

Validity of Gout Diagnoses in Administrative Data

LESLIE R. HARROLD,¹ KENNETH G. SAAG,² ROBERT A. YOOD,¹ TED R. MIKULS,³
SUSAN E. ANDRADE,¹ HASSAN FOUAYZI,¹ JUDITH DAVIS,¹ K. ARNOLD CHAN,⁴
MARSHA A. RAEBEL,⁵ ANN VON WORLEY,⁶ AND RICHARD PLATT⁷

Objective. To determine the utility of using administrative data for epidemiologic studies of gout by examining the validity of gout diagnoses in claims data.

Methods. From a population of ~800,000 members from 4 managed care plans, we identified patients who had at least 2 ambulatory claims for a diagnosis of gout between January 1, 1999 and December 31, 2003. From this group, a random sample of 200 patients was chosen for medical record review. Trained medical record reviewers abstracted gout-related clinical, laboratory, and radiologic data from the medical records. Two rheumatologists independently evaluated the abstracted information and assessed whether the gout diagnosis was probable/definite or unlikely/insufficient information. Discordant physician ratings were adjudicated by consensus. Based on record reviews, patients were also classified according to the American College of Rheumatology (ACR), Rome, and New York gout criteria and these results were compared with the physician global assessments.

Results. There were 121 patients rated as having probable/definite gout by physician consensus, leading to a positive predictive value of ≥ 2 coded diagnoses of gout of 61% (95% confidence interval 53–67). There was low concordance between physician assessments and established gout criteria including ACR, Rome, and New York criteria ($\kappa = 0.17, 0.16,$ and $0.20,$ respectively).

Conclusion. Use of administrative data alone in epidemiologic and health services research on gout may lead to misclassification. Medical record reviews for validation of claims data may provide an inadequate gold standard to confirm gout diagnoses.

KEY WORDS. Gout; Validity; Administrative data.

INTRODUCTION

Recent research has used administrative claims data from managed care organizations to assess the epidemiology and prevalence of gout (1). Although a potentially powerful research tool, these administrative databases have been created primarily for fiscal purposes to track health care utilization of enrollees of health plans. Their usefulness in research is dependent on the accuracy of the information they contain (2–4). Therefore, assessment of the validity of gout diagnoses in administrative data is essential (5).

We calculated the positive predictive value (PPV) of ≥ 2 ambulatory gout diagnoses in the administrative databases of 4 health maintenance organizations (HMOs) through systematic review of medical records using a random selection of health plan members. We hypothesized that in a substantial proportion of patients there would be insufficient information in the medical records to confirm a diagnosis of gout and that many patients would not meet criteria for the diagnosis of gout using the American College of Rheumatology (ACR), Rome, or New York gout criteria (6–8).

Supported by grants from TAP Pharmaceuticals and from the Agency for Healthcare Research and Quality (grant numbers HS-10391 and HS-10389).

¹Leslie R. Harrold, MD, MPH, Robert A. Yood, MD, Susan E. Andrade, ScD, Hassan Fouayzi, MS, Judith Davis, MPH: the Meyers Primary Care Institute, University of Massachusetts Medical School, Fallon Foundation, and Fallon Community Health Plan, Worcester, Massachusetts; ²Kenneth G. Saag, MD, MSc: University of Alabama at Birmingham; ³Ted R. Mikuls, MD, MSPH: University of Nebraska Medical Center, Omaha; ⁴K. Arnold Chan, MD, ScD: Channing Laboratory and Harvard School of Public Health, Boston, Massachusetts; ⁵Marsha A. Raebel, PharmD: Kaiser Permanente Colorado, Denver;

⁶Ann Von Worley, RN, BS, CCRP: Lovelace Clinic Foundation, Albuquerque, New Mexico; ⁷Richard Platt, MD, MS: Harvard Pilgrim Health Care, Boston, Massachusetts.

Dr. Saag has received consulting fees (less than \$10,000 each) from TAP Pharmaceuticals and Savient. Dr. Mikuls has received consulting fees (less than \$10,000) from TAP Pharmaceuticals.

Address correspondence to Leslie R. Harrold, MD, MPH, Department of Medicine, University of Massachusetts Medical School, 55 Lake Avenue North, Worcester, MA 01655. E-mail: HarroldL@ummhc.org.

