The Validity of Using Administrative Claims Data in Total Joint Arthroplasty Outcomes Research

Kevin J. Bozic, MD, MBA,*† Vanessa W. Chiu, MPH,*† Steven K. Takemoto, PhD,*† Jordan N. Greenbaum, MD, MBA,‡ Thomas M. Smith, DO,§ Seth A. Jerabek, MD,‡ and Daniel J. Berry, MD§

Abstract: The purpose of this study was to evaluate concordance between administrative and clinical diagnosis and procedure codes for revision total joint arthroplasty (TJA). Concordance between administrative and clinical records was determined for 764 consecutive revision TJA procedures from 4 hospitals. For revision total hip arthroplasty, concordance between clinical diagnoses and administrative claims was very good for dislocation, mechanical loosening, and periprosthetic joint infection (all $\kappa > 0.6$), but considerably lower for prosthetic implant failure/breakage and other mechanical complication (both $\kappa < 0.25$). Similarly, for revision total knee arthroplasty diagnoses, concordance was very good for periprosthetic fracture, periprosthetic joint infection, mechanical loosening, and osteolysis (all $\kappa > 0.60$), but much lower for implant failure/breakage and other mechanical complication (both $\kappa < 0.24$). Concordance for TJA-specific procedure codes was very good only for revision total knee arthroplasty patellar component revisions and tibial insert exchange procedures. Total (all-component) revisions were overcoded for hips (00.70) and undercoded for knees (00.80). Improved clinical documentation and continued education are needed to enhance the value of these codes. **Keywords:** total joint arthroplasty, administrative claims, *ICD-9-CM*, health services research.

Administrative claims data are increasingly being used to define inclusion criteria and patient outcomes for orthopedic health services and clinical outcomes research [1-7]. Unlike research studies that rely on institution- or surgeon-specific clinical outcomes data, administrative claims data offer the advantage of being able to include large numbers of patients from a broad range of geographic and practice settings, thus increasing the sample size and the generalizability of the results. However, administrative claims are coded by nonclinical (administrative) personnel, based on their interpretation of the clinical record; and the correlation between administrative diagnosis and procedure codes and clinical records has not yet been verified in total joint arthroplasty (TJA). Furthermore, before 2005, the *International Classification of Diseases, Ninth Revision, Clinical Modification* (*ICD-9-CM*) diagnosis and procedure codes related to TJA were too vague to discriminate between different causes of failure and types of revision TJA procedures.

In 2005, based on a request from a group of orthopaedic surgeons and health services researchers, the Center for Medicare and Medicaid Services and the National Center for Health Statistics implemented a series of new *ICD-9-CM* diagnosis and procedure codes related to failed TJAs and revision TJA procedures. The motivations for the requested coding changes were (1) to facilitate quality improvement through a better understanding of the mechanisms of failure following TJA, (2) to provide more accurate and descriptive data inputs for the American Joint Replacement Registry project, and (3) to enable more appropriate prospective payments (eg, Diagnosis Related Groups) to hospitals that reflect actual resource utilization related to specific types of revision TJA procedures. Although these new codes were introduced in October 2005, it is unclear how commonly and accurately they are being used by surgeons and administrative coding personnel and therefore how appropriate they are for use in clinical outcomes research. The purpose of this study was to evaluate

*From the *Department of Orthopaedic Surgery, University of California, San Francisco, California; †Philip R. Lee Institute for Health Policy Studies, San Francisco, California; ‡Department of Orthopaedic Surgery, Massachusetts General Hospital, Boston, Massachusetts; and §Department of Orthopaedic Surgery, Mayo Clinic, Rochester, Minnesota.*
Submitted January 4, 2010; accepted April 1, 2010.
No benefits or funds were received in support of this study.
Reprint requests: Kevin J. Bozic, MD, MBA, UCSF Department of Orthopaedic Surgery and Philip R. Lee Institute for Health Policy Studies 500 Parnassus, MU 320W San Francisco, CA 94143-0728.

0883-5403/2506-0013$36.00/0
doi:10.1016/j.arth.2010.04.006

concordance between administrative and clinical diagnosis and procedure codes for revision TJA procedures.

