The Nationwide Inpatient Sample database does not accurately reflect surgical indications for fusion

Clinical article

Yakov GoloGorsky, M.D., John J. Knightly, M.D., John H. Chi, M.D., M.P.H., and Michael W. Groff, M.D.

1 Department of Neurosurgery, Brigham and Women’s Hospital, Harvard Medical School, Boston, Massachusetts; and 2 Atlantic Neurosurgical Specialists, Atlantic Neuroscience Institute, Morristown, New Jersey

Object. The rates of lumbar spinal fusion operations have increased dramatically over the past 2 decades, and several studies based on administrative databases such as the Nationwide Inpatient Sample (NIS) have suggested that the greatest rise is in the general categories of degenerative disc disease and disc herniation, neither of which is a well-accepted indication for lumbar fusion. The administrative databases classify cases with the International Classification of Disease, Ninth Revision, Clinical Modification (ICD-9-CM). The ICD-9-CM discharge codes are not generated by surgeons but rather are assigned by trained hospital medical coders. It is unclear how accurately they capture the surgeon’s indication for fusion. The authors sought to compare the ICD-9-CM code(s) assigned by the medical coder to the surgeon’s indication based on a review of the medical chart.

Methods. A retrospective review was undertaken of all lumbar fusions performed at our institution by the department of neurosurgery between 8/1/2011 and 8/31/2013. Based on the authors’ review, the indication for fusion for each case was categorized as spondylolisthesis, deformity, tumor, infection, nonpathological fracture, pseudarthrosis, adjacent-level degeneration, stenosis, degenerative disc pathology, or disc herniation. These surgeon diagnoses were compared with the primary ICD-9-CM codes that were submitted to administrative databases.

Results. There were 178 lumbar fusion operations performed for 170 hospital admissions. There were 44 hospitalizations in which fusion was performed for tumor, infection, or nonpathological fracture; the remaining 126 were for degenerative diagnoses. For these degenerative cases, the primary ICD-9-CM diagnosis matched the surgeon’s diagnosis in only 61 of 126 degenerative cases (48.4%). When both the primary and all secondary ICD-9-CM diagnoses were considered, the indication for fusion was identified in 100 of 126 cases (79.4%).

Conclusions. Characterizing indications for fusion based solely on primary ICD-9-CM codes extracted from large administrative databases does not accurately reflect the surgeon’s indication. While these databases may accurately describe national rates of lumbar fusion surgery, the lack of fidelity in the source codes limits their role in accurately identifying indications for surgery. Studying relationships among indications, complications, and outcomes stratified solely by ICD-9-CM codes is not well founded.

Key Words: NIS • administrative databases • data fidelity • ICD-9-CM • lumbar fusion • indications

Large administrative databases have assumed a major role in population-based studies examining health care delivery. Two of the largest databases include the Medicare Provider Analysis and Review (MedPAR) database and the Nationwide Inpatient Sample (NIS). The MedPAR database includes 100% of Medicare hospital claims, while the NIS is a component of the Healthcare Cost and Utilization Project based at the federal Agency for Healthcare Research and Quality (http://www.hcup-us.ahrq.gov/nisoverview.jsp). The NIS collects data from states that themselves receive data from individual hospitals regarding their inpatient hospitalizations. Over time, the NIS has received input from an increasing number of states, and in 2011 the database included data on 8 million hospital stays, drawing from 1045 hospitals located in 46 states (http://www.hcup-us.ahrq.gov/nisoverview.jsp). Thus, the database has become increasingly more representative of the national population.

Both of these large databases use the International
Classification of Disease, Ninth Revision, Clinical Modification (ICD-9-CM) procedure and diagnosis codes to identify and categorize surgeries to describe trends in volume, rates of surgery, and complications. Because the ICD-9-CM system was designed primarily for billing purposes, it may not be sufficient for understanding health care policy research, in particular indications for procedures. Despite this limitation, numerous studies in the past 2 decades have exploited the NIS and MedPAR databases to not only document rising rates of lumbar fusion but also demonstrate trends for specific lumbar spine diagnoses. As the conclusions from these investigations significantly influence the current debate on health care policy, the data on which they are based must be critically evaluated.

The processing of information in medical records, which is then entered into databases for later analysis, follows a typical sequence in most hospitals. Trained medical coders abstract the clinical information in the medical record and the discharge summary. Numerical codes for diagnoses, procedures, and complications are assigned according to the ICD-9-CM. One primary and as many as 19 secondary codes are assigned for each hospitalization. These codes are then collated into a discharge abstract, which is reported to state or federal databases. Unfortunately, the rich data set found in the patient record that motivates the assignment of the ICD-9-CM codes is not available to researchers studying the NIS.

