

Osteoarthritis and Cartilage



Brief Report

Accuracy of VA databases for diagnoses of knee replacement and hip replacement

J.A. Singh †‡§||*, S. Ayub ||¶

† Birmingham VA Medical Center and University of Alabama, Birmingham, AL, United States

‡ Center for Surgical Medical Acute care Research and Transitions (C-SMART), Birmingham VA Medical Center, Birmingham, AL, United States

§ Departments of Health Sciences Research and Orthopedic Surgery, Mayo Clinic School of Medicine, Rochester, MN, United States

|| VA Medical Center, Minneapolis, MN, United States

¶ Division of Rheumatology, University of Minnesota, Minneapolis, MN, United States

ARTICLE INFO

Article history:

Received 6 April 2010

Accepted 4 October 2010

Keywords:

Hip replacement
Knee replacement
Validity
Databases
Veterans Affairs

SUMMARY

Purpose: To examine the validity of International Classification of Diseases-Ninth Revision (ICD-9) and Current Procedural Terminology (CPT) codes for knee replacement and hip replacement in Veterans Affairs (VA) databases.

Methods: From a cohort of veterans who received health care at Minneapolis VA Medical Center and/or affiliated medical facilities, we obtained four random samples of 50 patients each with: neither hip nor knee replacement code, knee replacement code only, hip replacement code only and both knee and hip replacement codes. The gold standard was documentation of knee or hip replacement surgery in patient medical records. Accuracy of ICD-9 or CPT code for knee and hip replacement was assessed by calculating sensitivity, specificity, positive and negative predictive values (PPV and NPV).

Results: Of the 200 patients, medical records were available for 166:140 (70%) had complete medical records and 26 (13%) had incomplete medical records. Knee replacement codes were accurate with excellent PPV of 95%, sensitivity of 95%, specificity of 96% and NPV of 96%. Hip replacement codes were accurate with excellent PPV of 98%, sensitivity of 96%, specificity of 99% and NPV of 96%. Sensitivity analyses that included incomplete charts had little impact on these estimates. The procedure dates found in VA databases matched exactly with medical records in 96%.

Conclusions: The ICD-9 and CPT codes for knee replacement and hip replacement in VA databases are valid. These codes may be used to identify cohorts of veterans with knee replacement and hip replacement for research studies.

Published by Elsevier Ltd on behalf of Osteoarthritis Research Society International.

Introduction and summary

Knee replacements and hip replacements are one of the commonest surgeries performed in the US and in the VA hospitals^{1,2}. The Department of Veterans Affairs (VA) is the largest health care delivery system in the US. VA databases have been used in a variety of high-quality health services research and outcomes studies^{3–6}. Although diagnostic inaccuracy (under/over-coding)⁷ and incomplete documentation⁸ are limitations of these databases, administrative/clinical database codes have been found to be valid for acute myocardial infarction⁹, hepatitis C¹⁰ and spondyloarthritis¹¹. Our objective was to assess the validity of International Classification of Diseases-Ninth Revision (ICD-9) or Current Procedural Terminology

(CPT) codes for knee and hip replacement and dates of surgery in Minneapolis VA databases. We obtained four random samples of 50 patients each with: neither hip nor knee replacement code, knee replacement code only, hip replacement code only and both knee and hip replacement codes. The gold standard was documentation of knee or hip replacement surgery in patient medical records. Accuracy of ICD-9 or CPT code for knee and hip replacement was assessed by calculating sensitivity, specificity, positive and negative predictive values (PPV and NPV). Of the 200 patients, medical records were available for 166:140 (70%) had complete medical records and 26 (13%) had incomplete medical records. Knee replacement codes were accurate with excellent PPV of 95%, sensitivity of 95%, specificity of 96% and NPV of 96%. Hip replacement codes were accurate with excellent PPV of 98%, sensitivity of 96%, specificity of 99% and NPV of 96%. Sensitivity analyses that included incomplete charts had little impact on these estimates. The procedure dates found in VA databases matched exactly with medical records in 96%.