Submitted for publication January 13, 2006; accepted in revised form May 8, 2006.

PATIENTS AND METHODS

Identification of study cohort. The study population included members from 4 health plans that participate in the HMO Research Network Center for Education and Research on Therapeutics (9). We identified our cohort using a previously developed data set that was comprised of a random sample of ~200,000 members per HMO from January 1, 1999 to December 31, 2003. The sampling scheme and demographic distribution of this population have been described elsewhere (10). The data set included computerized information on utilization of health care services including membership information, pharmacy dispensing data, and selected hospital and ambulatory diagnoses and procedures. Institutional review boards at each participating organization approved this study.

We identified members from the data set who met the following criteria for enrollment into the cohort: at least 19 years of age and ≥ 2 ambulatory visits associated with a gout diagnosis (health care encounters whereby the provider on the billing form codes the visit as being for the management of gout) at least 30 days apart. The identified International Classification of Diseases, Ninth Revision (ICD-9) codes for gout diagnosis were 274.0, 274.1, 274.8, and 274.9. Patients selected for chart review were identified using 2 methodologies to select a range of patients with both active and well-controlled gout. To select patients with general gout, in the first sample, each of the health plans identified a random sample of 25 individuals ($n = 100$) with ≥ 2 encounters for gout at least 30 days apart and continuous enrollment in the health plan with drug benefits during the period 6 months prior to the first encounter through 3 months following the second encounter between January 1, 1999 and December 31, 2003 for medical record reviews. Among identified members, any encounter during the study period coded by a health care provider as being for gout was abstracted. To select patients with presumably active gout based on claims data (meaning repeated encounters associated with a gout diagnosis without gout therapy), in the second sample, we identified members who had ≥ 2 ambulatory visits for gout at least 30 days apart in 12 consecutive months who were continuously enrolled with pharmacy benefits and did not receive a urate-lowering therapy during the period 3 months prior to and 3 months following the consecutive 12-month period. A random sample of 25 members in each health plan ($n = 100$) was selected for medical record abstraction; all health care encounters, regardless of diagnosis, were abstracted over the period beginning 3 months prior to through 3 months following the consecutive 12-month period with ≥ 2 gout encounters.

Medical record abstraction process. A medical record abstraction tool was developed and pilot tested. The tool collected information relevant to confirming a diagnosis of gout and included all elements from the ACR criteria (Appendix A) for the diagnosis of gout (6). Trained medical record abstractors systematically abstracted information on presenting joint symptoms, history of podagra, history of episodic arthritis symptoms, family history of gout,

alcohol use, physical examination findings including presence of tophi, and performance of arthrocentesis and results of crystal examination. De-identified copies of results of all relevant radiologic and laboratory studies (radiographs of the extremities as well as serum urate levels and synovial fluid analyses including cell counts, crystal analysis, and culture results) during the study period were obtained. The medical record abstractors were trained and certified in using de-identified medical records of individuals with gout prior to the abstraction process. Using the de-identified records of 4 individuals with gout, the data elements from the abstractions performed by the medical record abstractors were compared with those abstracted by a rheumatologist investigator (LRH) with an average of 98% agreement.

Physician reviewers. Pairs of rheumatologist investigators (LRH, TRM, KGS, and RAY) independently evaluated the abstracted information and, based on their clinical experience, provided their global assessment as to whether the gout diagnosis was probable/definite or unlikely/insufficient information. Disagreements of gout ratings between investigators were resolved by consensus. Investigator interrater reliability on gout ratings was assessed using a random sample of 20 patients whose chart abstractions were reviewed by all 4 investigators. Kendall's coefficient comparing all 4 investigators was 0.61 (0.4–0.6 is considered moderate agreement and 0.6–0.8 is considered substantial agreement) (11).

