Accuracy of Diagnostic Coding for Sarcoidosis in Electronic Databases: A Population-Based Study

Patompong Ungprasert, MD, MS[1,2,*], **Eric L. Matteson, MD, MPH**[1,3], and **Cynthia S. Crowson, MS**[1,4]

[1]Division of Rheumatology, Department of Internal Medicine, Mayo Clinic College of Medicine and Science, Rochester, MN 55905, USA [2]Division of Rheumatology, Department of Medicine, Faculty of medicine Siriraj hospital, Mahidol University, Bangkok 10700, Thailand [3]Division of Epidemiology, Department of Health Science Research, Mayo Clinic College of Medicine and Science, Rochester, MN 55905, USA [4]Division of Biomedical Statistics and Informatics, Department of Health Science Research, Mayo Clinic College of Medicine and Science, Rochester, MN 55905, USA

Abstract

PURPOSE—Epidemiologic study of sarcoidosis utilizing electronic databases has been increasingly popular. However, the accuracy of diagnostic codes for sarcoidosis is unknown.

METHODS—The medical record-linkage system of the Rochester Epidemiology Project was searched to identify all potential adult cases of sarcoidosis between January 1, 1995 and December 31, 2013 in Olmsted County, Minnesota, using the International Classification of Diseases, Ninth Revision (ICD-9) code 135 (sarcoidosis). Complete medical records of those potential cases were individually reviewed. The diagnosis of sarcoidosis was confirmed by the presence of non-caseating granuloma on histopathology, radiographic findings of intrathoracic sarcoidosis and compatible clinical presentations. Positive predictive value (PPV) was estimated as the number of patients verified to have sarcoidosis divided by the number of patients with a diagnostic code for sarcoidosis.

RESULTS—The study cohort included 366 patients with at least one code for sarcoidosis. Of these, 224 cases of confirmed sarcoidosis were identified, resulting in PPV of 61.2% (95% CI, 56.0% – 66.2%). A total of 268 patients in the database had a code for sarcoidosis on least two occasions separated by at least 30 days. Of these, there were 205 cases of confirmed sarcoidosis. The PPV for having the code at least twice was 76.5% (95% CI, 71.0% – 81.4%).

---

*Corresponding author; P.U.: P.Ungprasert@gmail.com, Division of Rheumatology, Mayo Clinic, 200 First Avenue SW, Rochester, MN 55905, USA, Tel 507-284-8450 Fax 507-284-0564.

CONCLUSIONS—The PPV of ICD-9 code for sarcoidosis is relatively low and, thus, further verification is required for studies using electronic databases.

Introduction

Epidemiologic and clinical research utilizing electronic databases have been increasingly popular over the past two decades [1]. The advantages of this approach include the convenience and speed in conducting analyses on large number of subjects which would allow detection of small-size associations and provide more precise effect estimates. However, the accuracy of diagnostic codes is generally limited. Studies on diagnostic codes for common pulmonary diseases, such as chronic obstructive pulmonary disease, asthma and venous thromboembolism, revealed positive predictive values of only 50% to 70% [2–4]. The current study was conducted with the aim to assess the accuracy of diagnostic coding for sarcoidosis.

Methods

Approval for this study was obtained from the institutional review boards of Mayo Clinic and (14-008651; approved November 5th, 2014) Olmsted Medical Center (012-OMC-15, approved March 23rd, 2015). The need for informed consent was waived. This study utilized the resources of the Rochester Epidemiology Project (REP) which is a medical record-linkage system that collects diagnostic codes of all clinical encounters (inpatient, outpatient and emergency room visit) of Olmsted County, Minnesota residents with local providers (the Mayo Clinic, the Olmsted Medical Center, local nursing homes and few private practitioners). The diagnoses made by healthcare providers at each visit are obtained from billing data. The system allows a virtually complete identification of all clinically recognized cases of the disease of interest in the community [5]. The medical record-linkage system was searched to identify all potential adult cases (age >18 years) of sarcoidosis between January 1, 1995 and December 31, 2013 using the International Classification of Diseases, Ninth Revision (ICD-9) code 135 (sarcoidosis). Complete medical records of those potential cases were individually reviewed. The diagnosis of sarcoidosis was confirmed by the presence of non-caseating granuloma on histopathology, radiographic findings of intrathoracic sarcoidosis (bilateral hilar adenopathy and/or interstitial infiltration) and compatible clinical presentations. Patients with evidence other granulomatous diseases such as tuberculosis and histoplasmosis were excluded. The only exception for the histopathological confirmation was stage I pulmonary sarcoidosis that required only the presence of bilateral hilar adenopathy on imaging study. Isolated extra-thoracic sarcoidosis was also included after exclusion of other possible etiologies of granulomatous inflammation. Descriptive statistics were used to summarize the data. Positive predictive value (PPV) was estimated as the number of patients verified to have sarcoidosis divided by the number of patients with a diagnostic code for sarcoidosis. Ninety-five percent confidence intervals (CI) were calculated using the exact binomial method. Logistic regression models

were used to examine differences in PPV according to age, sex and calendar year. Additional analysis was conducted to estimate the PPV for patients with the occurrence of the code for at least twice.

