

Improving discharge data fidelity for use in large administrative databases

YAKOV GOLOGORSKY, M.D.,¹ JOHN J. KNIGHTLY, M.D.,² YI LU, M.D., PH.D.,¹
JOHN H. CHI, M.D., M.P.H.,¹ AND MICHAEL W. GROFF, M.D.¹

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

Object. Large administrative databases have assumed a major role in population-based studies examining health care delivery. Lumbar fusion surgeries specifically have been scrutinized for rising rates coupled with ill-defined indications for fusion such as stenosis and spondylosis. Administrative databases classify cases with the *International Classification of Diseases, Ninth Revision, Clinical Modification* (ICD-9-CM). The ICD-9-CM discharge codes are not designated 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 first sought to compare the ICD-9-CM code(s) assigned by the medical coder according to the surgeon's indication based on a review of the medical chart, and then to elucidate barriers to data fidelity.

Methods. A retrospective review was undertaken of all lumbar fusions performed in the Department of Neurosurgery at the authors' institution between August 1, 2011, and August 31, 2013. Based on this review, the indication for fusion in each case was categorized as follows: spondylolisthesis, deformity, tumor, infection, nonpathological fracture, pseudarthrosis, adjacent-level degeneration, stenosis, degenerative disc disease, or disc herniation. These surgeon diagnoses were compared with the primary ICD-9-CM codes that were generated by the medical coders and submitted to administrative databases. A follow-up interview with the hospital's coders and coding manager was undertaken to review causes of error and suggestions for future improvement in data fidelity.

Results. There were 178 lumbar fusion operations performed in the course of 170 hospital admissions. There were 44 hospitalizations in which fusion was performed for tumor, infection, or nonpathological fracture. Of these, the primary diagnosis matched the surgical indication for fusion in 98% of cases. The remaining 126 hospitalizations were for degenerative diseases, and of these, the primary ICD-9-CM diagnosis matched the surgeon's diagnosis in only 61 (48%) of 126 cases of degenerative disease. When both the primary and all secondary ICD-9-CM diagnoses were considered, the indication for fusion was identified in 100 (79%) of 126 cases. Still, in 21% of hospitalizations, the coder did not identify the surgical diagnosis, which was in fact present in the chart. There are many different causes of coding inaccuracy and data corruption. They include factors related to the quality of documentation by the physicians, coder training and experience, and ICD code ambiguity.

Conclusions. Researchers, policymakers, payers, and physicians should note these limitations when reviewing studies in which hospital claims data are used. Advanced domain-specific coder training, increased attention to detail and utilization of ICD-9-CM diagnoses by the surgeon, and improved direction from the surgeon to the coder may augment data fidelity and minimize coding errors. By understanding sources of error, users of these large databases can evaluate their limitations and make more useful decisions based on them.

(<http://thejns.org/doi/abs/10.3171/2014.3.FOCUS1459>)

KEY WORDS • administrative database • data fidelity • ICD-9-CM • diagnosis code • lumbar fusion • accuracy

THE use of diagnosis codes from the International Classification of Diseases (ICD) has been expanded from its original purpose of classifying morbidity and mortality information for statistical pur-

poses to diverse sets of applications in health research, health care policy, and health care finance.¹⁷ Currently in its ninth iteration, the *International Classification of Diseases, Ninth Revision, Clinical Modification* (ICD-9-CM) contains more than 12,000 diagnosis codes. Discharge codes assigned through the medical coding process are probably the most powerful descriptors of the patient's hospital course once the full description of the medical record has been left behind. Because medical coding serves as an important nexus between the primary data sources and the subsequent data usage, inaccuracies introduced by

Abbreviations used in this paper: ALD = adjacent-level degeneration; BMI = body mass index; DDD = degenerative disc disease; DRG = diagnosis-related group; ICD-9-CM = *International Classification of Diseases, Ninth Revision, Clinical Modification*; MedPAR = Medicare Provider Analysis and Review; NIS = Nationwide Inpatient Sample.

low-quality coding will of necessity confound any secondary analysis.²¹ Increasing data quality at the coding level will conversely result in improving data fidelity in downstream usages.