Statistical analyses. Initial analyses were performed separately for the 2 samples identified for chart review by comparing the demographic characteristics, treating providers, associated comorbidities, and gout treatments of patients rated as probable/definite with those rated as unlikely/insufficient information using chi-square test or Fisher's exact test for discrete variables and *t*-test or Wilcoxon Mann-Whitney rank-sum test for continuous variables. There were no differences in the results between the 2 samples; therefore, we combined both populations for the final analyses. We assessed the PPV of ≥ 2 gout diagnoses from the administrative database using the investigators' rating of definite/probable gout as the gold standard. We also examined the impact on the PPV of more restricted definitions of the study population based on utilization data such as increasing the number of visits with a gout diagnosis or requiring a dispensing of allopurinol, as well as seeing a rheumatologist.

Lastly, we compared the numbers of persons who met the ACR, Rome, and New York criteria for the diagnosis of gout (Appendix A) (6–8). Using kappa statistics, we compared patients who met these criteria and those who did not with the physician global assessments.

RESULTS

There were 3,866 health plan members who had ≥ 2 encounters associated with an ICD-9 gout diagnosis code who met the selection criteria. From these, 200 patients were randomly identified for chart review. The majority of

Table 1. Characteristics of patients stratified by the physician global assessment*

Characteristics	Probable/definite (n = 121)	Unlikely/insufficient information (n = 79)	P
Age, mean \pm SD years	62 \pm 15	69 \pm 13	< 0.001
Male sex	91 (75)	60 (76)	0.90
Type of care for gout			
Rheumatology	22 (18)	2 (3)	< 0.001
Internal medicine	102 (84)	69 (87)	0.55
Family practice	14 (12)	12 (15)	0.46
Other	64 (53)	29 (37)	< 0.05
Mean \pm SD number of outpatient visits for gout	9 \pm 8	6 \pm 7	< 0.01
Gout-associated comorbidities			
Hypertension	84 (69)	64 (81)	0.07
Dyslipidemia	65 (54)	41 (52)	0.80
Coronary heart disease	30 (25)	18 (23)	0.74
Peripheral arterial disease	10 (8)	10 (13)	0.31
Diabetes mellitus	27 (22)	30 (38)	< 0.05
Nephrolithiasis	9 (7)	4 (5)	0.50
Renal insufficiency	20 (17)	9 (11)	0.31
Adverse reaction to allopurinol	5 (4)	2 (3)	0.71
Treatments for gout			
Glucocorticoid injections (intraarticular or bursal)	19 (16)	4 (5)	< 0.05
Nonsteroidal antiinflammatory drugs	105 (87)	59 (75)	< 0.05
Colchicine	59 (49)	33 (42)	0.33
Allopurinol	40 (33)	26 (33)	0.98
Oral glucocorticoids	46 (38)	25 (32)	0.36
Clinical features			
Presenting with acute arthritic symptoms	111 (92)	44 (56)	< 0.001
History of podagra	78 (64)	15 (19)	< 0.001
History of episodic joint swelling	8 (7)	7 (9)	0.56
Family history of gout	10 (8)	2 (3)	0.13
Any alcohol use	40 (33)	22 (28)	0.44
Presence of tophi	16 (13)	0 (0)	< 0.01
Identification of monosodium urate crystals (tophi, joint fluid, or bursal fluid)	20 (17)	0 (0)	< 0.001

* Values are the number (percentage) unless otherwise indicated.

patients were men (76%), and the mean \pm SD age was 64 \pm 15 years. There were 121 patients rated as having probable/definite gout by the physician reviewers and 79 rated as unlikely or insufficient information (Table 1). Patients rated as having probable/definite gout were more likely to be younger, to have seen a rheumatologist, and to have received either glucocorticoid injections or oral glucocorticoids.

The PPV of ≥ 2 coded diagnoses of gout was 61% (Table

2). Increasing the number of visits associated with a gout diagnosis to at least 3 or 4 did not substantially improve the PPV. Restricting the population to persons who received allopurinol considerably worsened the PPV. Limiting the population to persons who were evaluated by a rheumatologist increased the PPV; however, the denominator was very small.

