Published in final edited form as:

J Am Acad Dermatol. 2008 December ; 59(6): 981–985. doi:10.1016/j.jaad.2008.08.034.

Potential Misclassification of Psoriasis Patients in Electronic Databases

Murat Icen, MD¹, Cynthia S. Crowson, MS¹, Marian T. McEvoy, MD², Sherine E. Gabriel, MD MSc¹, and Hilal Maradit Kremers, MD MSc¹

¹Department of Health Sciences Research, College of Medicine, Mayo Clinic, Rochester Minnesota

²Department of Dermatology, College of Medicine, Mayo Clinic, Rochester Minnesota

Abstract

Background—Electronic claims and medical record databases are increasingly used for observational studies of psoriasis. The purpose of this study is to assess the validity of psoriasis diagnostic codes in an electronic database.

Methods—This study was performed in a population-based setting in Olmsted County, Minnesota where all diagnoses and procedures from all health care providers in a large community are indexed and recorded in an electronic database. The database was searched for subjects aged ≥ 18 years with diagnostic codes consistent with psoriasis for the time period 1/1/1976–1/1/2000. The complete medical records of all subjects were reviewed manually for validation of psoriasis diagnoses.

Results—We reviewed the complete medical records of 2556 subjects with at least 1 diagnostic code consistent with psoriasis. Based on medical record review, 1458 (57.0%) subjects were confirmed to have psoriasis of whom the majority (81%) were confirmed by a dermatologist. The most commonly used diagnostic codes for psoriasis was ICD9 696.1 (psoriasis, not otherwise specified) with a positive predictive value of 68.7% (95% confidence intervals [CI]: 66.5%, 70.9%). Increasing frequency of codes in a given time window was associated with positive predictive values. However, positive predictive value for only one code in a 5-year time window was still as high as 60% (95% CI: 57%, 63%).

Conclusions—Electronic identification of psoriasis subjects by diagnostic codes alone may lead to misclassification in database studies.

Keywords

Epidemiology; psoriasis; misclassification; healthcare databases

Introduction

Clinical research using data from electronic claims and medical record databases increased significantly in the past decade(1). This approach provides convenience and speed in

Corresponding author: Hilal Maradit Kremers, MD MSc, Mayo Clinic, Department of Health Sciences Research, Division of Epidemiology, 200 First St. SW, Rochester, MN 55905, telephone: 1 507 266 3169, fax: 1 507 284 1516, maradit@mayo.edu.

Conflict of interest: Funded by an unrestricted research grant from Amgen Inc. Dr. Maradit Kremers has received consulting fees, speaking fees, and/or honoraria (less than \$10,000) from Amgen.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

performing studies on large numbers of subjects for not only studies of medical outcomes but also for health care delivery and quality of care purposes(2). Opportunities to use claims and medical record databases are expected to increase in the future and have a significant impact on treatment, patient care and healthcare policy decisions. Therefore, it is imperative to assess the validity of data in electronic claims and medical record databases(3).

There have been several recent reports on the epidemiology and outcomes of psoriasis that were carried using medical record databases, including the General Practice Research Database (GPRD)(4-8) in the United Kingdom and Maccabi Healthcare Services(9) and Clalit Health Services(10) databases in Israel. In the GPRD studies, the investigators were able to validate psoriasis diagnoses in 90% of the subjects by evaluation of prescription medications(7), or in 82% by directly contacting the treating general practitioners to confirm the diagnosis in a subset of 564 subjects(8). However, case validation by drug exposure may still result in misclassification of psoriasis subjects, since the same topical treatments are widely prescribed for a number of dermatological conditions.

We have recently completed a population based retrospective study to investigate trends in incidence of psoriasis in Olmsted County, Minnesota(11). As part of this study, we manually reviewed the complete medical records of all subjects who had diagnostic codes consistent with psoriasis in an electronically available diagnostic database that includes all medical and surgical diagnoses retrieved from the complete medical records. We capitalized on this unique resource to assess the validity of electronically recorded psoriasis diagnostic codes.