There are many potential sources of error interposed between the surgeon's diagnosis (accepted as the gold standard) and the nosological diagnosis code arrived at by the medical coder. We will focus this paper on the typical process of elective lumbar fusion operations for degenerative diseases. The data trail begins in an outpatient physician-patient interaction, after which the physician chooses appropriate ICD codes for the patient's relevant diagnosis. Often, this diagnosis is required for insurance precertification for outpatient radiological tests or procedures, such as MRI examinations or epidural steroid injections. The diagnosis code(s) that the surgeon enters after the first patient interaction potentially stays with the patient throughout the interval of care. When an operation is planned, the Current Procedural Terminology (CPT) procedure codes defined by the American Medical Association are added to the ICD diagnosis codes. As the patient progresses through the hospitalization, additional sources of patient information such as the history and physical, progress notes, operative report, radiological reports, and eventually the discharge summary are added to the medical record. After discharge, the entire medical chart is transferred to medical records.

The processing of information in medical records, which is then entered into administrative 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. In our hospital, 1 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.^{6,17} Of necessity, the rich data set found in the patient record that prompts the assignment of the ICD-9-CM codes is not available to researchers studying large administrative databases, and therefore, researchers studying these large databases perpetuate any errors that were created at the coding level.

Large administrative databases have assumed a major role in population-based studies examining health care delivery.^{6-8,16} Two of the largest include the Medicare Provider Analysis and Review (MedPAR) database and the Nationwide Inpatient Sample ([NIS] <http://www.hcup-us.ahrq.gov/nisoverview.jsp>).⁴ The MedPAR database includes 100% of Medicare hospital claims, whereas the NIS is a component of the Healthcare Cost and Utilization Project based at the federal Agency for Healthcare Research and Quality. The NIS collects data from the states, which receive data from individual hospitals regarding 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. Thus, this database has become progressively more representative of the national population. Because the ICD-9-CM system was not designed for research purposes, it may not be sufficient for understanding health care policy research,

and in particular indications for procedures.¹⁹ Despite this limitation, numerous studies in the past 2 decades have exploited the NIS and MedPAR databases not only to document rising rates of lumbar fusion, but also to demonstrate trends for specific lumbar diagnoses.³ Because 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, and all sources of data corruption must be identified to mitigate the "garbage in, garbage out" phenomenon.

As practitioners at a hospital that contributes to the NIS, we examined the quality of the data that are submitted to the database by the medical coders, and compared it to the information in the medical record. The aim of this study was multifold: first, 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 (Y.G.) with domain-specific knowledge in the area of lumbar fusion. By contrast, the trained medical coders do not have domain-specific training and focus their efforts on a smaller subset of the medical record. Next, we wanted to clarify how often any of the secondary ICD-9-CM codes are in agreement with the surgical indication. Finally, we wished to elucidate the main error sources during the ICD diagnostic coding process from patient admission to diagnostic code assignment.

Methods

This study retrospectively examines the demographic, diagnostic, and coder-related data in 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 2 of the authors (J.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 diagnoses for the remaining cases were obtained, and the medical charts were reviewed by a fellowship-trained spine surgeon with no involvement in the cases (Y.G.). Admissions for multiple lumbar fusion operations were counted only once, because only one discharge abstract was compiled per hospitalization.

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 (isthmus and degenerative), deformity (coronal or sagittal), tumor, infection, non-pathological fracture (trauma), pseudarthrosis, adjacent-level degeneration (ALD) after prior lumbar fusion, stenosis, or degenerative disc disease (DDD). The stenosis category included the presence of central or foraminal stenosis without any of the other diagnoses listed above, or patients in whom adequate decompression would result in iatrogenic instability (that is, 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

Improving data fidelity in administrative databases

single-level DDD in whom the indication for fusion approximated the criteria outlined by the Swedish Trial.^{10,11} Namely, these patients must have 1) pain duration for at least 1 year; 2) back pain more pronounced than leg pain; 3) high disability scores on the EuroQol-5 Dimensions metric; 4) and/or be on work leave/disability; and 5) degenerative changes only at L4-5 or L5-S1 on CT and/or MRI studies.

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 in the "spondylolisthesis" rather than "spinal stenosis" category. In cases where the surgical diagnosis was never identified, we attempted to identify the source of the error. After all hospitalization discharge data were reviewed, a structured interview with 4 of the coders and the coding manager was undertaken. Coder-related data collected included the number of coders, their coding credentials, experience level, specialized training (domain-specific), information regarding which parts of the hospital chart they have access to, their algorithm for deconstructing a hospital chart and selecting primary versus secondary diagnoses, and which resources are available to them in case of uncertainty. Approval to conduct this study was obtained from the Institutional Review Board of the Brigham and Women's Hospital.