There was low agreement between physician assessments and the ACR, Rome, and New York criteria ($\kappa =$

Table 2. Positive predictive value (PPV) of gout diagnoses using various selection criteria*

Selection criteria (limiting the population to the conditions below)	Numerator/denominator for PPV	PPV	95% CI
≥ 2 visits associated with a gout ICD-9 code	121/200	61	53–67
≥ 3 visits associated with a gout ICD-9 code	97/151	64	56–72
≥ 4 visits associated with a gout ICD-9 code	84/126	67	58–75
≥ 1 dispensings of allopurinol	26/66	39	28–52
Seen by a rheumatologist	22/24	92	73–99

* All patients were considered to have probable/definite gout based on physician review and were selected based on ≥ 2 ICD-9 codes at least 30 days apart. 95% CI = 95% confidence interval; ICD-9 = International Classification of Diseases, Ninth Revision.

Table 3. A comparison between meeting criteria for gout and the physician global assessment*

Physician global assessment	ACR†		Rome‡		New York§	
	Yes	No	Yes	No	Yes	No
Probable/definite	26	95	29	92	30	91
Unlikely/insufficient information	1	78	4	75	1	78

* ACR = American College of Rheumatology.
† $\kappa = 0.17$.
‡ $\kappa = 0.16$.
§ $\kappa = 0.20$.

0.17, 0.16, and 0.20, respectively) (Table 3). Examples of situations whereby patients were rated as having probable/definite gout by physician reviewers but did not meet the established gout criteria included patients presenting with podagra and hyperuricemia as well as patients with tophi receiving allopurinol without any acute gout symptoms.

DISCUSSION

Our population-based study found the PPV of ≥ 2 gout diagnoses to be 61%, and the PPV did not substantially increase when limiting the population to members with at least 3 or 4 visits with a gout diagnosis. Interestingly, there was low concordance between physician assessments and established diagnostic gout criteria. Our results are similar to results of other studies that have shown relatively low PPVs using administrative claims data to identify patients with rheumatologic conditions such as osteoarthritis (60%) (4,12) and rheumatoid arthritis (59%) (13).

There are several possible reasons why the PPV of ≥ 2 gout diagnoses was low in this patient population. The diagnosis of gout is often less definite than the diagnosis of other conditions such as myocardial infarction or diabetes mellitus where less invasive tests may confirm the diagnosis (14). Also, the managing physicians may assign a diagnostic code of gout before the diagnosis is firmly established, given that the condition is episodic in nature (15). Currently, there are no widely accepted criteria for the diagnosis of gout, except visualization of intracellular monosodium urate crystals in synovial fluid. Aspiration of joints and/or tophi may have been underutilized because the majority of patients with gout were cared for by primary care physicians who may not have the training or expertise in the procedure. In addition, primary care physicians may choose to treat gout empirically without obtaining synovial fluid for crystal confirmation of the diagnosis.

Our findings are similar to findings from other studies that have demonstrated that the PPV of rheumatic diagnoses is influenced by the medical specialty of the health care provider (4,16). Both Harrold et al and Katz et al found that the PPV of an osteoarthritis diagnosis was $>80\%$ in rheumatology practices (4,16). This may in part be due to differences between providers in documentation of clinical information. On average, there was much greater documentation related to gout in evaluations by

rheumatologists as compared with primary care physicians, because the rheumatologists often focused on the musculoskeletal system in their notes. Although evaluation by a rheumatologist was not a sensitive method to identify patients with gout, the high specificity might have potential value in future research using claims data. Lack of documentation of the indications for urate-lowering therapy was an additional difficulty in assigning a rating for each patient on the likelihood of gout. This was particularly problematic in patients with chronic stable gout. For example, in situations where patients were receiving continuous urate-lowering therapy without evidence of tophi based on the medical records and no acute joint symptoms during the period under study, physician reviewers were unable to confirm the diagnosis of gout. This occurred in one-third of patients rated as unlikely/insufficient information and explains in part why the overall PPV of ≥ 2 gout diagnoses was so low in persons receiving allopurinol. If those asymptomatic patients receiving urate-lowering therapy had been considered instead to have probable/definite gout, the PPV of ≥ 2 encounters associated with a gout diagnosis would have increased to 75%.