As we practice at a hospital that contributes to the NIS, we sought to examine the quality of the data that is submitted to the database and compare it to the information in the medical record. The aim of this study was twofold: First, our goal was to evaluate the accuracy of the primary ICD-9-CM diagnosis code in reflecting the true indication for fusion surgery as documented in the medical record. This surgical diagnosis was derived from a careful review of the entire medical record by an individual with domain-specific knowledge in the area of lumbar fusion. By contrast, the trained medical coders have less domain-specific training and focus their efforts on a small portion of the medical record. This surgical diagnosis was derived from a careful review of the entire medical record by an individual with domain-specific knowledge in the area of lumbar fusion. By contrast, the trained medical coders have less domain-specific training and focus their efforts on a small portion of the medical record. Second, we wanted to clarify how often any of the secondary ICD-9-CM codes are in agreement with the surgical indication.

### Methods

Approval to conduct this study was obtained from the Institutional Review Board of the Brigham and Women's Hospital. This study retrospectively examined the demographic, diagnostic, and coder-related data of 170 consecutive hospitalizations involving 168 patients undergoing lumbar fusion between August 1, 2011, and August 31, 2013, in one department at a single tertiary care hospital. The billing records from all operations performed by two of the authors (J.H.C. and M.W.G.) were identified and reviewed. All cases that did not involve fusion of the lumbar spine were excluded. All of the hospitalization ICD-9-CM–based diagnoses for the fusion cases were obtained, and the medical charts were reviewed by a panel consisting of 3 fellowship-trained surgeons. Interrater reliability (IRR) coefficients were calculated. To evaluate the possibility of an undue impact of the surgeon performing the case, direct comparison of that surgeon to each other member of the group was calculated. It bears mention that in our study the operating surgeon had no influence on the outcome variable, the ICD-9 code, which was independently established by the coders in each case. Admissions with multiple lumbar fusion operations were counted only once.

Demographic data including sex, age, body mass index (BMI), smoking status, and surgical indication for fusion were recorded. Indication for fusion was categorized as follows: spondylolisthesis (isthmic and degenerative), deformity (coronal or sagittal), tumor, infection, nonpathological fracture (trauma), pseudarthrosis, adjacent-level degeneration after prior lumbar fusion, stenosis, or degenerative disc disease (DDD). The stenosis category included 1) the presence of central or foraminal stenosis without any of the other aforementioned diagnoses or 2) cases in which adequate decompression would result in iatrogenic instability (i.e., removal of 50% of the joints bilaterally or 100% unilaterally). The DDD category included 2 subgroups: 1) patients with twice-recurrent disc herniations after index discectomy and 2) patients with single-level DDD in whom the indication for fusion approximated the criteria outlined by the Swedish Trial. Namely, these patients must have a) pain duration of at least 1 year, b) back pain that is more pronounced than leg pain, c) high disability scores on the EQ-5D metric and/or be on work leave/disability, and d) degenerative changes only at L4–5 or L5–S1 on CT and/or MRI.

Operative data including number of levels fused and surgical approach taken were collected. Surgical approach was broken into four categories: 1) posterolateral, 2) transfornaminal lumbar interbody fusion (TLIF), posterior lumbar interbody fusion (PLIF), or posterior vertebral column resection (PVCR) with placement of a cage, 3) anterior or lateral, and 4) circumferential (360°).

The surgeon's indication for fusion was identified for each case and compared against the primary and all secondary ICD-9-CM diagnoses. It is important to note that the surgical indication for fusion was not necessarily the primary indication for surgery. For example, a patient presenting with neurogenic claudication due to lumbar stenosis secondary to spondylolisthesis would be classified as having “spondylolisthesis” rather than “spinal stenosis.”

### Results

Between August 1, 2011, and August 31, 2013, two of the authors (J.H.C. and M.W.G.) performed 178 operations involving fusion of the lumbar spine across 170 admissions. In 8 patients, 2 operations were performed in a single hospitalization: 2 for staged procedures and 6 for revision of misplaced hardware. Of the 170 admissions, 44 were for fusion relating to tumor, infection, or nonpathological fracture, and 126 were for degenerative diagnoses. One patient was hospitalized on 3 separate occasions for lumbar fusion: an index case performed for spondylolisthesis, and 2 subsequent operations were conducted for pseudarthrosis and hardware failure. Currently at our institution there are 12 inpatient coders. Coders participate in a training program and have the following coding credentials: CCS, or Certified Cod-
ing Specialist; RHIT, or Registered Health Information Technician; and RHIA, Registered Health Information Administrator. The CCS credential is earned after some
one takes a 9-month-long coding course, passes a credential-
ing exam, and maintains yearly continuing education credits. The RHIT designation is an associate’s degree program, while the RHIA designation is a bachelor’s program. Coding at our institution is performed within 5 days of a patient’s discharge. Coders have access to the entire inpatient medical record, including preoperative office visits and medical assessments that directly apply to the index hospitalization. Findings identified on radiological films may be used as confirmation but may not be coded unless confirmed by the surgical team. For example, if the surgeon mentions spondylolisthesis but does not specify the level, the coders may use the radiology report for clarification. However, if there is no mention of spondylolisthesis by the surgical team, the coders cannot use this information in their discharge abstract.