Results

Between August 1, 2011, and August 31, 2013, 168 patients underwent 178 lumbar fusions in the course of 170 separate hospitalizations. One patient had 3 separate hospitalizations for lumbar fusion: an index case performed for spondylolisthesis, and 2 subsequent operations for pseudarthrosis and hardware failure. There were 76 men and 92 women, with a median age of 57.5 years (range 21-96 years), and a median BMI of 28.5 (range 15.8-51.4). Thirty-three (19.4%) were active smokers at the time of surgery. The median number of levels treated was 3 (range 2-9). A summary of demographic data is presented in Table 1; medians rather than means are presented because several subcategories have few patients.

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 ALD (n = 2). Thus, of the 170 admissions, 44 were for fusion relating to tumor, infection, or nonpathological fracture, and 126 were for degenerative diagnoses.

In 98% of cases whose indication for fusion related to tumor, infection, or nonpathological fracture, the primary ICD-9-CM discharge code accurately captured the surgical indication. However, in the remaining 126 cases for which the ICD-9-CM diagnosis is considered to be degenerative in nature, the primary ICD-9-CM code accurately reflected the surgical indication in only 61 (48.4%)

of cases. This finding may suggest that diagnoses that are more objective such as tumor, infection, or fracture have less coder interpretation error than those with more nuanced and subjective diagnoses associated with degenerative conditions. When the secondary codes in the discharge abstract were also considered, the likelihood of finding the surgical diagnosis increased significantly but remained disappointingly low (79.4%). Even when all diagnoses were considered, medical coders did not identify the surgical indication for fusion in 26 (20.6%) of 126 cases (Table 2).

During this study period, our institution employed 12 coders who were responsible for classifying inpatient hospitalizations. Coding was done remotely with coders spread geographically across the US, but the coding manager was located on-site. Coders were responsible for all inpatient stays at our hospital, and were not disease- or department-specific (that is, department of neurosurgery coders or spine coders). None had advanced domain training or particular expertise in spine operations. In situations in which the coder needed assistance to resolve vague or conflicting information in the medical chart, that individual had access to 2 designated Data Quality Specialists for assistance. In addition, a corporate coder trainer is available for consultation. At their discretion, coders are free to contact the physician-author of any clinical report or note in the chart. In general, coders are loath to contact the surgeon in cases of diagnosis ambiguity, and in fact neither of the senior surgeons has ever been contacted by the coders.

The American Health Information Management Association (AHIMA), the governing body for health information professionals, designates 2 types of certification: R.H.I.T. (Registered Health Information Technician) or R.H.I.A. (Registered Health Information Administrator). The R.H.I.T. designation is an associate's degree program, whereas the R.H.I.A. designation is a bachelor's degree program. Additionally, the C.C.S. (Certified Coding Specialist) credential is earned after someone takes a 9-month coding course and passes a credentialing examination, and then maintains yearly continuing education credits, which is most commonly done after achieving a bachelor's degree in an unrelated field. Eleven of our 12 coders have the C.C.S. credential, and 2 have both the R.H.I.T. and C.C.S. credentials. The coders' median experience was 21 years, and the range was 10-27 years.

Coding at our institution is performed within 5 days of discharge. At the time of the study, all coding was done manually. More recently, and in anticipation of ICD-10, our institution has begun moving to computer-assisted coding, in which a computer scans the medical chart for key terms and makes coding suggestions. The coder must then validate the codes identified, prioritize them, and assign appropriate codes.

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 studies 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 acknowledgment of spondylolisthesis by

TABLE 1: Summary of demographic data in 170 admissions for lumbar fusion*

Surgeon's Dx	No.	Median Age (yrs)	Median BMI (kg/m ²)	Median No. of Levels	% Male	% Smokers
spondylolisthesis	82	63	28.6	2	39.0	17.1
tumor	20	53.5	25.4	5	60	5
nonpathological Fx (trauma)	19	53	22.7	5	63.2	31.6
DDD	18	45	29.3	2	50	33.3
deformity	11	59	29.5	4	27.3	9.1
pseudarthrosis	9	55	32.0	3	66.7	22.2
infection	5	54	27.6	3	40	40
severe stenosis	4	67	28.8	2.5	0	25
ALD	2	61	30.75	2	100	0
total, range	170	57.5, 21–96	28.5, 15.8–51.4	3, 2–9	45.9	19.4

* Dx = diagnosis; Fx = fracture.

the surgical team, the coders cannot use this information in their coding determination. The algorithm for deconstructing the hospital record follows a stereotyped progression, and most coders begin with the discharge summary, corroborate codes in the operative report, and scan daily progress notes.

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. In circumstances where two or more interrelated conditions each meet criteria for primary diagnosis, there is no hierarchy or prioritization for one code over another. For example, a patient who suffers from lumbar stenosis and degenerative spondylolisthesis can have either diagnosis as primary. In general, the primary code is most commonly derived from the discharge summary.