The low concordance between physician global assessments and the well-established Rome, New York, and ACR criteria is most likely related to the nature of our study. Those criteria were established to evaluate patients prospectively in a clinical setting or using epidemiologic surveys (1). Documentation in medical records was not sufficient to assess all the elements included in the criteria. For example, the New York criteria include as one criterion "a clear history and/or observation of a good response to colchicine, defined as a major reduction in objective signs of inflammation within 48 hours of the onset of the therapy" (8), and the ACR criteria include a criterion that requires confirmation that the maximum inflammation developed within 1 day (6). Such details are not commonly found in general medical record documentation. Both clinicians and researchers would benefit from the development of more simple criteria to enable a standard approach to diagnose gout.

The strengths of the study include sampling patients based on 2 different methodologies. This enabled assessment as to whether our sampling strategy was related to the resulting outcomes, which it was not. We also identified patients from 4 different health plans in the US, thus limiting the impact of regional differences in the evaluation and documentation of care related to gout. Limitations of the study include a restricted period for medical record abstraction: 18 months for patients not receiving urate-lowering therapy and having ≥ 2 encounters for gout in a 12-month period, and up to 5 years for those with ≥ 2 encounters over the entire study period. Because gout is an episodic condition that may recur infrequently and some episodes may potentially be managed without seeking medical attention, this period may not be adequate. However, abstraction over a longer period was not feasible due to resources and changes in health plan enrollment by patients. In addition, we had a limited sample size of 200, although using 95% confidence intervals for our estimate of the PPV, the precision surrounding our estimate was

$\pm 7\%$. Lastly, our study was performed in 4 US HMOs and our results may not be generalizable to other health plans or other systems of health care delivery.

Use of administrative databases enables rapid identification of a large number of patients; however, these databases have inherent limitations. Validation of the conditions of interest to be studied using administrative data is necessary. Medical record reviews for validation of claims data may be an inadequate gold standard to confirm gout diagnoses. Although the use of administrative databases can be a powerful resource for epidemiologic and health services research, it is important to recognize the limitations of using such data.

ACKNOWLEDGMENTS

We thank Jackie Fuller, MPH, Jim Livingston, MBA, and Parker Pettus, MS, for leading the data management and computer programming effort. We thank Kimberly Lane, MPH, and Kimberly Hill, MS, for project coordination. We acknowledge the project managers, chart abstractors, and programmers from each of the organizations.

AUTHOR CONTRIBUTIONS

Dr. Harrold had full access to all of the data in the study and takes responsibility for the integrity of the data and the accuracy of the data analysis.

Study design. Drs. Harrold, Saag, Yood, Andrade, Chan, Raebel, and Platt, and Ms Davis.

Acquisition of data. Drs. Harrold, Mikuls, Chan, and Raebel, and Mr. Fouayzi, Ms Davis, and Ms Von Worley.

Analysis and interpretation of data. Drs. Harrold, Saag, Yood, Mikuls, Andrade, Chan, and Raebel, and Mr. Fouayzi.

Manuscript preparation. Drs. Harrold, Saag, Yood, Mikuls, Raebel, and Platt, and Ms Von Worley.

Statistical analysis. Drs. Harrold and Andrade, and Mr. Fouayzi.

Critical manuscript review. Ms Davis and Dr. Chan.

ROLE OF THE STUDY SPONSOR

TAP Pharmaceuticals was not involved in the study design, data collection, data analysis, or manuscript preparation. Although it was understood prior to initiating the study that the results would be submitted for publication, the manuscript was not reviewed by the study sponsor prior to submission.