* Address correspondence and reprint requests to: Jasvinder A. Singh, University of Alabama, Faculty Office Tower 805B, 510 20th Street S, Birmingham, AL 35294, United States. Tel: 1-205-934-8158.

E-mail address: Jasvinder.md@gmail.com (J.A. Singh).

Methods

For this validation study, we used data from our survey of all veterans in the Upper Midwest Network with a VA health care encounter ($n = 70,334$), details described elsewhere¹². Of these, 1241 had undergone prior knee or hip replacement, as documented by presence of an ICD-9 or CPT code for knee or hip replacement (00.70–00.76, 00.8–00.84, 81.51–81.55; 27,437, 27,438, 27,440–27,443, 27,445–27,447, 27,486, 27,487, 27,125, 27,130, 27,132, 27,134, 27,137, 27,138 and 27,236). We obtained four random samples of 50 patients each with: neither hip nor knee replacement code, knee replacement code only, hip replacement code only and both knee and hip replacement codes. This combined list of 200 patients with names in alphabetic order were provided to a physician (SA) trained in chart abstraction, who was blinded to the ICD- and CPT-codes as well as how the sample was obtained. The Institutional Review Board at the Minneapolis VA Medical Center approved the study.

We used a standardized data extraction form to abstract demographics (age and gender), the date of knee or hip replacement, laterality (right, left), type of replacement (total or partial, primary or revision), underlying diagnosis and the procedure details from the operative and other clinical notes. Documentation of knee or hip replacement surgery in patients' medical records was the gold standard for patient having undergone a knee or a hip replacement. All medical records (in- and out-patient visits) were reviewed starting from the first available encounter. We retrieved complete VA medical records including paper and computerized records for 140 patients (70%) and incomplete medical records for 26 patients (13%). No medical records were available for 34 patients (17%). Thus, the study sample consisted of 166 patients (83%). Main analyses were performed for the 140 with complete charts and sensitivity analyses included 166 patients. This chart retrieval rate is similar to other studies of validity of diagnoses in the VA health care system⁷.

We compared the administrative data definition of presence of ICD-9 or CPT code to the gold standard of chart documentation of knee or hip replacement for each patient. We calculated sensitivity, specificity, PPV and NPV and Kappa statistic for administrative data. Sensitivity was the fraction of those with knee replacement or hip replacement according to the gold standard that were correctly identified as positive by the data definition, respectively. Specificity was the fraction of those without joint replacement (knee or hip) according to the gold standard that were correctly identified as negative by the data definition. PPV was the proportion of those with positive test definition that meet the gold standard definition of medical chart documentation, according to the gold standard. NPV was the proportion of those with negative test definition that did not meet the gold standard definition. The kappa coefficient was used to describe agreement (beyond chance) between the medical record documentation of knee or hip replacement diagnosis (gold standard) and the four database definitions (range, 0–1; 1 = perfect agreement)¹³.

A receiver operating characteristic (ROC) curve analysis plotted the true positive rate against the false positive rate for the different possible cut-points of a case definition¹⁴. An ROC curve shows how any increase in sensitivity is accompanied by a decrease in specificity. The area under the ROC measured discrimination of the database test definition, i.e., its ability to correctly classify those with and without respective joint replacement status. The 45° diagonal line represents the null hypothesis and a test definition that is no better than random will overlap the diagonal.

Because one may arrive at different conclusions depending on the relative importance given to sensitivity or specificity using the classic methods described above, we performed Bayesian analysis. Specificity and sensitivity can be regarded as utility measures of a test procedure under two unknown states of nature, i.e., having or not having the disease. A weighted average of these two quantities is the Bayes utility of a test. Bayes values for each diagnosis definition were calculated by giving a range of importance (P ; value ranging from 0 to 1) to sensitivity and $(1 - P)$ to specificity, where 0 indicates the least importance and 1 indicates the maximum importance. For example, if sensitivity is most critical, one would choose the method with the highest sensitivity, i.e., P of 1. However, in various situations sensitivity and specificity have different weights of importance. Linear combinations of sensitivity and specificity for different values of P were graphed. The analyses were performed using SPSS software, version 11.5 (SPSS, Chicago, IL) and S-plus 2000 (Mathsoft, Seattle, WA).