A structured interview was undertaken with 4 coders and the coding manager, and the medical records of all

cases in which the surgical diagnosis was never identified were thoroughly reviewed. In these 26 cases, the error was identified as the fault of the surgeon in 3 cases (that is, never mentioning the indication for fusion in the operative report or office visit—Fig. 1); the fault of the coder in 3 cases (the indication was clearly shown, but the coder failed to code it—Fig. 2); and combined responsibility was identified in the vast majority of the cases (20 of 26). In 12 other cases, the words “instability” or “destabilizing” were used by the surgeon in the operative report in the sections for diagnosis or indications. There is no corresponding ICD-9-CM code, with the possible exception of 724.6 (Disorders of sacrum [including instability of lumbosacral joint]). In 4 other instances, the diagnosis of spondylolisthesis was mentioned but buried deep within the operative report or in a preoperative office note, both of which were available to the coder, but not in a conspicuous location. Although these are technically errors on the coder's part, we assumed that the surgeon was equally at fault for not bringing this diagnosis, the actual indication for fusion, to the coder's attention. In the last 4 cases, error

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

Surgeon's Dx	No.	Primary ICD-9-CM Code Captures Surgeon's Dx	Secondary ICD-9-CM Code Captures Surgeon's Dx	Surgeon's Dx Not Captured
nondegenerative disease				
tumor	20	20/20	0/20	0/20
nonpathological Fx (trauma)	19	19/19	0/19	0/19
infection	5	4/5	1/5	0/5
subtotal	44	43/44 (97.7%)	1/44 (2.3%)	0/44 (0%)
degenerative disease				
spondylolisthesis	82	29/82	34/82	19/82
DDD	18	18/18	0/18	0/18
deformity	11	4/11	5/11	2/11
pseudarthrosis	9	9/9	0/9	0/9
severe stenosis	4	1/4	0/4	3/4
ALD	2	0/2	0/2	2/2
subtotal	126	61/126 (48.4%)	39/126 (31%)	26/126 (20.6%)

Improving data fidelity in administrative databases

Partners HealthCare System, Inc.
BRIGHAM AND WOMENS HOSPITAL
A Teaching Affiliate of Harvard Medical School
75 Francis Street, Boston, MA 02115

MRN: [REDACTED] (BWH)
[REDACTED]
Date of Birth: [REDACTED]
Age: 60 yrs. Sex: F

Operative Report

Date of Operation: [REDACTED] Report Status: Signed

SURGEON: [REDACTED] M.D.

ASSISTANT:
[REDACTED] MD

PREOPERATIVE DIAGNOSIS:
L4 through S1 spondylosis.

POSTOPERATIVE DIAGNOSIS:
L4 through S1 spondylosis.

OPERATION:
L4 through S1 laminectomy and medial facetectomy, posterior lumbar interbody fusion, bone marrow aspirate, pedicle screw fixation segmental.

ANESTHESIA:
General.

ESTIMATED BLOOD LOSS:
500 mL.

COMPLICATIONS:
None.

INDICATIONS FOR PROCEDURE:
This 59-year-old woman had a longstanding history of neurogenic claudication and axial back pain. Conservative therapy and activity modifications had been unsuccessful in ameliorating her symptoms. She was taken to the operating room with informed consent.

DESCRIPTION OF PROCEDURE:
The patient was taken to the OR and while on the stretcher, vascular access was secured and general endotracheal anesthesia

Fig. 1. Representative example of a poorly written operative report. Note that the surgeon uses vague diagnosis terms such as "spondylosis" (underlined). There is no ICD-9-CM correlate to the term "axial back pain" used in the "Indications for Procedure" section, and the fact that this patient suffered from demonstrated spondylolisthesis and dynamic instability on flexion-extension radiographs was not mentioned. The primary discharge code for this hospitalization was 724.03 (spinal stenosis, lumbar region, with neurogenic claudication).

was attributed to a combination of inadequacy of the ICD-9-CM diagnosis codes, insufficient surgeon's explanation, and coder error. For example, there is no ICD-9-CM code to describe a patient with a twice-recurrent disc herniation at a particular level, or a code to describe ALD after prior lumbar fusion.