Methods

This is a retrospective study using the data resources of the Rochester Epidemiology Project. As described previously, Rochester Epidemiology Project (REP) is a centralized medical records linkage system that encompasses all healthcare delivered to residents of Olmsted County, Minnesota(12). All healthcare providers in Olmsted County utilize a unit medical record system whereby all medical information on each individual is accumulated within a single dossier. Furthermore, detailed indices containing all clinical, surgical and pathological diagnoses are recorded electronically. The diagnostic indexing system of the REP is an informatics tool that captures and classifies diagnostic information from the patients' complete medical records (i.e. inpatient and outpatient medical records from all healthcare providers). The diagnoses assigned at each visit are coded and indexed continuously. The indices are then used for accessing patient information in the original medical records through the medical records linkage system and identification of patients based on clinical criteria. The diagnostic system dates back to beginning of the 20th century and a variety of coding systems have been used over time. Starting in 1975, coding was based on the Hospital Adaptation of the International Classification of Diseases (ICD), Second Edition (HICDA). Mayo Clinic modified this code to increase its specificity to over 34,000 codes and additional digits were added to identify individual diseases. More recently, indexing is based on electronic capture of final diagnoses from the electronic medical records.

We searched the computerized indices of the Rochester Epidemiology Project to identify all potential psoriasis subjects aged ≥ 18 years between January 1, 1976 and January 1, 2000. Psoriasis codes used in the study were ICD9 codes 696.1 (psoriasis not otherwise specified), 696.2 (psoriasiform dermatitis), 696.3 (pityriasis rosea), 696.5 (other and unspecified pityriasis), 696.8 (other psoriasis and similar disorders) and HICDA codes 06961210 (psoriasis not otherwise specified), 06961211 (dermatitis, psoriasiform), 06961212 (psoriasis, inverse), 06961213 (seborrheic/sebopsoriasis), 06961220 (psoriasis, vulgaris), 06961310 (psoriasis, guttate), 06961410 (psoriasis, pustular) and 06961510 (erythroderma, psoriaticum). The complete medical records of all potential subjects were reviewed by 2 trained nurse abstractors

and the primary author. The psoriasis diagnoses were validated by either a dermatologist's confirmation of psoriasis in the medical records, description of the psoriatic lesions by the reporting physician, or the result of a skin biopsy, whenever possible. Agreement between abstractors was high with a kappa statistic of 0.87 for psoriasis diagnosis and 0.84 for prevalence versus incidence status. When the diagnosis of psoriasis was doubtful, records of the subject were reviewed by the dermatologist co-investigator (MTM). Subjects who were first diagnosed with psoriasis during the study time period were defined as "incident" whereas those who had psoriasis prior to study period or prior to moving into Olmsted County were defined as "prevalent" psoriasis subjects.

Percentages were used to summarize the data. Positive predictive value (PPV) was estimated as the number of patients verified to have psoriasis divided by the total number of patients screened for psoriasis in each category. Confidence intervals for PPV were obtained using exact binomial methods. Since the disease course, self-management practices and the frequency of physician visits for psoriasis may be different in prevalent subjects, compiling incident and prevalent subjects may obscure the predictive value analyses for code type and frequency. Therefore, we focused only on incident psoriasis subjects for the primary analyses and included the prevalent subjects for secondary analysis.

Results

We identified a total of 2,556 potential subjects with diagnostic codes consistent with psoriasis between 1976 and 2000. After manual screening of the medical records from all healthcare providers, 1,458 (57.0%) psoriasis subjects were confirmed as incident psoriasis. Of the remaining 1,098 subjects, 62.4% had diagnoses other than psoriasis (e.g., parapsoriasis, pityriasis rosea, psoriasiform dermatitis) and 37.6% were prevalent subjects, i.e. psoriasis prior to study time frame or moved to Olmsted County within the study time frame with pre-existing psoriasis.

The most commonly used diagnostic code for psoriasis was ICD9 696.1 'psoriasis not otherwise specified' in 1770 (69.2%) of subjects (Table 1). The next common code was HICDA 06961220 'psoriasis, vulgaris' in 116 (4.5%) subjects. The PPV for code 696.1 and code 06961220 were 68.7% (95% CI: 66.5, 70.9) and 94.0% (95% CI: 88.0, 97.5), respectively. ICD9 696.2 'psoriasiform dermatitis' was the third common code used in 107 (4.1%) subjects but was associated with a PPV of 18.7% (95% CI: 11.8, 27.4). Specific HICDA codes for less common psoriasis subtypes HICDA 06961310 ('psoriasis, guttate' 54 subjects [2.1%]), HICDA 06961410 ('psoriasis, pustular' 41 subjects [1.6%]) and HICDA 06961213 ('seborrhiassis/sebopsoriasis' 38 subjects [1.4%]) were rare but were associated with PPV of 77.8% (95% CI: 64.4, 88.0), 90.2% (95% CI: 76.9, 97.3) and 84.2% (95% CI: 68.8, 94.0), respectively.