Overall, there were 168 patients undergoing 178 lumbar fusions across 170 separate hospitalizations. There were 76 men and 92 women whose median age was 57.5 years (range 21–96 years), and the median BMI was 28.5 (range 15.8–51.4). Thirty-three patients (19.4%) were active smokers at the time of surgery. The median number of levels operated was 3 (range 2–9). Demographic data are summarized in Table 1. Median rather than mean values are presented because several subcategories have few patients. The surgical approach to fusion was described as posterolateral in 68 hospitalizations, TLIF/PLIF/PVCR plus a cage in 87, anterior or lateral in 8, and circumferential in 7.

The most common surgical indications for fusion were, in descending order, spondylolisthesis (n = 82), tumor/pathological fracture (n = 20), trauma/nonpathological fracture (n = 19), DDD (n = 18), deformity (n = 11), pseudarthrosis (n = 9), infection (n = 5), stenosis (n = 4), and adjacent-level disease (n = 2). An IRR analysis was performed to assess the degree to which the coders consistently assigned a given surgical indication for fusion to each subject in the study. Because all coders rated all subjects, and there were no missing data, the most appropriate coefficients for IRR analysis were Fleiss’ \( \kappa \), Krippendorff’s \( \alpha \), and Light’s arithmetic mean of all paired Cohen’s \( \kappa \) coefficients. The resulting coefficients were as follows: Fleiss’ \( \kappa = 0.825 \), Krippendorff’s \( \alpha = 0.826 \), and Light’s \( \kappa = 0.825 \). These coefficients all indicate near-perfect coder agreement in assigning surgical indication for fusion based on review of the medical chart.

Overall, the primary ICD-9-CM code most accurately captured the surgical indication in only 104 (61.2%) of 170 cases. One of the secondary codes correctly described the surgical indication in another 40 cases (23.5%). Thus, the ICD-9-CM hospital codes failed to identify the correct surgical indication for fusion in 15.3% of hospitalizations (Table 2). There is an emerging consensus in the literature that several ICD-9-CM diagnosis categories, including tumor, infection, and trauma, are well-accepted indications for fusion. These diagnoses were listed as the primary ICD-9-CM code in 44 of the admissions in our data set, and in 98% of these cases, the primary ICD-9-CM code accurately captured the surgical indication. This paper focuses on the remaining 126 cases for which the ICD-9-CM diagnosis is considered to be degenerative in nature. In this important subset, the primary ICD-9-CM code accurately reflected the surgical indication in only 61 cases (48.4%). However, when the secondary codes in the NIS database were also considered, the likelihood of finding the surgical diagnosis increased approximately twofold (79.4%). For example, when the primary ICD-9-CM diagnosis category was spinal stenosis, the surgical indication for fusion was appropriately identified in only 4 (13.3%) of 30 cases. When all secondary diagnoses were searched, however, the indication was identified in 60% of cases. Similarly, when the ICD-9-CM diagnosis category was DDD, the rationale for fusion was more commonly found in the secondary diagnoses. Even when all diagnoses were considered, the surgical indication for fusion was not identified in 20.6% of cases.

**Illustrative Cases**

**Case 1**

This 42-year-old cable television technician injured his back while using a shovel, and this resulted in a persistent significant sharp pain on the right side of the back radiating down the right leg. He reported numbness, tin-

<table>
<thead>
<tr>
<th>Surgeon’s Diagnosis</th>
<th>No. of Cases</th>
<th>Median Age (yrs)</th>
<th>Median BMI (kg/m²)</th>
<th>Median No. of Levels</th>
<th>Male (%)</th>
<th>Smokers (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>spondylolisthesis</td>
<td>82</td>
<td>63</td>
<td>28.6</td>
<td>2</td>
<td>39.0</td>
<td>17.1</td>
</tr>
<tr>
<td>pathological fracture—tumors</td>
<td>20</td>
<td>53.5</td>
<td>25.4</td>
<td>5</td>
<td>60.0</td>
<td>5</td>
</tr>
<tr>
<td>nonpathological fracture—tumors</td>
<td>19</td>
<td>53</td>
<td>22.7</td>
<td>5</td>
<td>63.2</td>
<td>31.6</td>
</tr>
<tr>
<td>DDD</td>
<td>18</td>
<td>45</td>
<td>29.3</td>
<td>2</td>
<td>50.0</td>
<td>33.3</td>
</tr>
<tr>
<td>deformity</td>
<td>11</td>
<td>59</td>
<td>29.5</td>
<td>4</td>
<td>27.3</td>
<td>9.1</td>
</tr>
<tr>
<td>pseudarthrosis</td>
<td>9</td>
<td>55</td>
<td>32.0</td>
<td>3</td>
<td>66.7</td>
<td>22.2</td>
</tr>
<tr>
<td>infection</td>
<td>5</td>
<td>54</td>
<td>27.6</td>
<td>3</td>
<td>40.0</td>
<td>40</td>
</tr>
<tr>
<td>severe stenosis</td>
<td>4</td>
<td>67</td>
<td>28.8</td>
<td>2.5</td>
<td>0</td>
<td>25</td>
</tr>
<tr>
<td>adjacent-level degeneration</td>
<td>2</td>
<td>61</td>
<td>30.75</td>
<td>2</td>
<td>100.0</td>
<td>0</td>
</tr>
<tr>
<td>total (range)</td>
<td>170</td>
<td>57.5 (21–96)</td>
<td>28.5 (15.8–51.4)</td>
<td>3 (2–9)</td>
<td>45.9</td>
<td>19.4</td>
</tr>
</tbody>
</table>
NIS database does not reflect indications for fusion