Discussion

The use of ICD diagnosis codes has greatly expanded from its original purpose of classifying morbidity and mortality information for statistical purposes to a diverse set of applications in health research, health care policy, and health care finance.¹⁷ Codes assigned through the medical coding process are fundamental to processes of health services research and methods of quality improvement. Because medical coding serves as an important nexus between the primary data sources and many of their secondary data usages, inaccuracy or variation present in low-quality coding will detract from the quality of such secondary use.²¹ Health care policy research has become increasingly reliant on large administrative databases (<http://www.hcup-us.ahrq.gov/nisoverview.jsp>) that use the ICD-9-CM system, such as the NIS,⁴⁻⁸ because these databases allow researchers to identify, track, and analyze national trends in health care utilization, charges, quality, and outcomes.^{2-8,16,19}

Increased attention to code accuracy has occurred

both as a result of the application of ICD codes for purposes other than those for which the classifications were originally designed as well as because of their widespread use for making important funding, clinical, and research decisions.^{17,21} In the 1980s the prospective payment system using diagnosis-related groups (DRGs) was implemented, and increased scrutiny of diagnostic accuracy began.²¹ Hsia et al. reported in the *New England Journal of Medicine* that ICD-9-CM diagnostic code inaccuracy sufficient to change the hospitalization's DRG was approximately 20%.¹⁴ Lloyd and Rissing examined physician and coding errors in the medical records of 5 Veterans Administration hospitals, and found 22% frank diagnosis error. They identified 3 sources of error: physician (62%), coder (35%), and keypunch (3%). The authors projected that there were 0.81 coding errors in the average abstract. If the errors were corrected in the abstracts, it would change 19% of records for DRG purposes.¹⁵ Studies in the 1990s found rates similar to those of the 1980 studies, with error rates ranging from 0% to 70%, but most falling between 20% and 50%.¹⁷ The wide variation in error rates is due largely to differences across study methods and to the many different sources of errors that influence code accuracy.^{1,13,17}

O'Malley et al. reported the most systematic and extensive analysis of ICD-9-CM coding inaccuracy. They examined potential sources of errors at each step of the inpatient ICD-9-CM coding process. They found that multiple factors contribute to the inaccuracy of ICD-9-

Partners HealthCare System, Inc.
BRIGHAM AND WOMENS HOSPITAL
A Teaching Affiliate of Harvard Medical School
75 Francis Street, Boston, MA 02115

MRN: [REDACTED] (BWH)
Date of Birth: [REDACTED]
Age: 65 yrs. Sex: M

Operative Report

Date of Operation: [REDACTED] Report Status: Signed

SURGEON: [REDACTED] M.D.

ASSISTANT:
[REDACTED] M.D., Neurosurgery resident.

PREOPERATIVE DIAGNOSIS:
Grade I spondylolisthesis with lumbar stenosis, L4-L5.

POSTOPERATIVE DIAGNOSIS:
Grade I spondylolisthesis with lumbar stenosis, L4-L5.

OPERATION:
L4-L5 transforaminal lumbar interbody fusion, left-sided approach, L4-L5 posterior spinal instrumentation, L4-L5 posterior spinal arthrodesis and placement of the L4-L5 intervertebral biomechanical cage, L4-L5 left sided hemilaminotomy for decompression.

ANESTHESIA:
General.

INDICATIONS FOR PROCEDURE:
This is a 65-year-old male with progressive claudication symptoms with sciatic pain with standing and walking with significant lumbar stenosis at L4-L5 with a grade I spondylolisthesis with instability on flexion radiographs. Informed consent was obtained for the procedure above and the patient was brought to the operating room on the date above.

DESCRIPTION OF PROCEDURE:
After anesthesia was induced, the patient was positioned prone on

Fig. 2. Representative example of a well-written operative report that highlights spondylolisthesis prior to lumbar stenosis (underlined). Note that the vague term “instability” is used in the “Indications for Procedure” section, but is immediately preceded by the more precise term of “spondylolisthesis.” Despite this, the coder failed to identify spondylolisthesis as a discharge code. The primary discharge code for this hospitalization was 721.3 (lumbosacral spondylosis without myelopathy).

CM coding, including amount and quality of information at admission, communication among patients and providers, the clinician’s knowledge and experiences with the illness, the clinician’s attention to details, variance in the electronic and written records, coder training and experience, facility quality-control efforts, and unintentional and intentional coder errors (for example, misspecification, unbundling, missequencing, and upcoding).¹⁷

With the adoption of the far more detailed and complex ICD-10 system, coding inaccuracy persists. Gibson and Bridgman assessed the accuracy of diagnostic coding performed using the ICD-10 in general surgery by comparing codes ascribed by hospital coders to codes ascribed by expert external coders. They found errors of coding in 29%, of which 8% were at the most serious level (that is, wrong ICD-10 chapter). They reported that 78% of errors occurred between the outpatient medical record and the admission form, and that 29% of records had inaccurate diagnostic codes.¹² This study highlights the significance of codes assigned by the surgeon before the admission has even started.

