis of risk factors associated with inpatient adverse events after femoral shaft fractures in three commonly used national clinical databases

<table>
<thead>
<tr>
<th></th>
<th>NIS (n = 3951)</th>
<th>NSQIP (n = 663)</th>
<th>NTDB (n = 20,501)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Odds ratio (95% CI)</td>
<td>p-value</td>
<td>Odds ratio (95% CI)</td>
</tr>
<tr>
<td>Alcoholism</td>
<td>1.22 (0.84 to 1.77)</td>
<td>0.296</td>
<td>0.00 (omitted)</td>
</tr>
<tr>
<td>Smoking</td>
<td>0.60 (0.44 to 0.81)†</td>
<td>0.001†</td>
<td>1.68 (0.77 to 3.69)</td>
</tr>
<tr>
<td>Diabetes</td>
<td>1.40 (1.14 to 1.72)†</td>
<td>0.001†</td>
<td>0.41 (0.18 to 0.95)</td>
</tr>
<tr>
<td>Hypertension</td>
<td>1.04 (0.86 to 1.26)</td>
<td>0.699</td>
<td>1.54 (0.82 to 2.89)</td>
</tr>
<tr>
<td>Bleeding</td>
<td>1.87 (1.41 to 2.48)†</td>
<td>&lt;0.001†</td>
<td>2.06 (1.03 to 4.12)</td>
</tr>
<tr>
<td>Obesity</td>
<td>1.63 (1.25 to 2.14)†</td>
<td>&lt;0.001†</td>
<td>1.15 (0.60 to 2.19)</td>
</tr>
<tr>
<td>Disseminated cancer</td>
<td>2.09 (1.00 to 4.34)†</td>
<td>0.049†</td>
<td>1.04 (0.38 to 2.85)</td>
</tr>
</tbody>
</table>

Age (yrs)

<table>
<thead>
<tr>
<th></th>
<th>NIS (n = 3951)</th>
<th>NSQIP (n = 663)</th>
<th>NTDB (n = 20,501)</th>
</tr>
</thead>
<tbody>
<tr>
<td>18 to 29</td>
<td>0.57 (0.39 to 0.84)</td>
<td>0.005</td>
<td>0.00 (omitted)</td>
</tr>
<tr>
<td>30 to 39</td>
<td>0.86 (0.56 to 1.31)</td>
<td>0.477</td>
<td>0.00 (omitted)</td>
</tr>
<tr>
<td>40 to 49</td>
<td>1.26 (0.82 to 1.92)</td>
<td>0.29</td>
<td>2.52 (0.33 to 19.38)</td>
</tr>
<tr>
<td>50 to 59</td>
<td>Reference</td>
<td>Reference</td>
<td>Reference</td>
</tr>
<tr>
<td>60 to 69</td>
<td>1.38 (0.96 to 2.00)</td>
<td>0.086</td>
<td>3.53 (0.74 to 16.81)</td>
</tr>
<tr>
<td>70 to 79</td>
<td>1.53 (1.07 to 2.18)†</td>
<td>0.019†</td>
<td>4.81 (1.06 to 21.74)</td>
</tr>
<tr>
<td>80 to 89</td>
<td>2.84 (2.02 to 3.99)†</td>
<td>&lt;0.001†</td>
<td>3.82 (0.81 to 17.92)</td>
</tr>
<tr>
<td>90 +</td>
<td>4.12 (2.82 to 6.01)†</td>
<td>&lt;0.001†</td>
<td>5.56 (1.13 to 23.42)</td>
</tr>
</tbody>
</table>

* Categories were omitted from National Surgical Quality Improvement Program (NSQIP) analysis that perfectly predicted absence of adverse events in multivariate analysis. Nationwide Inpatient Sample (NIS); American College of Surgeons National Trauma Data Bank Research Data Set (NTDB)

† Indicates statistically significant odds ratio with positive association with inpatient adverse events
‡ Indicates statistically significant odds ratio with negative association with inpatient adverse events CI, confidence interval
significant demographic differences of trauma populations in national databases all used for orthopaedic trauma research.

As expected, the differences in populations also lead to differences in outcomes although, interestingly, mortality was not significantly different between populations. This is likely to be because of the low overall mortality of femoral fractures. In our theoretical study of risk factors for inpatient adverse events, there were several differences in the simulated results. Most striking was the change in directionality of the association of certain risk factors (diabetes, older age) with adverse events, depending on the database studied. This illustrates the importance of understanding the database population being studied before attempting to generalise results of a study to other clinical populations.

The differences in database populations highlight their relative strengths and weaknesses database (Table II). Based on sampling methodology, NIS is the most nationally representative and comprehensive sample. NSQIP and NTDB include only data from voluntarily participating institutions and so could be argued to be biased towards hospitals that have interests in quality monitoring and improvement. By contrast the Hospital Cost and Utilization Project takes data on all discharges from a nationally representative sample of community hospitals (approximately 20% of all those in the United States). While this is a relative strength of the NIS dataset, the quality of the billing code data it contains has been called into question. It contains solely ICD-9 billing code diagnoses, as opposed to NSQIP and NTDB which use diagnoses abstracted from patient notes for comorbidity and adverse event coding. Inconsistencies have been noted in ICD-9 coding when comparing databases with administratively-coded versus chart-abstracted data. As administrative billing codes are subject to economic and political pressures, there is potential for under-reporting or over-reporting certain diagnoses.