<table>
<thead>
<tr>
<th>Surgeon’s Diagnosis</th>
<th>No. of Cases</th>
<th>The Primary ICD-9-CM Diagnosis Code Captures</th>
<th>A Secondary ICD-9-CM Diagnosis Code Captures</th>
<th>Surgeon’s Diagnosis Not Captured</th>
</tr>
</thead>
<tbody>
<tr>
<td>pathological fracture—tumors</td>
<td>20</td>
<td>20/20 (100%)</td>
<td>0/20 (0%)</td>
<td>0/20 (0%)</td>
</tr>
<tr>
<td>nonpathological fracture (trauma)</td>
<td>19</td>
<td>19/19 (100%)</td>
<td>0/19 (0%)</td>
<td>0/19 (0%)</td>
</tr>
<tr>
<td>infection</td>
<td>5</td>
<td>4/5 (80%)</td>
<td>1/5 (20%)</td>
<td>0/5 (0%)</td>
</tr>
<tr>
<td>nondegenerative diagnoses subtotal</td>
<td>44</td>
<td>43/44 (97.7%)</td>
<td>1/44 (2.3%)</td>
<td>0/44 (0%)</td>
</tr>
<tr>
<td>spondylolisthesis</td>
<td>82</td>
<td>29/82 (35.4%)</td>
<td>34/82 (41.5%)</td>
<td>19/82 (23.5%)</td>
</tr>
<tr>
<td>DDD</td>
<td>18</td>
<td>18/18 (100%)</td>
<td>0/18 (0%)</td>
<td>0/18 (0%)</td>
</tr>
<tr>
<td>deformity</td>
<td>11</td>
<td>4/11 (36.4%)</td>
<td>5/11 (45.5%)</td>
<td>2/11 (18.2%)</td>
</tr>
<tr>
<td>pseudarthrosis</td>
<td>9</td>
<td>9/9 (100%)</td>
<td>0/9 (0%)</td>
<td>0/9 (0%)</td>
</tr>
<tr>
<td>severe stenosis</td>
<td>4</td>
<td>1/4 (25%)</td>
<td>0/4 (0%)</td>
<td>3/4 (75%)</td>
</tr>
<tr>
<td>adjacent-level degeneration</td>
<td>2</td>
<td>0/2 (0%)</td>
<td>0/2 (0%)</td>
<td>2/2 (100%)</td>
</tr>
<tr>
<td>degenerative diagnoses subtotal</td>
<td>126</td>
<td>61/126 (48.4%)</td>
<td>39/126 (31.1%)</td>
<td>26/126 (20.6%)</td>
</tr>
<tr>
<td>total</td>
<td>170</td>
<td>104/170 (61.2%)</td>
<td>40/170 (23.5%)</td>
<td>26/170 (15.3%)</td>
</tr>
</tbody>
</table>

Table 2: Accuracy of discharge ICD-9-CM diagnosis codes in capturing surgical indication for fusion, stratified by surgical diagnosis

Fig. 1. Sagittal (left) and axial (right) T2-weighted MR images demonstrating Grade I–II spondylolisthesis at L4–5 with unroofing of the posterior disc causing bilateral neural foraminal narrowing and lateral recess stenosis.

Fig. 2. Inverted lateral dynamic radiographs, flexion (left) and extension (right) views, showing bilateral spondylolisthesis (arrow) with a movable subluxation.

Gling, and cramping pain that was associated with weakness in the right leg. On physical examination he had Grade 4/5 strength in his right anterior tibialis and extensor hallucis longus muscles, as well as hypesthesia to light touch in the L-4 and L-5 right dermatomes. MRI demonstrated an L4–5 Grade I–II spondylolisthesis, with unroofing of the posterior disc that caused bilateral neural foraminal narrowing, and lateral recess stenosis (Fig. 1). There was also evidence of a bilateral spondylolisthesis with a movable subluxation shown on dynamic radiographs (Fig. 2). Nonoperative therapy including physical therapy, lumbar epidural steroid injections, and chiropractic manipulation failed to resolve his symptoms.

The patient underwent a minimally invasive right-sided TLIF, placement of an interbody cage, and reduction of the spondylolisthesis. Postoperatively, he experienced resolution of his low-back pain and radiculopathy. The primary ICD-9-CM diagnosis for this admission was 722.10 (displacement of lumbar intervertebral disc without myelopathy). The fourth secondary diagnosis was 738.4, acquired spondylolisthesis (Table 3). That is, for an investigator assuming that the primary diagnosis is the indication for fusion, this example would lead to the erroneous conclusion that fusion was done for a poorly recognized indication for fusion, disc herniation, whereas the surgical indication of spondylolisthesis is in fact generally accepted.