Our series is very much in line with the rates of coding error mentioned above. Less than 50% of lumbar fusions carried the correct primary diagnosis. When all secondary diagnoses were considered, an error of 20% remained. In our series, sequencing error was the most common one identified. Sequencing error occurs when two interrelated diagnoses are both listed, but the manifestation of the primary disorder is placed as the primary instead of a secondary diagnosis.^{17,18} For example, a patient has respiratory failure as a manifestation of congestive heart failure. The

congestive heart failure should be the principal diagnosis and the respiratory failure the secondary diagnosis. In our series, lumbar stenosis was often listed as the primary diagnosis, and spondylolisthesis as a secondary code. In fact, it seems that the lumbar stenosis is a manifestation of the spondylolisthesis, and this missequencing is especially salient when researchers mining large administrative databases study only primary codes as surrogates for indication for fusion surgery. Most sequencing errors are not intentional and may comprise the commonest kind of errors in hospital discharge abstracts,^{15,17} a finding that we found in our series as well.

Another major source of error was coder ignorance of complex domain-specific terminology. In particular, the triad of spondylosis, spondylolysis, and spondylolisthesis was highlighted by the coders in an interview, as well as the vague term “instability.” Although all of our coders had more than 10 years of experience in coding, none had domain-specific advanced training, and all coded for the entire hospital rather than for our department exclusively. Finally, preadmission diagnosis codes, originally assigned by the surgeon in the outpatient setting, were more likely to end up as the primary discharge code than were other codes identified from the medical record by the coders. A plausible explanation for this finding is that the original note and diagnosis code were later largely copied by the residents and physician extenders in the history and physical note, which itself was reproduced in the discharge summary, again penned by the residents and physician extenders. Because physicians treat many patients simultaneously and carry heavy workloads, the time and attention

Improving data fidelity in administrative databases

physicians dedicate to checking the accuracy of the codes varies tremendously.¹⁷

Another source of error stems from the complexity and ambiguity inherent in the ICD-9-CM system. Take for example 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 and colleagues^{10,11} 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. Although data validity and diagnosis fidelity in highly regulated clinical trials^{9,20} with active auditing systems may be excellent, this is not necessarily true in routine clinical practice.

Because ICD-9-CM discharge codes are not generated by surgeons, but are rather assigned by trained hospital medical coders, any data corruption on the coding level will be perpetuated in large administrative databases that collect discharge data, such as the NIS. The net effect of coding errors on the analysis of information obtained from administrative databases is unpredictable.^{7,8} Even a small degree of misclassification will have potent effects when studying large databases, but errors of the magnitude described in this 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.

There are several suggestions that can be implemented to improve coding accuracy. They can be broadly divided into either coder education or surgeon behavior modification. One idea would be to have advanced domain-specific training to improve coder familiarity with medical terminology. For example, coders can attend a workshop in which they are taught to delineate between spondylosis, spondylolysis, and spondylolisthesis, understand the significance of words such as pseudarthrosis or instability, and to focus their attention on the operative report sections for diagnosis and indications. In addition, having department-specific or spine-specific coders may improve coding accuracy because of increased coder familiarity with technical terminology. Coding managers can instruct coders that in instances in which two or more interrelated codes each meet criteria for primary diagnosis, a hierarchy of coding may be instituted, and the code that justifies the “higher-level” procedure should be selected. For example, spondylolisthesis should be chosen ahead of lumbar stenosis (Table 3). Finally, with the advent of computer-assisted coding, scanning for key words like spondylolisthesis will be made simpler, and usage of appropriate diagnosis codes will be enhanced, especially with the adoption of the more complex ICD-10 system.

Surgeons will need to be educated on the significance of outpatient code selection, because the codes that they sometimes choose perfunctorily during the initial outpatient interaction may follow the patient over the entire

course of the illness. Vague codes like spondylosis and lumbago should be avoided in favor of more precise codes, especially if a surgical intervention is considered. If possible, for fusion surgery the primary code should be the indication for fusion. This can be stated explicitly in the operative report sections for diagnosis and indications or, if the surgeon is aware of the appropriate ICD code, it may be placed in the aforementioned sections. Care should be taken to list appropriate diagnoses in a hierarchical order that suggests the indication for fusion prior to one for surgery in general. Finally, the surgeon is urged to carefully review the discharge summary, which is often the foundation for the coder’s understanding of the admission. Ultimately, with the exception of rare clerical errors, the surgeon is responsible for the accuracy of data regarding the patient’s hospital stay.