The NSQIP is primarily geared towards elective surgery and most acute trauma and transplant procedures are excluded. Review of NSQIP data collection guidelines for surgical clinical reviewers indicates that most high energy mechanisms of injury, such as motor vehicle accidents and firearms are excluded from the database. Lower-energy mechanisms included consist of falls from standing or up to three steps’ height, resulting in non-penetrating, single-bone or single organ system injury. The population with fractures in NSQIP is, therefore, likely to be skewed towards patients with pre-existing comorbidities such as osteoporosis or metastatic disease. Nevertheless, studies

### Table II. Comparison of the three databases studied

<table>
<thead>
<tr>
<th>Overseeing organisation</th>
<th>Nationwide Inpatient Sample (NIS)</th>
<th>National Surgical Quality Improvement Program (NSQIP) Participant User Data File (PUF)</th>
<th>National Trauma Data Bank (NTDB) Research Data Set (RDS)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Earliest year available</td>
<td>Healthcare Cost and Utilization Project (HCUP)</td>
<td>American College of Surgeons</td>
<td>American College of Surgeons</td>
</tr>
<tr>
<td>Participating hospitals</td>
<td>Stratified systematic sample from all HCUP hospitals (equivalent to 20% of all discharges from United States community hospitals)</td>
<td>NSQIP participant institutions (requires annual membership renewal and NSQIP surgical clinical reviewer on staff)</td>
<td>Any voluntarily participating institution</td>
</tr>
<tr>
<td>Patient inclusion criteria</td>
<td>All inpatient discharges</td>
<td>Systematic sampling (every 8 days) of hospital’s daily surgical log. Hospital may include general/vascular cases only or multispecialty. Most acute trauma, transplant, and minor surgical cases excluded.</td>
<td>All trauma cases (based on ICD-9 diagnosis) resulting in inpatient admission or death.</td>
</tr>
<tr>
<td>Data collection methods</td>
<td>Hospital billing codes</td>
<td>Chart review by NSQIP-trained surgical clinical reviewer with regular inter-rater reliability audits</td>
<td>Institution dependent (chart abstraction/billing code data)</td>
</tr>
<tr>
<td>Number of cases (2011)</td>
<td>8,023,590</td>
<td>442,149</td>
<td>773,299</td>
</tr>
<tr>
<td>Number of hospitals (2011)</td>
<td>1,049</td>
<td>315</td>
<td>744</td>
</tr>
<tr>
<td>Number of variables reported:</td>
<td>187</td>
<td>252</td>
<td>116</td>
</tr>
<tr>
<td>Diagnosis coding</td>
<td>ICD-9</td>
<td>ICD-9</td>
<td>ICD-9</td>
</tr>
<tr>
<td>Number of diagnoses reported</td>
<td>Multiple</td>
<td>Single preoperative diagnosis</td>
<td>Multiple</td>
</tr>
<tr>
<td>Procedure coding</td>
<td>ICD-9</td>
<td>Current procedural terminology</td>
<td>ICD-9</td>
</tr>
<tr>
<td>Number of procedures reported</td>
<td>Multiple</td>
<td>Multiple</td>
<td>Multiple</td>
</tr>
<tr>
<td>Procedure timing</td>
<td>By day</td>
<td>By day</td>
<td>By hour</td>
</tr>
<tr>
<td>Pre-hospital data</td>
<td>Not available</td>
<td>Not available</td>
<td>From initial emergency medical services dispatch</td>
</tr>
<tr>
<td>Post-discharge data</td>
<td>Not available</td>
<td>Up to 30 post-operative days</td>
<td>No</td>
</tr>
<tr>
<td>Comorbidity data source</td>
<td>ICD-9 diagnosis codes</td>
<td>Chart-review by surgical clinical reviewers using NSQIP comorbidity definitions</td>
<td>Institution dependent (Chart-review/ICD-9 codes). NTDB recommended comorbidity definitions</td>
</tr>
<tr>
<td>Adverse events data source</td>
<td>ICD-9 diagnosis codes</td>
<td>Chart-review by surgical clinical reviewers using NSQIP adverse event definitions</td>
<td>Institution dependent Chart-review/ICD-9 codes. NTDB recommended adverse event definitions</td>
</tr>
</tbody>
</table>

THE BONE & JOINT JOURNAL
have been published using this data to draw generalised conclusions.\textsuperscript{3,14,16,17} While its data collection system is robust and revolves around trained surgical case reviewers and regular inter-rater reliability audits, trauma researchers should consider carefully whether it genuinely reflects the population about which they wish to draw conclusions.

The NTDB, by contrast, gives arguably the best representation of the acute, high-energy fracture patient population. All patients admitted through the emergency department at participating institutions with a trauma-specific ICD-9 diagnosis code (800.0 to 959.9) meet NTDB inclusion criteria.\textsuperscript{3} This excludes pathologic fractures (ICD-9 733.1X), stress fractures (733.9X), and nonunion or malunion (733.8X). This may explain why the femoral shaft fracture population in NTDB is younger.

No database serves as a benchmark for the study of trauma populations but investigation of each of them may have merit. Femoral shaft fractures represent a spectrum of trauma populations but investigation of each of them may have merit. Femoral shaft fractures represent a spectrum of injuries resulting from both low- and high-energy mechanisms, implicitly reflected by the bimodal distribution of age seen in NIS. As a true national sample it is ideal for national-level demographic and epidemiological studies.

Each database also offers some unique data permitting different analyses. The NSQIP has several unique operative variables such as operative time, anaesthetic, and transfusion while NTDB contains detailed pre-hospital information, such as ambulance timings and mechanism of injury. The NIS is maybe more limited owing to the constraints of ICD-9 but is useful because of its nationally weighted sampling.

The primary limitation of this study is the lack of ability to match patients between databases which would permit more detailed comparison and more insight into whether the differences were driven by which institutions were included, or which patients. Owing to the large sample sizes, statistically significant results may be drawn from each database but be due to type I error. It is therefore crucial that researchers understand and take steps to mitigate this.

Supplementary material
Data elements tables are available alongside the online version of this article at www.bjj.boneand-joint.org.uk

References