Case 2

This 74-year-old woman with medical history noteworthy for rheumatoid arthritis and chronic steroid use, Type II diabetes mellitus, hypertension, osteoporosis, and 3 prior lumbar surgeries, including an L2–5 laminectomy, presented with a several-year history of progressive and debilitating low-back pain with burning in her anterior thighs bilaterally, greater on the left side than the right. Preoperative CT and MRI demonstrated degenerative lumbar levoscoliosis, anterolisthesis of L-2 on L-3, retrolisthesis of L-3 on L-4, anterolisthesis of L-5 on S-1, and an old T-12 compression fracture. At L2–3, there was severe DDD that, superimposed on the spondylolisthesis, resulted in severe neural foraminal stenosis bilaterally (Fig. 3). Nonoperative therapy did not result in durable relief.
<table>
<thead>
<tr>
<th>ICD-9-CM</th>
<th>No. of Dx</th>
<th>Code</th>
<th>Description</th>
<th>Code</th>
<th>Description</th>
<th>Code</th>
<th>Description</th>
<th>Code</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>primary diagnosis code</td>
<td>1</td>
<td>722.10</td>
<td>displacement of lumbar intervertebral disc w/o myelopathy</td>
<td>722.52</td>
<td>degeneration of lumbar or lumbosacral intervertebral disc</td>
<td>724.02</td>
<td>spinal stenosis, lumbar region, w/o neurogenic claudication</td>
<td>721.3</td>
<td>lumbosacral spondylosis w/o myelopathy</td>
</tr>
<tr>
<td>secondary diagnoses codes</td>
<td>2</td>
<td>584.9</td>
<td>renal failure, acute, unspecified</td>
<td>724.02</td>
<td>spinal stenosis, lumbar region, w/o neurogenic claudication</td>
<td>202.80</td>
<td>other malignant lymphomas, unspecified site, extranasal &amp; solid organ sites</td>
<td>414.01</td>
<td>coronary atherosclerosis of native coronary artery</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>276.51</td>
<td>dehydration</td>
<td>722.83</td>
<td>postlaminectomy syndrome, lumbar spine</td>
<td>738.4</td>
<td>acquired spondylolisthesis</td>
<td>V45.82</td>
<td>postprocedural percutaneous transluminal coronary angioplasty</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>738.4</td>
<td>acquired spondylolisthesis</td>
<td>338.29</td>
<td>other chronic pain</td>
<td>780.62</td>
<td>postprocedural fever</td>
<td>272.4</td>
<td>other &amp; unspecified hyperlipidemia</td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>782.0</td>
<td>symptoms involving skin &amp; other integumentary tissue</td>
<td>401.9</td>
<td>hypertension, unspecified</td>
<td>401.9</td>
<td>hypertension, unspecified</td>
<td>401.9</td>
<td>hypertension, unspecified</td>
</tr>
<tr>
<td></td>
<td>6</td>
<td>E920.4</td>
<td>accidents caused by other hand tools &amp; implements</td>
<td>250.00</td>
<td>diabetes mellitus, Type II or unspecified type, not stated as uncontrolled</td>
<td>414.01</td>
<td>coronary atherosclerosis of native coronary artery</td>
<td>V58.63</td>
<td>long-term (current) use of antiplatelet/antithrombotic</td>
</tr>
<tr>
<td></td>
<td>7</td>
<td>E016.0</td>
<td>activities involving digging, shoveling, &amp; raking</td>
<td>786.09</td>
<td>other respiratory abnormalities</td>
<td>272.0</td>
<td>pure hypercholesterolemia</td>
<td>V58.66</td>
<td>long-term (current) use of aspirin</td>
</tr>
<tr>
<td></td>
<td>8</td>
<td>E000.0</td>
<td>civilian activity done for income or pay</td>
<td></td>
<td></td>
<td>530.81</td>
<td>gastroesophageal reflux, no esophagitis</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>9</td>
<td>275.41</td>
<td>hypocalcemia</td>
<td></td>
<td></td>
<td>600.00</td>
<td>benign prostatic hypertrophy w/o urinary obstruction</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>10</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>274.9</td>
<td>gout, unspecified</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>11</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>V15.5</td>
<td>personal history of injury presenting hazards to health</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>12</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>V87.41</td>
<td>personal history of antineoplastic chemotherapy</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

* Boldface text identifies the surgical indication for fusion. In Case 4, surgical indication was not captured. Dx = diagnoses.
NIS database does not reflect indications for fusion

The patient underwent a staged left retroperitoneal direct lateral interbody discectomy and fusion, placement of a cage at L2–3, a posterior L2–4 revision laminectomy for decompression, and L2–4 posterolateral spinal arthrodesis and deformity correction. Her leg pain has resolved, while her low-back pain is much improved but not completely resolved. The primary ICD-9-CM diagnosis for this admission was 722.52 (degeneration of lumbar or lumbosacral intervertebral disc). The second and third secondary diagnoses were 737.39 (other kyphoscoliosis and scoliosis) and 722.83 (postlaminectomy syndrome, lumbar spine), respectively (Table 3). Although the primary diagnosis relates to disc degeneration, the driving indication for two-stage fusion was deformity as well as severe foraminal stenosis necessitating near-complete facetectomy.