To examine whether the dates of knee or hip replacement surgery in VA administrative databases are accurate, we identified the cohort of patients that underwent knee or hip replacement during the fiscal years 1992–1998, since the surgery dates were available for this time-period in the administrative databases. 94 patients of the original cohort of 140 patients with complete charts qualified (41 with no knee/hip replacement procedures and five with procedures outside the study period 1992–1998). Date difference was calculated in days for difference between the dates from administrative database and gold standard chart documentation (most often from the operative or anesthesia note).

Results

There were 200 patients with ICD-9 and CPT codes for knee and/or hip replacement. As described above, main analyses were done for patients with complete charts ($n = 140$). The mean age in this cohort of 140 patients was 69.3 (standard deviation, 12.8), and 98.6% were men (138/140).

Sensitivity, specificity, predictive values, kappa statistics and ROC curve

Among the 140 patients with complete medical records, 63 had knee replacement and 64 had hip replacement according to the VA electronic databases (Table I: sum of columns for true and

Table I
Validity of ICD-9 and CPT codes for knee and hip replacement as compared to the gold standard of medical record documentations

	True positives	False positives	True negatives	False negatives	Sensitivity (95% CI)	Specificity (95% CI)	PPV (95% CI)	NPV (95% CI)	Kappa (95% CI)
<i>Knee replacement</i>									
Complete charts ($n = 140$)	60/140	3/140	74/140	3/140	95.2% (92%, 99%)	96.1% (93%, 99%)	95.2% (92%, 99%)	96.1% (93%, 99%)	0.91 (0.84, 0.98)
All charts ($n = 166$)	74/166	7/166	82/166	3/166	96.1% (93%, 99%)	92.1% (88%, 97%)	91.4% (87%, 96%)	96.5% (94%, 99%)	0.88 (0.81, 0.95)
<i>Hip replacement</i>									
Complete charts ($n = 140$)	63/140	1/140	73/140	3/140	95.5% (92%, 99%)	98.6% (97%, 100%)	98.4% (96%, 100%)	96.1% (93%, 99%)	0.94 (0.89, 1.00)
All charts ($n = 166$)	76/166	6/166	81/166	3/166	96.2% (93%, 99%)	93.1% (88%, 97%)	92.7% (89%, 97%)	96.4% (94%, 99%)	0.89 (0.82, 0.96)

false positives). Using the medical chart as the gold standard, of these 140 patients, 63 had knee replacement and 66, hip replacement (numbers derived from cross-tabulations; not depicted in Table 1).

ICD-9 and CPT codes for knee replacement were accurate with excellent PPV of 95%, sensitivity of 95%, specificity of 96% and NPV of 96%, Kappa of 0.91 [95% confidence interval (CI), 0.84–0.98]. ICD-9 and CPT codes for hip replacement were accurate with excellent PPV of 98%, sensitivity of 96%, specificity of 99% and NPV of 96%, Kappa of 0.94 (95% CI, 0.89–1.00) (Table 1). Sensitivity analyses that included incomplete charts had little impact on these estimates. The area under the ROC curve (95% CI) for database definition of knee replacement was 0.875 (0.824, 0.926) and for hip replacement was 0.882 (0.832, 0.931), compared to the medical record documentation [Fig. 1 (A) and (B)].

The Bayesian approach showing the weighted averages of sensitivity and specificity against various weights is presented in

Fig. 1(C). For example, in Fig. 1(C, D), the line for knee replacement represents the values of $0.952 \times P + 0.961 \times (1 - P)$ and the line for hip replacement represents the values of $0.955 \times P + 0.986 \times (1 - P)$ [sensitivity $\times P +$ specificity $\times (1 - P)$] for different values of P . If we gave the most importance to sensitivity, then $P = 1$ and if we give most importance to specificity, then $P = 0$. Interpretation for all other values of P can be made from the graph. In general the Bayesian value is high for almost the entire spectrum of P from 0 to 1.