We then assessed the PPV of the frequency of codes in 1, 2, 3 and 5 year time windows (Figure 1). For a one year period, regardless of the code used, having a single code for psoriasis had a PPV of 67% (95% CI 65, 69) which increased to 70% (95% CI 63, 76) in presence of 2 codes and 84% (95% CI 74, 91) in presence of ≥ 3 psoriasis codes. When we examined the frequency of codes over two years, one psoriasis code over a 2-year time window had a PPV of 64% (95% CI: 62, 66) which increased to 78% (95% CI: 73, 82) when 2 codes were present and 86% (95% CI: 79, 90) when ≥ 3 codes were present. Figure 1 illustrates the decrease in PPV for a single diagnostic code for psoriasis and increase in PPV of 2 and ≥ 3 diagnostic codes as the time frame increases.

Although the PPV value of a single code decreased as the time window increased, the PPV for only one code in a 5-year time window was still 60% (95% CI 57, 63), indicating that at least

60% of these subjects had psoriasis, even though they only have 1 psoriasis code in a 5-year time window.

In secondary analyses, we examined PPV for both incident and prevalent psoriasis subjects. By including the prevalent subjects, the PPV for the most common code ICD9 696.1 increased from 68.7% (95% CI: 66.5, 70.9) to 72.9% (95% CI: 67.5, 74.6).

Discussion

In this study, we report on the validity of psoriasis diagnostic codes in an electronic medical record database. Both the type and frequency of psoriasis codes influence the validity of identification of psoriasis subjects. The PPV remains as high as 60% for a single psoriasis code in a 5-year time window, indicating that the relapsing and remitting disease course in psoriasis may challenge electronic identification of psoriasis subjects in database studies. These findings indicate that misclassification of psoriasis subjects can be substantial in electronic database studies. Future claims and medical record database studies in psoriasis should take into account the extent to which diagnoses are validated, such as through review of original medical records, and sensitivity analyses should evaluate the potential for bias due to misclassification in explaining study findings.

Electronic claims and medical record databases are increasingly used in recent years in drug utilization reviews, epidemiologic studies of risk factors and outcomes, investigations on risk adjustment and health economics, physician profiling and quality of care(1,13). These computerized databases offer unique advantages. They contain information literally on millions of subjects(5). The vast amount of information on routine clinical care contained in these databases makes it possible to run comprehensive analyses to identify unknown, real world associations in the clinical data and the size of the databases allows investigation of rare outcomes. Systematic and comprehensive data collection saves time and money when undertaking studies and since they do not require informed consent, decreases the likelihood of selection bias due to non-response(2,14,15). Concerns about the validity of results obtained by electronic medical record review arise from at least three possible sources of error which have all been demonstrated in previous investigations. These include inaccurate diagnoses (13), missing or incomplete data(16) and inaccuracies in data entry(17).

Psoriasis is a chronic systemic inflammatory disorder with characteristic skin lesions that can be recognized and diagnosed fairly easily. Epidemiologic studies in psoriasis are relatively challenging, due partly to lack of universally accepted diagnostic criteria(18), and partly to the remitting relapsing disease course and heterogeneity in clinical presentation. The same reasons also create difficulties in relying on electronic psoriasis diagnoses in database studies.

