Identification and Confirmation of IPF Cases in an Electronic Insurance Claims Database

Daina Esposito, MPH[1], Stephan Lanes, PhD, MPH[1], Gaurav Deshpande, PhD[1], Crystal N. Holick, ScD, MPH[1], Daniel Mines, MD, MSCE[1], Sean O'Quinn, MPH[2], Trung N. Tran, MD, PhD[2]

[1]HealthCore, Inc., Wilmington, Delaware; [2]MedImmune, LLC, Gaithersburg, Maryland

BACKGROUND

- Idiopathic pulmonary fibrosis (IPF) is a progressive and irreversible interstitial lung disease (ILD) with limited pharmacologic therapy options and a dismal prognosis.
- Epidemiologic data on IPF are limited. Estimates of the prevalence of this rare condition range widely from 2 to 43 per 100,000 in the general population.
- Clinical definitions have evolved over time. The most recent consensus guidelines (2011) require exclusion of other known causes of ILD and either (1) the presence of a usual interstitial pneumonia (UIP) pattern on high-resolution computer tomography (HRCT) in patients not subjected to lung biopsy or (2) specific combinations of HRCT and lung biopsy patterns. Diagnosis also requires input from pulmonologists, radiologists and pathologists.
- While some studies have sought to identify IPF in administrative databases, many studies precede development of the most recent guidelines for diagnosis, and the performance of the algorithms applied is unknown.
- Developing and estimating the positive predictive value (PPV) of clinically sensible algorithms to identify IPF cases in a large claims database will improve validity of observational studies of IPF in administrative databases and allow for better understanding of IPF incidence, prevalence, and patient characteristics.

OBJECTIVES

- To estimate the incidence and prevalence of IPF in the US population.
- To assess the positive predictive value (PPV) of claims-based algorithms to identify IPF (work is on-going).

METHODS

- In collaboration with a panel of pulmonary disease experts, we developed two algorithms to identify IPF cases:
  1. Sensitive algorithm (Figure 1):
     - At least one diagnosis of IPF made by a physician
     - Patient age ≥50 years
     - No diagnoses indicative of alternative causes recorded after the date of the last recorded diagnosis of IPF but within six months of the first physician-assigned diagnosis of IPF
  2. Specific algorithm (Figure 2):
     - Met criteria for the sensitive algorithm
     - At least one diagnosis of IPF was made by a pulmonologist
     - For incident IPF cases (i.e., patients with at least 12 months of health plan eligibility prior to the first diagnosis of IPF):
       - Diagnosis of IPF recorded after HRCT or lung biopsy AND at least two of the following criteria are met:
         1. Age ≥65 years
         2. At least two diagnoses at least three months apart
         3. At least one hospitalization with IPF as the principal diagnosis
         4. At least one claim for pulmonary function tests
         5. Diagnosis at least seven days after antinuclear antibody (ANA) or rheumatoid factor (RF) serology
         6. Lung transplantation
     - For prevalent IPF cases (i.e., patients with less than 12 months of health plan eligibility prior to the first diagnosis of IPF), at least two of the 7 criteria (criteria 1 to 6 mentioned above and diagnosis of IPF recorded after HRCT or lung biopsy).
- Both algorithms were applied to the HealthCore Integrated Research Database (HIRD℠), which includes claims information from one of the largest commercially insured populations in the US.
- We characterized cases in terms of demographic characteristics, comorbidities, and factors that qualify patients as probable IPF cases.
- Incidence and prevalence of IPF were calculated based on each case definition, stratified by age and gender.
- Confirmation of case status using medical record review is in progress. We are currently completing the first stage of a two-stage approach. A sample of 100 medical records of patients that met IPF cases definitions as described in Figures 1 and 2 are undergoing review by a panel of three experts in ILD. Based on findings from this round of review, we will refine our specific case definition to improve the algorithm's PPV. We will then determine the PPV of the data-derived algorithm in a new group of 50 medical records. We will apply the PPV of the sensitive algorithm to correct our estimates of incidence and prevalence.