Although in this paper we focused on errors influencing code accuracy, the goal was not to disparage ICD codes in general or the large administrative databases that use them. The ICD codes are invaluable tools for research, reimbursement, and policy making. Large administrative databases such as the NIS that use these codes provide powerful and unique advantages 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 act as a benchmark for performance. However, policy making and research conclusions made by studying these codes are improved when code accuracy is well understood and taken into account. By heightening their awareness of potential error sources, users can better evaluate the applicability and limitations of codes in their own context, and thus use ICD codes in optimal ways. With improved coding accuracy, the downstream usage of large administrative databases can take on greater significance and influence health care policy in a meaningful manner.

Conclusions

Large administrative databases have assumed a major role in population-based studies examining health care delivery. Health care policy is increasingly reliant on these databases for high-quality research data. It is well accepted that discharge codes generated by medical coders are considered one of the upstream points for data quality. Inaccuracy or variation present in low-quality coding will be carried forward to secondary use as in administrative databases. Errors that differentiate the ICD code from the true disease include both random and systematic measurement errors. Increasing data quality at the coding level will result in improving the data fidelity in downstream applications.

There are many different causes of coding inaccuracy and data corruption. They include factors related to quality of documentation by the physicians, factors related to coder training and experience, and factors that relate to ICD code ambiguity. There are numerous points of data input within these episodes of care, starting from the out-

TABLE 3: The ICD-9-CM codes for which fusion is generally indicated and which one should be chosen as the primary code

Broad Indication Category	ICD-9-CM Group or Code(s)	Description
tumor	140–239	all neoplasms
	733.1, 733.10, 733.13	pathological Fx
infection of spinal column	015.0	tuberculosis of vertebral column (Pott disease)
	324.1	intraspinous abscess
	730	osteomyelitis
Fx/trauma	805	Fx of vertebral column w/o mention of spinal cord injury
	806	Fx of vertebral column w/ spinal cord injury
	839	vertebral dislocations
instability	724.6	disorders of sacrum, including instability of lumbosacral joint
	738.4	acquired spondylolisthesis
	756.11	spondylolysis, lumbosacral region
	756.12	spondylolisthesis
deformity	737	curvature of spine
	738.5	other acquired deformity of back or spine
	754.2	congenital musculoskeletal deformities of spine
revision/complication	733.8	malunion & nonunion of Fx, pseudarthrosis
	996.4	mechanical complication of internal orthopedic device implant & graft
	996.6	infection & inflammatory reaction due to internal prosthetic device implant & graft
	996.7	other complications due to internal prosthetic device, implant, & graft

patient examination room, through booking and scheduling processes, admission to the hospital, to the final assignment of diagnosis codes, at which “errors” can be carried through, leading to erroneous reporting.

Researchers, policymakers, payers, and physicians should note these limitations when reviewing studies performed using hospital claims data. Critical analysis of these data sets needs to be augmented with auditing methodologies or sensitivity analyses. At the very least, researchers should consider all secondary codes when analyzing large administrative databases. Increased domain-specific coder training, the surgeon’s meticulous attention to detail and avoidance of vague diagnoses, and increased direction from the surgeon to the coder (for example, by using ICD-9-CD terminology in the medical record), may improve data fidelity and minimize coding error. By understanding sources of error, users can evaluate the limitations of the classifications and make more informed decisions based on them.

Acknowledgment

We thank and acknowledge Cynthia Vieira, C.C.S., Partners Healthcare Trainer Coding Manager, for her help clarifying the coding scheme used at our institution.

Disclosure

Dr. Chi is a consultant for DePuy Spine. Dr. Groff is a consultant for DePuy and for Biomet Spine.

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

References

1. Bossuyt PM, Reitsma JB, Bruns DE, Gatsonis CA, Glasziou PP, Irwig LM, et al: Towards complete and accurate reporting of studies of diagnostic accuracy: the STARD initiative. **BMJ** 326:41–44, 2003
2. Cherkin DC, Deyo RA, Volinn E, Loeser JD: Use of the International Classification of Diseases (ICD-9-CM) to identify hospitalizations for mechanical low back problems in administrative databases. **Spine (Phila Pa 1976)** 17:817–825, 1992
3. Deyo RA, Gray DT, Kreuter W, Mirza S, Martin BI: United States trends in lumbar fusion surgery for degenerative conditions. **Spine (Phila Pa 1976)** 30:1441–1447, 2005
4. Deyo RA, Mirza SK, Martin BI, Kreuter W, Goodman DC,