There are a number of possible explanations for the relatively low PPV of the diagnostic codes for psoriasis. First, psoriasis may remain as a differential diagnosis in initial and repeated physician visits but may be dismissed later on. Yet, it will remain as a false positive psoriasis diagnosis in the electronic records. Second, subjects may be misdiagnosed and treated as psoriasis. Access to complete medical records enables us to ascertain the case status by controlling not only the diagnoses but also the descriptions of the lesions over the entire life span of the subject. We were able to confirm or reject a doubtful diagnosis with the diagnosis of a dermatologist, whenever available. Third, the most commonly used code is a non-specific psoriasis code which does not even specify the type of psoriasis. Indeed, our analyses indicated that codes specifying the type of psoriasis displayed higher sensitivity with higher PPV. A fourth reason for the low PPV may be the long study period. As our results show, PPV of number of diagnostic codes change over time. We analyzed the PPV for 1, 2 and 3 or more diagnostic codes over time (Figure). Presence of one single diagnostic code consistent with

psoriasis in a one year time frame was associated with a PPV of 67% and it decreased steadily over 5 years. PPV of three or more codes, on the other hand, increased over time. PPV of two diagnostic codes, however, increased from 70% over one year to 78% over two years and did not show a change between two and five years (76%). Our study clearly demonstrates that first, even patients who received 3 or more diagnostic codes consistent with psoriasis over time have 10 to 16% chance of being misdiagnosed or misclassified and second, a single psoriasis code in a 5-year time window time frame may indicate a true psoriasis flare. Finally, low PPV in our study may be related to the inclusion criteria of the study as we excluded prevalent psoriasis subjects. Inclusion of prevalent psoriasis subjects resulted in higher PPV (increased from 68.7% to 72.9%).

In a recent study(8) using the General Practice Research Database (similarly a medical records database), the investigators were able to validate psoriasis diagnoses in 82% of subjects with more than one code over a 2-year period. This estimate is in fact very similar to ours. We estimated PPV of 2 diagnostic codes as 78% and of 3 or more codes as 86% over two years.

PPV of diagnostic codes has important implications in epidemiological studies. The primary outcome of interest in our study was also PPV, rather than sensitivity or specificity, because our aim was to assess whether subjects with diagnostic codes truly had psoriasis. In an incidence study such as ours, a low PPV (and a high negative predictive value) may result in over-estimation of the true incidence of the disease. In a study that examines risk for certain outcomes among subjects with psoriasis, a low PPV may result in underestimation of the true risk for the outcome if the psoriasis cohort contains subjects who were misclassified as psoriasis.

A number of potential limitations should be considered in interpretation of these results. Each electronic database is unique and therefore, our validation results may not apply to other claims and medical record databases. The remitting relapsing course of psoriasis and the individual variability in the extent of the lesions present challenges in case ascertainment. A true psoriatic lesion diagnosed by a general practitioner may resolve completely by the time the patient is examined by a dermatologist for another lesion in the same site, resulting in the case to be classified as a false positive. Including psoriasis in the differential diagnosis (possible psoriasis) in the first visit but later dismissing it may be another. During the long time frame of the study, changes in coding practices and coding systems can affect PPV over time. In earlier years of this study, diagnoses were recorded only once per year and this might have artificially overestimated the PPV for a single code in a one year period. Strengths of our study include the community based setting, long follow-up period, approximately 10 years prior to and 10 years following psoriasis incidence date, and most importantly, access to the original medical records from all health care providers and virtually complete chart review of all 2556 potential subjects.

In conclusion, variable disease course may lead to misclassification of psoriasis subjects in electronic claims and medical record databases. Our findings emphasize the importance of validation of psoriasis diagnoses in database studies.

Acknowledgments

We would like to thank Rita Black, Deb Olson and Diane Carlson for data abstraction, Justin Carlin for database management, David Tines for computer programming and Mitch Baias for record retrieval.

Funded by NIH Grant AR30582, The Rochester Epidemiology Project, Dr. W. A. Rocca, PI.