Methods

Retrospective clinical and administrative claims data were collected from 4 high-volume TJA centers (University of California, San Francisco; Mayo Clinic, Rochester, MN; Massachusetts General Hospital, Boston, MA; Newton-Wellesley, Newton, MA). The following *ICD-9-CM* procedure codes were used to evaluate consecutive revision total hip arthroplasty (THA) and revision total knee arthroplasty (TKA) procedures at each institution beginning October 1, 2006: 00.70 (femoral and acetabular component revision THA), 00.71 (acetabular component revision THA), 00.72 (femoral component revision THA), 00.73 (isolated acetabular liner/femoral head revision THA), 81.53 (revision THA, not otherwise specified), 00.80 (all-component revision TKA), 00.81 (tibial component revision TKA), 00.82 (femoral component revision TKA), 00.83 (patellar component revision TKA), 00.84 (isolated tibial insert revision), and 81.55 (revision TKA, not otherwise specified). The cause of TJA failure for each procedure was determined by the *ICD-9-CM* diagnosis code, including 996.41 (mechanical loosening), 996.42 (dislocation), 996.43 (prosthetic joint implant failure/breakage), 996.44 (periprosthetic fracture), 996.45 (periprosthetic osteolysis), 996.46 (articular bearing surface wear), 996.47 (other mechanical complication of prosthetic joint implant), and 996.66 (periprosthetic joint infection).

Concordance between administrative codes and clinical documentation compiled from each institution based on operative notes and discharge summaries was examined using 3 different measures of concordance: (1) sensitivity (the proportion of diagnoses and procedures from the clinical notes that have a corresponding administratively coded diagnosis or procedure code); (2) positive predictive value (PPV) (the proportion of claims that correctly identified a particular diagnosis or procedure); and (3) the κ statistic, used to describe the degree of concordance with established grading criteria as follows: κ less than 0.20, poor; κ = 0.20 to 0.39, fair; κ = 0.40 to 0.59, moderate; κ = 0.60 to 0.79, very good; κ greater than 0.80, excellent [8]. As many as 4 diagnoses were abstracted from the clinical notes and up to 11 *ICD-9* diagnoses codes were extracted from administrative claims. Descriptions of procedures from the clinical record were compared with primary and secondary *ICD-9* procedure codes. Reviewers of clinical documentation were blinded to the corresponding administrative claims data.

Table 1. Concordance of Revision Hip Diagnoses and Diagnosis Codes

ICD-9-CM Diagnosis Code	κ	Sensitivity	PPV	% Cases	% Codes
996.41 Mechanical loosening	0.67	67.4	90.7	36.7%	27.3%
996.42 Dislocation	0.70	73.8	77.6	20.4%	19.4%
996.43 Prosthetic joint implant failure/breakage	0.21	34.6	22.5	6.6%	10.2%
996.44 Periprosthetic fracture	0.59	60.0	64.3	7.7%	7.1%
996.45 Periprosthetic osteolysis	0.41	58.3	39.6	9.2%	13.5%
996.46 Articular bearing surface wear	0.39	36.7	58.1	12.5%	7.9%
996.47 Other mechanical complication	0.17	25.7	24.3	8.9%	9.4%
996.66 Periprosthetic joint infection	0.67	59.3	87.5	15.1%	10.2%

Table 2. Concordance of Revision Hip Procedures and Procedure Codes

ICD-9-CM Procedure Code	κ	Sensitivity	PPV	% Cases	% Codes
00.70 Femoral and acetabular component revision	0.37	83.5	42.2	23.2%	45.9%
00.71 Acetabular component revision	0.44	42.9	79.0	28.6%	15.8%
00.72 Femoral component revision	0.53	51.5	70.8	16.8%	12.2%
0.73 Acetabular liner/femoral head revision	0.16	13.5	73.7	26.5%	4.8%
81.53 Revision THA, NOS	0.08	16.7	10.3	4.6%	7.4%

NOS indicates not otherwise specified.

Results

The most common reasons for revision THA procedures were mechanical loosening (996.41), dislocation (996.42), and infection (996.66) (Table 1). Concordance was very good for these 3 diagnoses (all κ >0.65), but poor for implant failure/breakage (996.43) (κ = 0.21) and other mechanical complication (996.47) (κ = 0.17). The sensitivity of clinical diagnoses having a corresponding administrative claim ranged from a high of 74% for dislocation to a low of 26% for other mechanical complication. Claims for mechanical loosening and infection were correctly coded (PPV) for 91% of cases. Concordance was lower for revision THA procedure codes (all κ < 0.55) than for diagnosis codes (Table 2). Lower κ values could be attributed to overuse of the 00.70 (femoral and acetabular component revision) code as indicated by the low PPV value for this code and low sensitivity for the isolated acetabular component revision (00.71) and isolated femoral (00.72) and head/liner exchange procedures (00.73).

Very good concordance was noted for revision TKA diagnoses of mechanical loosening (996.41), periprosthetic fracture (996.44), osteolysis (996.45) and infection (996.66) (all κ > 0.6); but a large number of cases were incorrectly coded as having other mechanical complication (996.47) (κ = 0.24) and implant failure/breakage (κ = 0.22) (Table 3). The sensitivity of clinical diagnoses having a corresponding administrative claim was not as high for revision TKA diagnoses as it was for revision THA diagnoses, with only periprosthetic frac-