REFERENCES

- Wallace KL, Riedel AA, Joseph-Ridge N, Wortmann R. Increasing prevalence of gout and hyperuricemia over 10 years among older adults in a managed care population. *J Rheumatol* 2004;31:1582-7.
- Iezzoni LI. Assessing quality using administrative data [review]. *Ann Intern Med* 1997;127:666-74.
- 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;14:555-8.
- Harrold LR, Yood RA, Andrade SE, Reed JI, Cernieux J, Strauss W, et al. Evaluating the predictive value of osteoarthritis diagnoses in an administrative database. *Arthritis Rheum* 2000;43:1881-5.
- Rawson NS, D'Arcy C. Assessing the validity of diagnostic information in administrative health care utilization data: experience in Saskatchewan. *Pharmacoepidemiol Drug Saf* 1998;7:389-98.
- Wallace SL, Robinson H, Masi AT, Decker JL, McCarty DJ, Yu TF. Preliminary criteria for the classification of the acute arthritis of primary gout. *Arthritis Rheum* 1977;20:895-900.
- Kellgren JH, Jeffrey MR, Ball J, editors. *The epidemiology of chronic rheumatism*. Oxford: Blackwell; 1963. p. 327.
- Bennett PH, Wood PH, editors. *Population studies of the rheumatic diseases: proceedings of the third international symposium, New York, Jun 5-10, 1966*. Amsterdam: Excerpta Medica Foundation; 1968. p. 457-8.
- Platt R, Davis R, Finkelstein J, Go AS, Gurwitz JH, Roblin D, et al. Multicenter epidemiologic and health services research on therapeutics in the HMO Research Network Center for Education and Research on Therapeutics. *Pharmacoepidemiol Drug Saf* 2001;10:373-7.
- Chan KA and the HMO Research Network CERTs Patient Safety Study Investigators. Development of a multipurpose dataset to evaluate potential medication errors in ambulatory setting. In: Henriksen K, Battles JB, Marks ES, Lewin DI, editors. *Advances in patient safety: from research to implementation*. Vol. 2. Rockville (MD): Agency for Healthcare Research and Quality; 2005.
- Sackett DL, Haynes RB, Guyatt GH, Tugwell P. *Clinical Epidemiology: a basic science for clinical medicine*. 2nd ed. Boston: Little, Brown and Company; 1991. p. 30.
- Gabriel SE, Crowson CS, O'Fallon WM. A mathematical model that improves the validity of osteoarthritis diagnoses obtained from a computerized diagnostic database. *J Clin Epidemiol* 1996;49:1025-9.
- Gabriel SE. The sensitivity and specificity of computerized databases for the diagnosis of rheumatoid arthritis. *Arthritis Rheum* 1994;37:821-3.
- Terkeltaub RA. Gout: epidemiology, pathology and pathogenesis. In: Klippel JH, Crofford L, Stone JH, Weyand CM, editors. *Primer on the rheumatic disease*. 12th ed. Atlanta: Arthritis Foundation; 2001. p. 307-12.
- Quam L, Ellis LB, Venus P, Clouse J, Taylor CG, Leatherman S. Using claims data for epidemiologic research: the concordance of claims-based criteria with the medical record and patient survey for identifying a hypertensive population. *Med Care* 1993;31:498-507.
- Katz JN, Barrett J, Liang MH, Bacon AM, Kaplan H, Kieval RI, et al. Sensitivity and positive predictive value of Medicare Part B physician claims for rheumatologic diagnoses and procedures. *Arthritis Rheum* 1997;40:1594-600.

APPENDIX A: CRITERIA CREATED FOR THE DIAGNOSIS OF GOUT

American College of Rheumatology criteria (6)

Urate crystals in joint fluid

OR

Tophus (proven by microscopic evaluation, tissue biopsy, etc.)

OR presence of at least 6 of the following:

1. More than 1 attack of acute arthritis
2. Maximal inflammation developed within 1 day
3. Monarthritis attack
4. Redness observed over joints
5. First metatarsophalangeal joint painful or swollen
6. Unilateral first metatarsophalangeal joint attack
7. Unilateral tarsal joint attack
8. Tophus (suspected)
9. Hyperuricemia
10. Asymmetric swelling within a joint on radiograph
11. Subcortical cysts without erosions on radiograph
12. Joint fluid culture negative for organisms during attacks

Rome criteria (7)

Two of the following 4 criteria had to be present to make a diagnosis of gout:

1. Serum uric acid level of >7.0 mg/dl in men, or 6.0 mg/dl in women
2. Tophi
3. Uric acid crystals in synovial fluid or tissues
4. History of attacks of painful joint swelling of abrupt onset with remission within 2 weeks

New York criteria (8)

Uric acid crystals in synovial fluid or tissue (tophi, etc.)

OR the presence of ≥ 2 of the following criteria:

1. History or observation of at least 2 attacks of painful limb swelling with remission within 1–2 weeks
2. History or observation of podagra
3. Presence of tophus
4. History or observation of a good response to colchicine (major reduction in objective signs of inflammation within 24 hours of onset of therapy)