References

1. Wilchesky M, Tamblyn RM, Huang A. Validation of diagnostic codes within medical services claims. *J Clin Epidemiol* 2004;57(2):131–41. [PubMed: 15125622]
2. Schneeweiss S, Avorn J. A review of uses of health care utilization databases for epidemiologic research on therapeutics. *J Clin Epidemiol* 2005;58(4):323–37. [PubMed: 15862718]
3. Preen DB, Holman CD, Lawrence DM, Baynham NJ, Semmens JB. Hospital chart review provided more accurate comorbidity information than data from a general practitioner survey or an administrative database. *J Clin Epidemiol* 2004;57(12):1295–304. [PubMed: 15617956]
4. Gelfand JM, Berlin J, Van Voorhees A, Margolis DJ. Lymphoma rates are low but increased in patients with psoriasis: results from a population-based cohort study in the United Kingdom. *Archives of Dermatology* 2003;139(11):1425–9. [PubMed: 14623702]
5. Gelfand JM, Neimann AL, Shin DB, Wang X, Margolis DJ, Troxel AB. Risk of myocardial infarction in patients with psoriasis. *JAMA* 2006;296(14):1735–41. [PubMed: 17032986]
6. Brauchli YB, Jick SS, Curtin F, Meier CR. Association between beta-blockers, other antihypertensive drugs and psoriasis: population-based case-control study. *Br J Dermatol*. 2008
7. Neimann AL, Shin DB, Wang X, Margolis DJ, Troxel AB, Gelfand JM. Prevalence of cardiovascular risk factors in patients with psoriasis. *J Am Acad Dermatol* 2006;55(5):829–35. [PubMed: 17052489]
8. Huerta C, Rivero E, Rodriguez LA. Incidence and risk factors for psoriasis in the general population. *Arch Dermatol* 2007;143(12):1559–65. [PubMed: 18087008]
9. Shapiro J, Cohen AD, David M, Hodak E, Chodik G, Viner A, et al. The association between psoriasis, diabetes mellitus, and atherosclerosis in Israel: a case-control study. *J Am Acad Dermatol* 2007;56(4):629–34. [PubMed: 17157411]
10. Cohen AD, Sherf M, Vidavsky L, Vardy DA, Shapiro J, Meyerovitch J. Association between psoriasis and the metabolic syndrome. A cross-sectional study. *Dermatology* 2008;216(2):152–5. [PubMed: 18216477]
11. Icen M, Crowson CS, McEvoy MT, Dann FJ, Gabriel SE, Maradit Kremers H. Trends in incidence of psoriasis over three decades: a population based study. 2008 under review.
12. Melton L. History of the Rochester Epidemiology Project. *Mayo Clinic Proceedings* 1996;71:266–274. [PubMed: 8594285]
13. Peabody JW, Luck J, Jain S, Bertenthal D, Glassman P. Assessing the accuracy of administrative data in health information systems. *Med Care* 2004;42(11):1066–72. [PubMed: 15586833]
14. Harrison JH Jr. Introduction to the mining of clinical data. *Clin Lab Med* 2008;28(1):1–7. v. [PubMed: 18194715]
15. Suissa S, Garbe E. Primer: administrative health databases in observational studies of drug effects--advantages and disadvantages. *Nat Clin Pract Rheumatol* 2007;3(12):725–32. [PubMed: 18037932]
16. Hennessy S, Bilker WB, Weber A, Strom BL. Descriptive analyses of the integrity of a US Medicaid claims database. *Pharmacoepidemiol Drug Saf* 2003;12(2):103–11. [PubMed: 12647699]
17. Khwaja HA, Syed H, Cranston DW. Coding errors: a comparative analysis of hospital and prospectively collected departmental data. *BJU Int* 2002;89(3):178–80. [PubMed: 11856094]
18. Griffiths CE. Psoriasis: future research needs and goals for the twenty-first century. *Dermatol Clin* 2004;22(4):493–9. [PubMed: 15450345]

Figure 1. Positive predictive value (PPV) of frequency of codes for psoriasis

Table 1
Positive predictive value (PPV) of psoriasis diagnostic codes

Diagnosis Text	Diagnostic Code	N	PPV (95% CI)
Psoriasis, not otherwise specified	ICD 696.1 or HICDA 06961210	1770	68.7 (66.5, 70.9)
Psoriasis, vulgaris	HICDA 06961220	116	94.0 (88.0, 97.5)
Psoriasiform dermatitis	ICD 696.2 or HICDA 06961211	107	18.7 (11.8, 27.4)
Psoriasis, guttate	HICDA 06961310	54	77.8 (64.4, 88.0)
Psoriasis, pustular	HICDA 06961410	41	90.2 (76.9, 97.3)
Seborrhiasis/sebopsoriasis	HICDA 06961213	38	84.2 (68.8, 94.0)
Pityriasis and other psoriasis	Other ICD9 codes 696.3, 696.5 or 696.8 or HICDA 06961212	17	11.8 (1.5, 36.4)

Abbreviations: CI: confidence interval; ICD: International Classification of Diseases; HICDA: Hospital Adaptation of the International Classification of Diseases.