Figure 1. Flowchart for identification of patients with IPF, sensitive case definition

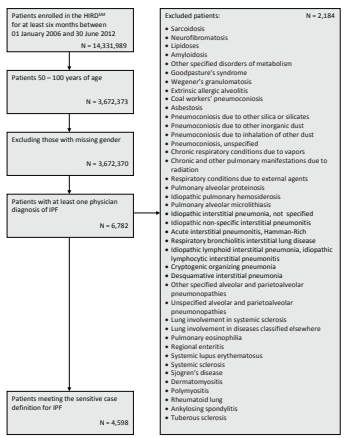
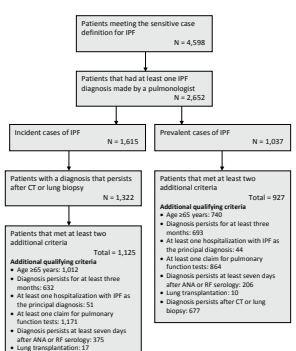


Figure 2. Flowchart for identification of patients with IPF, specific case definition



RESULTS

- From January 2006 through June 2012, 4,598 IPF patients were identified using the sensitive algorithm; among them, 2,052 (44.6%) met the definition for the specific algorithm.
- The mean age of cases meeting the sensitive algorithm was 73 years, and approximately half were male. Common comorbidities included coronary artery disease, pneumonia, chronic obstructive pulmonary disease and type 2 diabetes mellitus.
- The prevalence of IPF (2006 through 2012) was approximately 125.2 per 100,000 based on the sensitive algorithm and 55.9 per 100,000 based on the specific algorithm.
- Incidence was 31.9 per 100,000 person years (95% CI 30.9 – 32.9) and 12.5 per 100,000 person years (95% CI 11.9 – 13.1) based on the sensitive and specific algorithm respectively.
- Incidence and prevalence increased with advancing age.

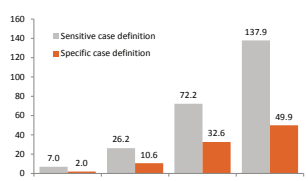
Table 1. Baseline characteristics of patients with IPF

	Sensitive case definition		Specific case definition	
	N	%	N	%
Total	4,598	100.0	2,052	100.0
Gender				
Male	2,303	50.1	1,095	53.4
Female	2,295	49.9	957	46.6
Age (mean, std, median)	73.1 (10.9), 74.0		73.1 (9.9), 74.0	
Age group				
50-59 years	609	13.2	217	10.6
60-69 years	1,094	23.8	497	24.2
70-79 years	1,459	31.7	748	36.5
≥80 years	1,436	31.2	590	28.8
US geographic region				
Northeast	856	18.6	261	12.7
South	826	18.0	385	18.8
Central	1,906	41.5	966	47.1
West	921	20.0	408	19.9
Unknown	89	1.9	32	1.6
Incident diagnosis	2,879	62.6	1,125	54.8
Comorbidities				
Pulmonary hypertension	491	10.7	223	10.9
Pulmonary embolism	186	4.0	77	3.8
Lung cancer	136	3.0	56	2.7
Coronary artery disease	1,629	35.4	737	35.9
Gastroesophageal reflux disease	1,126	24.5	508	24.8
Chronic obstructive pulmonary disease	1,218	26.5	522	25.4
Type 2 diabetes mellitus	1,242	27.0	521	25.4
Pneumonia	1,456	31.7	624	30.4
Lung infections	136	3.0	56	2.7

Table 2. Prevalence of IPF by case definition, age, and gender

	IPF diagnoses	Population at risk	Prevalence per 100,000 patients
Overall, broad case definition	4598	3,672,370	125.2
Gender			
Male	2,303	1,736,074	132.7
Female	2,295	1,936,296	118.5
Age			
50 - 59 years	609	1,974,076	30.8
60 - 69 years	1,094	1,023,997	106.8
70 - 79 years	1,459	429,338	339.8
≥80 years	1,436	244,959	586.2
Overall, restrictive case definition	2052	3,672,370	55.9
Gender			
Male	1,095	1,736,074	63.1
Female	957	1,936,296	49.4
Age			
50 - 59 years	217	1,974,076	11.0
60 - 69 years	497	1,023,997	48.5
70 - 79 years	748	429,338	174.2
≥80 years	590	244,959	240.9

Figure 3. Incidence of IPF per 100,000 person-years by age (years) and case definition



DISCUSSION

- Characterization of the incidence and prevalence of IPF using administrative data requires clinically sensible, validated case finding approaches. Given the challenges of studying such a rare condition, this is an important unmet need.
- We developed two algorithms to identify IPF cases in order balance false positive and false negative errors. One aims for higher sensitivity, and the other aims to improve the PPV based on requiring appropriate patent factors and evidence of a clinically sensible process of care seen in the automated data. This approach allows us to produce a reasonable range of plausible values of incidence and prevalence.
- Findings from the confirmation of our algorithms currently in-progress will provide refinement of the unconfirmed estimates presented.

Presented at OMOP-IMEDS Symposium 2013 | November 5 – 6 | Bethesda, Maryland, USA