Of the 94 patients who had knee and/or hip replacement surgery 1992–1998, there was a perfect match for date/s of surgery for 89 patients (95%). Of these 89 patients with perfect match, 57 (64%) had a single replacement surgery and 32 (36%) had multiple replacement surgeries – 24 patients with two procedures, seven patients with three procedures and one patient with five procedures.

There are five patients whose replacement procedure dates did not match completely with VA databases. For patient#1, the dates

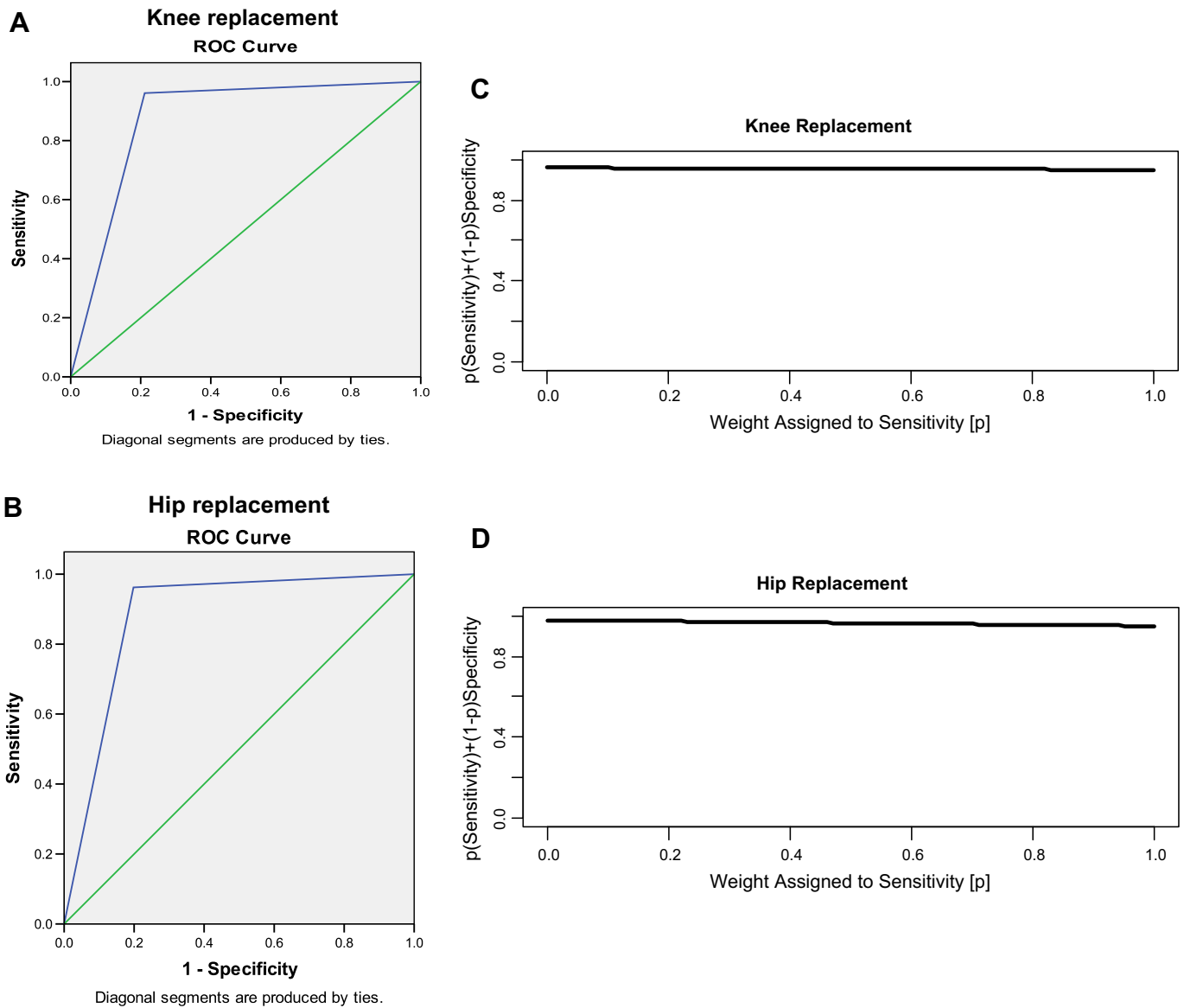


Fig. 1. Accuracy of ICD- and CPT-codes for knee replacement and hip replacement on receiver operating curve (1A and 1B) and using the Bayesian approach (1C and 1D). (A) and (B) The diagonal line represents the null hypothesis that the ICD-code is no better than chance. The farther away the line is from the diagonal, the better the definition is. (C) The x-axis represents values of P , the level of importance given to sensitivity (ranging from 0 to 1). The y-axis represents the Bayes values (for the respective diagnoses) based on a linear combination of sensitivity and specificity for different values of P , with a higher value being better.

of first and fourth procedure matched perfectly, but the second (left total knee replacement) and third (revision of left infected knee) procedure dates in databases differed by +303 and –210 days compared to medical records, respectively. For patient#2, the discrepancy for partial left hip replacement procedure was –154 days. For patient#3, first and third replacement surgery dates matched perfectly; for the second surgery (right total hip replacement), the date discrepancy was +2 days. For patient#4, the second replacement procedure date matched perfectly, but the first replacement data (left total knee replacement) was discrepant by –14 days. For patient#5, the second procedure date matched; for the first procedure (left total knee replacement), the date was missing in VA databases. Thus, from a total of the 143 knee or hip procedures, 137 (96%) procedure dates matched perfectly.

Discussion and conclusion

