Powers and Pitfalls of Using Administrative Data to Study the Epidemiology of Interstitial Lung Diseases

Margaret L. Salisbury

Division of Pulmonary and Critical Care Medicine, Department of Medicine, University of Michigan, Ann Arbor, Michigan

ORCID ID: 0000-0001-8217-6955 (M.L.S.).

Interstitial lung disease affects as many as one in 14 U.S. adults (1). Hypersensitivity pneumonitis (HP) is prevalent among these individuals, but few studies have estimated its incidence or prevalence in the general population. In the United Kingdom, incidence was estimated at 0.9 per 100,000 person-years, using a computerized longitudinal primary care database, with incident cases defined as those with a diagnosis of “extrinsic allergic alveolitis” entered after at least 1 year of database enrollment (2). Another study estimated the incidence of various interstitial lung diseases, using an interstitial lung disease registry and population data in a single county in New Mexico, finding incident interstitial lung disease in males at a rate of 31.5 per 100,000 person-years and females at 26.1 per 100,000; incident HP cases accounted for 1.5% of all interstitial lung disease (3).

In this issue of the *AnnalsATS*, Fernández Pérez and colleagues (pp. 460–469) use a commercial insurance database (MarketScan) to estimate the yearly (2004–2013) prevalence and cumulative incidence of HP (4). Prevalent HP cases had at least two *International Classification of Diseases, Ninth Revision* (ICD-9)-based HP claims in 12 months, without claims for select alternative interstitial lung diseases (notably, idiopathic pulmonary fibrosis was not excluded). Yearly HP prevalence ranged from 1.67 to 2.71 cases per 100,000 persons during the decade-long study. Extrapolating this estimate to the entire US population, the authors estimate

that 5,300–8,700 persons are living with HP in the United States. Chronic HP (identified when ICD-9 or prescriptions claims suggested disease activity for 6 or more months, or presence of a pulmonary fibrosis ICD-9 claim) had prevalence ranging from 0.91 to 1.70 per 100,000. Roughly half of all identified HP cases were considered chronic. Fibrotic HP (the subset of chronic HP with a pulmonary fibrosis claim) prevalence was 0.41–0.80 per 100,000.

The accuracy of the prevalent HP case definition was validated within the National Jewish Health database in two ways. First, several different HP case definition algorithms were evaluated among all adults seen at National Jewish Health during a 5-year period. Algorithm accuracy was assessed in a subset of 40 patients with HP identified by each algorithm, with gold standard HP diagnosis assigned via expert multidisciplinary case review. The “Algorithm A” definition (described here) produced the lowest false-positive rate (5%, or 2 of 40) and was therefore used in the MarketScan dataset. Algorithm A definitions for chronic and fibrotic HP were also associated with low false-positive rates of 6.3% and 0%, respectively. Notably, the sensitivity or false-negative rate (i.e., the proportion of patients with chronic or fibrotic HP missed by the algorithm) was not evaluated. Second, the Algorithm A HP case definition was applied to 80 patients selected from the National Jewish research databases (40 with HP and 40 with non-HP lung disease) and judged against the same

gold standard. Here, the case definition had 100% sensitivity and 95% specificity for gold standard HP. In this analysis, the calculated positive predictive value (i.e., the probability of having HP after identification as HP by the algorithm) is likely inflated by a high HP prevalence (50%).

Although validation of the ICD-9–based HP case definition is a clear strength of this study, it was performed only in the National Jewish Health database. The same group of interstitial lung disease experts adjudicating the gold standard HP diagnoses also assigned the HP ICD-9 codes within the database, potentially inflating the test characteristics of the algorithm definition as a result of similar practice across the group. Accuracy and sensitivity may differ if applied in a community or general pulmonary setting, including the MarketScan database to which it was applied in this study. For reference, an ICD-9 code-based algorithm for identification of incident and prevalent idiopathic pulmonary fibrosis was found to have a positive predictive value of 42% and a sensitivity of 55%, based on a multidisciplinary gold standard case definition applied to insurance claims data from a separate institution (5). The positive predictive value-corrected idiopathic pulmonary fibrosis incidence estimate, using an inclusive case-definition algorithm, was 5.6 per 100,000 person-years compared with an uncorrected incidence of 47 per 100,000 (5). Future development of a similar correction factor for a claims-based HP algorithm using an external database

(Received in original form December 29, 2017; accepted in final form January 3, 2018)

M.L.S. is supported by T32 HL007749-21.

Correspondence and requests for reprints should be addressed to Margaret L. Salisbury, M.D., University of Michigan, Pulmonary and Critical Care Medicine, 1500 East Medical Center Drive, 3916 Taubman Center, Ann Arbor, MI 48109-5360. E-mail: msalisbu@med.umich.edu.

Ann Am Thorac Soc Vol 15, No 4, pp 424–425, Apr 2018

Copyright © 2018 by the American Thoracic Society

DOI: 10.1513/AnnalsATS.201712-965ED

Internet address: www.atsjournals.org

may improve on estimates made by the current study.

The yearly cumulative incidence of HP was also estimated in this study. Incident cases had an HP claim “clean period,” defined by at least 6 months of health plan enrollment before the year containing the first HP claim without any HP claims during that time. HP incidence is estimated at 1.28–1.94 cases per 100,000 person years. The case definition for incident HP was not subjected to validation and may be inaccurate in an open population using administrative data. Of note, among the

4,680 prevalent HP cases in whom incidence could be assessed (meaning they were enrolled for 6 months during the preceding year), 95% met criteria for incident HP. This estimate seems incongruent with the idea that half of HP cases are chronic, and the median survival in patients even with fibrotic HP is at least 6 years. As noted by the authors, incidence would be better measured in a population-based study, and ideally a closed cohort.

In conclusion, this study by Fernández Pérez and colleagues estimates the

prevalence and incidence of hypersensitivity pneumonitis in a large database of persons with commercial health insurance coverage across the United States (4). Despite limitations, this estimate of the burden of HP, often a chronic disease, provides a vital piece of information on which future investigations of healthcare costs or efforts to better distribute resources may build. ■

Author disclosures are available with the text of this article at www.atsjournals.org.

References

- 1 Putman RK, Hatabu H, Araki T, Gudmundsson G, Gao W, Nishino M, *et al.*; Evaluation of COPD Longitudinally to Identify Predictive Surrogate Endpoints (ECLIPSE) Investigators; COPDGene Investigators. Association between interstitial lung abnormalities and all-cause mortality. *JAMA* 2016;315:672–681.
- 2 Solaymani-Dodaran M, West J, Smith C, Hubbard R. Extrinsic allergic alveolitis: incidence and mortality in the general population. *QJM* 2007;100:233–237.
- 3 Coultas DB, Zumwalt RE, Black WC, Sobonya RE. The epidemiology of interstitial lung diseases. *Am J Respir Crit Care Med* 1994;150:967–972.
- 4 Fernández Pérez ER, Kong AM, Raimundo K, Koelsch TL, Kulharni R, Cole AL. Epidemiology of hypersensitivity pneumonitis among an insured population in the United States: a claims-based cohort analysis. *Ann Am Thorac Soc* 2018;15:460–469.
- 5 Ley B, Urbania T, Husson G, Vittinghoff E, Brush DR, Eisner MD, *et al.* Code-based diagnostic algorithms for idiopathic pulmonary fibrosis. case validation and improvement. *Ann Am Thorac Soc* 2017;14:880–887.