Results

The study cohort included 366 patients with at least one code for sarcoidosis (mean age 49.7 years, 56% female, 85% Caucasian and 9% African-American). Of these, 224 cases of confirmed sarcoidosis were identified, resulting in PPV of 61.2% (95% CI, 56.0% – 66.2%). The PPVs by sex and age group are described in table 1. The PPV for females was significantly lower than males (56.4% vs 67.3%; p=0.034). No significant trends in PPV over calendar year (p=0.18) or age (p=0.55) were observed. A total of 268 patients in the database had a code for sarcoidosis on least two occasions separated by at least 30 days; of these, there were 205 cases of confirmed sarcoidosis. The PPV for having the code at least twice was 76.5% (95% CI, 71.0% – 81.4%).

Discussion

The current study is the first to utilize a population-based cohort to assess the accuracy of diagnostic coding for sarcoidosis. Similar to other pulmonary diseases [2–4], the PPV of ICD-9 coding for sarcoidosis was relatively low, which indicates that misclassification of patients with sarcoidosis is common in coding-based studies. Thus, the validity of the results of such studies depends on how vigorously the diagnosis of sarcoidosis was verified. Verification by individual medical record review is generally associated with the highest accuracy. However, this approach is often not feasible, particularly for studies using large electronic medical databases. The current study suggests that requiring at the occurrence of the diagnostic code at least twice could be a reasonable alternative as it improved the PPV to over 75%, although this approach missed 19 patients (8%) with sarcoidosis.

The PPV of ICD-9 coding for sarcoidosis among females in this study was lower than for males. Different patterns of healthcare utilization between the two sexes may be the contributing factors as 51% males with sarcoidosis in this cohort had pulmonary symptoms whereas only 36% of females had them [6], which suggests that sarcoidosis in females was incidentally found more often than in males.

The major strength of this study is the accuracy of the diagnosis of sarcoidosis that was individually verified by medical record, histopathology and radiographic study review. The population-based design also allows capturing of the full spectrum of the severity of sarcoidosis in the community, unlike referral-based cohort that tends to capture only cases with more severe disease. However, generalizability of the results to other databases could be limited as the pattern of diagnosis and coding could vary between healthcare systems. The clinical manifestations and severity of sarcoidosis varies considerably across ethnic groups [7, 8] and the patients in this study were predominately Caucasians. Finally, PPV is dependent on pre-test probability and could be significantly higher and lower in the populations/databases with higher and lower prevalence of sarcoidosis, respectively.

Conclusion

In conclusion, the PPV of ICD-9 code for sarcoidosis is relatively low and, thus, further verification is required for studies using electronic databases.

Acknowledgments

References

1. Szeto HC, Coleman RK, Gholami, et al. Accuracy of computerized outpatient diagnoses in a veteran affairs general medicine clinic. Am J Manag Care. 2002; 8:37–43. [PubMed: 11814171]

2. Yves, Lacasse, Daigle, JM., Martin, S., et al. Validity of chronic obstructive pulmonary disease diagnoses in a large administrative database. Can Respir J. 2012; 19:e5–9. [PubMed: 22536584]

3. Fang MC, Fan D, Sung SH, et al. Validity of using inpatient and outpatient administrative codes to identify acute venous thromboembolism. The CVRN VTE study. Med Care. 2016 Epub ahead of print.

4. Blais L, Lemiere C, Menzies D, et al. Validity of asthma diagnoses recorded in the medical services database of Quebec. Pharmacoepidemiol Drug Saf. 2006; 15:245–252. [PubMed: 16374899]

5. Rocca WA, Yawn BP, St Sauver JL, et al. History of the Rochester Epidemiology Project: Half a century of medical records linkage in a U.S. population. Mayo Clin Proc. 2012; 87:1202–1213. [PubMed: 23199802]

6. Ungprasert P, Crowson CS, Matteson EL. Influence of gender on epidemiology and clinical manifestations of sarcoidosis: A population-based retrospective cohort study 1976–2013. Lung. 2017; 195:87–91. [PubMed: 27709291]

7. Ramachandraiah S, Aronow W, Chandy D. Pulmonary sarcoidosis: an update. Postgrad Med. 2017; 129:149–158. [PubMed: 27766929]

8. Mirsaeidi M, Machado RF, Schraufnagel D. Racial difference in sarcoidosis mortality in the United States. Chest. 2015; 147:438–449. [PubMed: 25188873]

Table 1

Positive predictive value of at least one code for sarcoidosis

Group	Number of confirmed cases of sarcoidosis	Number of cases with at least one diagnostic code for sarcoidosis	Positive predictive value	95% confidence interval	
Overall	224	366	61.2	56	66.2
Female	115	204	56.4	49.3	63.3
Male	109	162	67.3	59.5	74.4
Age<40 years	55	89	61.8	50.9	71.9
Age 40–49 years	73	116	62.9	53.5	71.7
Age 50–59 years	57	86	66.3	55.3	76.1
Age 60+ years	39	75	52.0	40.2	63.7