In this study, we found excellent sensitivity, specificity, PPV and NPV of ICD-9 codes and CPT codes for knee and hip replacements in VA databases. The dates of surgery for knee and hip replacements in VA databases were also found to be accurate 96% of the time. To our knowledge there are no published studies of validation of ICD-9 and CPT codes for knee and hip replacements in VA databases for comparison. Thus, our study adds significantly to the current knowledge.

Several findings deserve further discussion. The high accuracy of codes for knee and hip replacement surgery in VA databases tested against the medical record gold standard is reassuring (ROC curve areas of 0.875 and 0.882 respectively). This is similar to high accuracy for spondyloarthritis¹⁰ and in contrast to the low accuracy rates for rheumatoid arthritis⁷ in VA databases. In conjunction with our other finding that dates in VA databases were 96% accurate, this means that one can identify cohort of knee and hip replacement using a simple valid database approach using ICD-9 and CPT codes in VA databases. These cohorts can be used for performing research studies of comparative effectiveness and/or of post-procedure complications.

Strengths of our study include selection of a random sample, good inter-observer agreement, and standardized data abstraction by a blinded physician. Our study findings may only be applicable to VA databases and accuracy of codes in other databases may differ. Similar validity studies need to be performed in other databases. Charts were available only for 83% of the sample and accuracy may have been different had all charts been available; residual bias is possible.

In conclusion, ICD-9 and CPT codes for knee and hip replacements in VA administrative databases are accurate with excellent sensitivity, specificity and PPV and NPV when compared with gold standard of medical record documentation. Dates of surgery also had high accuracy rates. The findings of this study imply that cohorts of VA patients with these procedures can be identified which can allow epidemiological and outcome studies in these populations.

Author contributions

SA – Data abstraction, assistance in data analyses, preparation and revision of manuscript.

JAS – Study concept and design, Ethics committee approval, Design of data abstraction form, Data analyses, and Preparation and revision of manuscript.