Case 3

This 76-year-old former construction worker had a longstanding history of episodic low-back pain. He reported that these episodes had increased in frequency and severity and now greatly limited his activities of daily living and life in general. He endorsed bilateral L-5 radiculopathies, worse on the right. He was neurologically intact. MRI demonstrated a Grade II–III anterolisthesis of L-5 on S-1, bilateral L-5 spondylolysis, and severe bilateral foraminal stenosis (Fig. 4). Nonoperative therapy including physical therapy and chiropractic manipulation failed to result in pain relief.

The patient underwent an L4–S1 Gill laminectomy and posterolateral arthrodesis with pedicle screw fixation that resulted in reduction to a Grade I spondylolisthesis. His low-back pain and radiculopathies resolved completely. The primary ICD-9-CM diagnosis for this admission was 724.02 (spinal stenosis, lumbar region, without neurogenic claudication). The third secondary diagnosis was 738.4, acquired spondylolisthesis (Table 3). Thus, an investigator would erroneously conclude that fusion was done for spinal stenosis, whereas the surgical indication was spondylolisthesis.

Case 4

This 65-year-old male physician had progressive claudication symptoms, low-back pain, bilateral L-5 radiculopathies with atrophy of the anterior tibialis muscles, and weakness in ankle dorsiflexion and the extensor hallucis longus muscle. MRI and CT demonstrated significant lumbar stenosis at L4–5 with a Grade I spondylolisthesis (Fig. 5). Dynamic radiographs demonstrated a 6-mm spondylolisthesis that increased to 12 mm on flexion (Fig. 6). Physical therapy did not alleviate his symptoms.

The patient underwent a minimally invasive left-sided TLIF, placement of an interbody cage, L4–5 laminectomy from the left side, and reduction of the spondylolisthesis.
He had complete resolution of his low-back pain and radiating symptoms and mild improvement in his dorsiflexion weakness. The primary ICD-9-CM diagnosis for this admission was 721.3 (lumbosacral spondylolisthesis without myelopathy). There were no other secondary diagnoses that were pertinent to the lumbar spine and obviously no justification for the fusion (Table 3). The surgical indication for fusion (spondylolisthesis with instability) was clearly noted in the office note, operative note, and radiology reports. This case furthers the concern that surgical indications are not captured reliably in the NIS database.

Discussion

Health care policy research is increasingly reliant on large administrative databases. There are unique advantages that large databases provide, including large, population-based studies of surgical trends, demographic characteristics, and major complications of common operations. The MedPAR database, for example, captures 100% of Medicare hospital claims and has unique patient identifiers that allow for linkage among files and identification of repeat hospitalizations. Similarly, the NIS database is the largest publicly available all-payer inpatient care database in the United States. It currently contains data from approximately 8 million hospital stays each year, drawing from 1045 hospitals located in 46 states, approximating a 20% stratified sample of United States hospitals (http://www.hcup-us.ahrq.gov/nisoverview.jsp). The NIS likewise includes hospital identifiers that permit linkages for long-term analysis. Researchers and policymakers often use the NIS to identify, track, and analyze national trends in health care utilization, charges, quality, and outcomes.

Both of these large databases use ICD-9-CM procedure and diagnosis codes to identify and categorize surgeries and hospitalizations. The ICD-9-CM is the official system of assigning codes to diagnoses and procedures associated with hospital utilization in the United States. Although designed primarily for billing purposes, the codes are now used in health-policy research because of their ubiquity and availability. Algorithms designed to mine large administrative databases by using ICD-9-CM codes to identify hospitalizations for mechanical low-back pain have been described. The conclusions based on studies employing these methods have been used in health care policy making and have far-reaching implications for patients, providers, insurance carriers, and society.

We have demonstrated that the primary diagnosis transcribed by the medical coder and submitted to the NIS does not accurately describe the indication for fusion in degenerative conditions of the lumbar spine. In 48% of cases that we analyzed, the primary diagnosis was not the indication for fusion. We would like to emphasize that this finding is not necessarily evidence of a broken system, but rather in many cases is simply a reflection of the simple fact that the primary code was never intended to be the indication for fusion. It would be logical for the primary code to reflect the primary indication for surgery, again not the indication for fusion, but even this goal does not receive significant emphasis.

As an example of a scenario in which the primary code does not reflect the indication for surgery even when the system is working perfectly, consider the case of a patient with lumbar stenosis and spondylolisthesis. The indication for fusion is neurogenic claudication, and therefore it would make sense to find lumbar stenosis as the primary code. Of course spondylolisthesis and instability are the indication for fusion, but these radiographic findings might well be asymptomatic and on their own would never be an indication for surgery, so it makes sense to find those codes listed as secondary. Nowhere in the coding guidelines is there the suggestion that the primary diagnosis must be the surrogate for fusion.