Improving data fidelity in administrative databases

- Jarvik JG: Trends, major medical complications, and charges associated with surgery for lumbar spinal stenosis in older adults. **JAMA** **303**:1259–1265, 2010
5. Deyo RA, Nachemson A, Mirza SK: Spinal-fusion surgery—the case for restraint. **N Engl J Med** **350**:722–726, 2004
 6. Faciszewski T: Spine update. Administrative databases in spine research. **Spine (Phila Pa 1976)** **22**:1270–1275, 1997
 7. Faciszewski T, Broste SK, Fardon D: Quality of data regarding diagnoses of spinal disorders in administrative databases. A multicenter study. **J Bone Joint Surg Am** **79**:1481–1488, 1997
 8. Faciszewski T, Jensen R, Berg RL: Procedural coding of spinal surgeries (CPT-4 versus ICD-9-CM) and decisions regarding standards: a multicenter study. **Spine (Phila Pa 1976)** **28**:502–507, 2003
 9. Fischgrund JS, Mackay M, Herkowitz HN, Brower R, Montgomery DM, Kurz LT: 1997 Volvo Award Winner in Clinical Studies. Degenerative lumbar spondylolisthesis with spinal stenosis: a prospective, randomized study comparing decompressive laminectomy and arthrodesis with and without spinal instrumentation. **Spine (Phila Pa 1976)** **22**:2807–2812, 1997
 10. Fritzell P, Hägg O, Jonsson D, Nordwall A, Swedish Lumbar Spine Study Group: Cost-effectiveness of lumbar fusion and nonsurgical treatment for chronic low back pain in the Swedish Lumbar Spine Study: a multicenter, randomized, controlled trial from the Swedish Lumbar Spine Study Group. **Spine (Phila Pa 1976)** **29**:421–434, 2004
 11. Fritzell P, Hägg O, Wessberg P, Nordwall A: 2001 Volvo Award Winner in Clinical Studies. Lumbar fusion versus nonsurgical treatment for chronic low back pain: a multicenter randomized controlled trial from the Swedish Lumbar Spine Study Group. **Spine (Phila Pa 1976)** **26**:2521–2534, 2001
 12. Gibson N, Bridgman SA: A novel method for the assessment of the accuracy of diagnostic codes in general surgery. **Ann R Coll Surg Engl** **80**:293–296, 1998
 13. Green J, Wintfeld N: How accurate are hospital discharge data for evaluating effectiveness of care? **Med Care** **31**:719–731, 1993
 14. Hsia DC, Krushat WM, Fagan AB, Tebbutt JA, Kusserow RP: Accuracy of diagnostic coding for Medicare patients under the prospective-payment system. **N Engl J Med** **318**:352–355, 1988
 15. Lloyd SS, Rissing JP: Physician and coding errors in patient records. **JAMA** **254**:1330–1336, 1985
 16. Martin BI, Mirza SK, Franklin GM, Lurie JD, MacKenzie TA, Deyo RA: Hospital and surgeon variation in complications and repeat surgery following incident lumbar fusion for common degenerative diagnoses. **Health Serv Res** **48**:1–25, 2013
 17. O'Malley KJ, Cook KF, Price MD, Wildes KR, Hurdle JF, Ashton CM: Measuring diagnoses: ICD code accuracy. **Health Serv Res** **40**:1620–1639, 2005
 18. Osborn CE: Benchmarking with national ICD-9-CM coded data. **J AHIMA** **70**:59–69, 1999
 19. Wang MC, Laud PW, Macias M, Nattinger AB: Strengths and limitations of International Classification of Disease Ninth Revision Clinical Modification codes in defining cervical spine surgery. **Spine (Phila Pa 1976)** **36**:E38–E44, 2011
 20. Weinstein JN, Lurie JD, Tosteson TD, Hanscom B, Tosteson AN, Blood EA, et al: Surgical versus nonsurgical treatment for lumbar degenerative spondylolisthesis. **N Engl J Med** **356**:2257–2270, 2007
 21. Zeng X, Bell PD: Determination of problematic ICD-9-CM subcategories for further study of coding performance: Delphi method. **Perspect Health Inf Manag** **8**:1b, 2011

Manuscript submitted February 14, 2014.

Accepted March 21, 2014.

Please include this information when citing this paper: DOI: 10.3171/2014.3.FOCUS1459.

Address correspondence to: Michael W. Groff, M.D., Department of Neurosurgery, Brigham and Women's Hospital, Harvard Medical School, 75 Francis St., Boston, MA 02115. email: mgroff@partners.org.