Conflicts of interest

There are no financial conflicts related to this work. JAS has received speaker honoraria from Abbott; research and travel grants

from Allergan, Takeda, Savient, Wyeth and Amgen; and consultant fees from Savient and URL pharmaceuticals. SA has no financial conflicts.

Acknowledgements

JAS's time was protected for research through a NIH CTSA Award 1 KL2 RR024151-01 (Mayo Clinic Center for Clinical and Translational Research) grant. The funding source had no contributions to design or conduct of study, preparation of manuscript or decision to submit it for publication. This material is the result of work supported by the resources and the use of facilities at the Birmingham VA Medical Center, Alabama, USA.

References

1. Kurtz S, Ong K, Lau E, Mowat F, Halpern M. Projections of primary and revision hip and knee arthroplasty in the United States from 2005 to 2030. *J Bone Joint Surg Am* 2007 Apr;89(4):780–5.
2. Khuri SF, Daley J, Henderson W, Barbour G, Lowry P, Irvin G, *et al.* The National Veterans Administration Surgical Risk Study: risk adjustment for the comparative assessment of the quality of surgical care. *J Am Coll Surg* 1995 May;180(5):519–31.
3. Cannon KT, Sarrazin MV, Rosenthal GE, Curtis AE, Thomas KW, Kaldjian LC. Use of mechanical and noninvasive ventilation in black and white chronic obstructive pulmonary disease patients within the Veterans Administration health care system. *Med Care* 2009 Jan;47(1):129–33.
4. Petersen LA, Normand SL, Leape LL, McNeil BJ. Regionalization and the underuse of angiography in the Veterans Affairs Health Care System as compared with a fee-for-service system. *N Engl J Med* 2003 May 29;348(22):2209–17.
5. Ashton CM, Soucek J, Petersen NJ, Menke TJ, Collins TC, Kizer KW, *et al.* Hospital use and survival among Veterans Affairs beneficiaries. *N Engl J Med* 2003 Oct 23;349(17):1637–46.
6. Cully JA, Zimmer M, Khan MM, Petersen LA. Quality of depression care and its impact on health service use and mortality among veterans. *Psychiatr Serv* 2008 Dec;59(12):1399–405.
7. Singh JA, Holmgren AR, Noorbaloochi S. Accuracy of Veterans Administration databases for a diagnosis of rheumatoid arthritis. *Arthritis Rheum* 2004 Dec 15;51(6):952–7.
8. Rabeneck L, Menke T, Simberkoff MS, Hartigan PM, Dickinson GM, Jensen PC, *et al.* Using the national registry of HIV-infected veterans in research: lessons for the development of disease registries. *J Clin Epidemiol* 2001 Dec;54(12):1195–203.
9. Petersen LA, Wright S, Normand SL, Daley J. Positive predictive value of the diagnosis of acute myocardial infarction in an administrative database. *J Gen Intern Med* 1999 Sep;14(9):555–8.
10. Kramer JR, Davila JA, Miller ED, Richardson P, Giordano TP, El-Serag HB. The validity of viral hepatitis and chronic liver disease diagnoses in Veterans Affairs administrative databases. *Aliment Pharmacol Ther* 2008 Feb 1;27(3):274–82.
11. Singh JA, Holmgren AR, Noorbaloochi S. Accuracy of the diagnoses of spondylarthritides in veterans affairs medical center databases. *Arthritis Rheum* 2007 May 15;57(4):648–55.
12. Singh JA, Borowsky SJ, Nugent S, Murdoch M, Zhao Y, Nelson DB, *et al.* Health-related quality of life, functional impairment, and healthcare utilization by veterans: veterans' quality of life study. *J Am Geriatr Soc* 2005 Jan;53(1):108–13.
13. Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics* 1977 Mar;33(1):159–74.
14. Hanley JA, McNeil BJ. The meaning and use of the area under a receiver operating characteristic (ROC) curve. *Radiology* 1982 Apr;143(1):29–36.