Assignment of the primary diagnosis followed standardized ICD-9-CM guidelines; namely, the circumstance of the inpatient admission always governs the selection of the primary diagnosis, and that diagnosis must be chiefly responsible for occasioning the admission of the patient to the hospital for care. Furthermore, in that particular example, the diagnosis chiefly responsible for occasioning admission for surgery is, in fact, lumbar stenosis. In circumstances where two or more interrelated conditions each meets criteria for primary diagnosis, there is no hierarchy or prioritization for one code over another.
NIS database does not reflect indications for fusion

The aim of the present study was not to assign blame to any particular party in examining the decay of data fidelity. An analysis of sources of error in this particular data set is the subject of another report in preparation. That being said, error in data fidelity can be broken down into 3 broad categories: errors that relate to the coder, the surgeon, and limitations of the classification system itself (ICD-9-CM). An example of coder error, for instance, includes a lack of familiarity with domain-specific terminology, such as the differences between spondylosis, spondylolysis, and spondylolisthesis.

Data corruption can also stem from a surgeon’s error. Surgeons often do not realize that the diagnosis hastily and perfunctorily marked on outpatient forms tends to follow patients throughout their hospitalization, and this diagnosis is very likely to end up as the primary diagnosis recorded in a patient’s medical record. In addition, operative reports tend to focus more on the details of surgery than on the supporting documentation regarding the indications for surgery. Vague diagnoses such as “instability” or “lumbar spondylosis” are frequently provided. Furthermore, discharge summaries are often penned by interns, residents, and physician extenders who have virtually no background in spine surgery and little knowledge of the surgical procedure or its indications. Because physicians often treat many patients simultaneously and have heavy workloads, the time and attention they dedicate to checking the accuracy of the codes vary tremendously.

Finally, another potential source of error stems from the complexity and ambiguity inherent in the ICD-9-CM classification system, which has several serious limitations. One difficulty is that many commonly used diagnostic terms have continued to evolve since the ICD-9-CM code was defined.7–9 Furthermore, the codes do not capture disease severity, prior treatments, or the neurological status of the patient.10 This lack of specificity, which may be inherent in any standardized coding system, limits the codes’ capacity for capturing surgical indications. By way of example, patients with a primary disc herniation, recurrent disc herniation, and twice-recurrent disc herniation may all have the same primary code. The care of these patients, type of operation, difficulty of operation, and range of complications related to that surgery, however, are vastly different. A patient presenting with a twice-recurrent disc herniation may benefit from a fusion at that level because of iatrogenic instability, but the ICD-9-CM does not contain a code to express this indication. Another example is the diagnosis of lumbar stenosis, which in and of itself does not indicate a need for fusion. However, in many cases in which adequate decompression would result in instability, there is again no code in the ICD-9-CM vocabulary to reflect the surgical indication for fusion. Furthermore, the stringent inclusion criteria and patient selection process described by Fritzell et al.11,12 in the Swedish trial of lumbar fusion for low-back pain/disc degeneration, a particularly controversial indication, cannot be expressed or captured in the ICD-9-CM system. While data validity and diagnosis fidelity in highly regulated clinical trials10,11 with active auditing systems may be excellent, this is not necessarily true in routine clinical practice. Frank errors due to incorrect or imprecise diagnostic coding are routinely passed on in the NIS or other large administrative databases.13,17–19

Finally, there is mounting evidence that ICD-9-CM diagnostic categories contain significantly heterogeneous patient populations. For example, within the diagnosis of DDD, admittedly a poor indication for fusion, Glassman et al. have established a subset of patients who experienced excellent outcomes after lumbar fusion.23 Unfortunately, the ICD-9-CM diagnostic categories are not capable of distinguishing this subset from the category as a whole.

Specific diagnostic shortcomings of ICD-9-CM notwithstanding, the underlying assumption that the ICD-9-CM diagnoses accurately reflect the actual medical record has never been demonstrated. In the original publication describing the use of the ICD-9-CM to identify hospitalizations for mechanical low-back pain from large administrative databases, the authors cautioned that the only way to ascertain this accuracy is “to abstract medical record data for a representative sample of the hospitalizations included in the study.”3 As a hospital that contributes to the NIS, this is precisely what we sought to do. In this series, the primary diagnosis did not accurately describe the indication for fusion in degenerative conditions of the lumbar spine.

The consequences of this finding are far reaching. Deyo and colleagues4–6 have critically examined the increasing rates of lumbar fusion for degenerative spinal disorders. Their analyses and conclusions were driven by subcategorization of these disorders into different groups based solely on the primary diagnosis code entered into the NIS.3 As our results indicate, the correct primary diagnostic code applicable to the fusion procedure was correct only 48% of the time. If the primary diagnosis codes obtained from the NIS are incorrect more than 50% of the time, the force of Deyo and colleagues’ conclusions is called into question. To gauge trends of indications for fusion more accurately, at the very least one would need to take into account the secondary diagnoses. Given that the secondary diagnoses are so readily available, it is disingenuous to consider only primary diagnosis codes in a critique of surgical indications. Table 4 summarizes the ICD-9-CM codes that we feel unambiguously support the indication for fusion.

The net effect of coding errors on the analysis of information obtained from administrative databases is unpredictable.3,9 Even a small degree of misclassification will have potent effects when studying large databases, but errors of the magnitude described in the present series make conclusions regarding lumbar fusion indications extremely tenuous. Relationships between diagnoses, procedures, complications, and outcomes are weakened, and as a result, the conclusions derived from studying flawed databases may be inaccurate. Policy recommendations based on such weak footing should be regarded with skepticism.

Our study has certain limitations that may prohibit its results from generalizability and application to the national arena. First and foremost, our data are gathered from a single department, at a single academic center, of one of approximately 1000 that contribute to the NIS, over a relatively short time frame. However, the type of investigation that this study required—review of the actual medical record rather than the discharge abstract and discharge codes—cannot be done on behalf of out-
Moreover, our patient population may be different from those commonly seen by our orthopedic spine colleagues or in community-based practices. Alternatively, our indications for fusion may be narrower than those used in other practices. Another limitation is the fact that our hospital does not provide for an algorithm or hierarchy of ICD-9-CM codes, and the diagnosis chosen as primary versus secondary is arbitrary. This coding practice may be different at other institutions or in centers using commercially available coding programs in lieu of hospital-based coders. Theoretically, the low rates of data fidelity that we have identified at our institution may be specific to our hospital alone and are not relevant to other institutions. However, we have no reason to believe that this is true. The types of coder, surgeon, and ICD-related errors described in this report are not unique to our hospital and are just as likely at other institutions. One of the most compelling points is, as emphasized earlier, that even when no errors occur, the primary diagnosis is a poor surrogate for the surgeon’s indication for fusion.

**Conclusions**

Undoubtedly, there are unique and powerful advantages that large databases provide such as allowing for large, population-based studies of surgical trends, including rates, underlying medical conditions, demographic characteristics, and various safety and outcome measures. These pooled data sets also help to illustrate variations on a local, regional, and national level to benchmark performance. There are, however, substantial limitations in their use in clinical and health-policy research due to the poor fidelity of the source data. We have shown here that the primary ICD-9-CM diagnosis code transcribed by the medical coder is not an accurate surrogate for indication for fusion in over half the cases of degenerative conditions of the lumbar spine. This finding calls into serious question published reports indicating that spinal fusion procedures are done predominantly in “unindicated” cases, and it certainly weakens any relationship between primary indication for fusion, outcomes, and complications reported. Errors in coding are amplified in large databases, and conclusions reached by utilizing these databases may be inherently flawed.

The underlying causes of these data corruptions are numerous and cannot be traced to a single source. Coders are not specialty specific and have varied different services they cover. The nuances of coding coupled with the multitude of possible ICD-9-CM codes that sound similar but carry different import regarding the appropri-
ate use of surgical fusion cannot be overstated. Inadequate documentation by physicians exacerbates these errors. There are numerous points of data input within these episodes of care starting from the exam room, through booking and scheduling processes, admission to the hospital, and then the final assignment of diagnosis codes where “errors” can be carried through, leading to erroneous reporting. Further research is needed to improve these coding “handoffs,” thereby minimizing the “garbage in/garbage out” decay of fidelity.

We have shown that all secondary diagnoses should be investigated when assessing indications for fusion. While data validity in well-regulated clinical trials with active auditing systems may be high, this is not necessarily true in routine clinical practice despite the best of intentions of all involved. Researchers, policymakers, payers, and physicians should note these limitations when reviewing studies using hospital claims data. Critical analysis of these data sets needs be augmented with auditing methodologies or sensitivity analyses. Unfortunately these added layers are cumbersome, expensive, and difficult to implement on a large scale. However, at a time when health care policymakers and insurance carriers are scrutinizing spine surgeons—largely in routine clinical practice despite the best of intentions of all involved. Researchers, policymakers, payers, and physicians should note these limitations when reviewing studies

Acknowledgments

We would like to acknowledge Meredith Olson Wallace, MPH, Senior Consultant, Brigham and Women’s Hospital Center for Clinical Excellence, for her assistance with generating the discharge data for this patient cohort. We thank Cynthia Vieira, CCS, Partners Healthcare Trainer Coding Manager, for her help clarifying the coding scheme used at our institution. Finally, we express gratitude to Sarah DiZoglio-Macrina, Billing Manager for the Department of Neurosurgery, for her assistance in identifying the patients studied in this manuscript.

Disclosure

Dr. Groff reports being a consultant for Biomet and DePuy Spine. Dr. Chi reports being a consultant for DePuy Synthes.

Author contributions to the study and manuscript preparation include the following. Acquisition of data: Golgorsky. Analysis and interpretation of data: Groff, Golgorsky. Drafting the article: Groff, Golgorsky, Knightly. Critical revising the article: Groff, Golgorsky, Knightly. Reviewed submitted version of manuscript: all authors. Approved the final version of the manuscript on behalf of all authors: Groff. Statistical analysis: Golgorsky. Study supervision: Groff.

References


Manuscript submitted December 5, 2013. Accepted August 29, 2014. Please include this information when citing this paper: published online October 17, 2014; DOI: 10.3171/2014.8.SPINE131113. Address correspondence to: Michael W. Groff, M.D., Department of Neurosurgery, Brigham and Women’s Hospital, 75 Francis St., Boston, MA 02115. email: mgroff@partners.org.