Respiratory Research

RESEARCH

Open Access

Epidemiology of idiopathic pulmonary fibrosis in central and Western Pennsylvania

Faina Linkov[1*†], Yue-Fang Chang[2†], Harshitha Ramanan[1], Richard S. Morgan[6], Kathleen M. McTigue[3], Anne E. F. Dimmock[4], Rebecca Bascom[4†] and Daniel J. Kass[5†]

Abstract

Background/rationale Idiopathic Pulmonary Fibrosis (IPF) is a chronic, progressive disease of unknown origin. Establishing the epidemiology of IPF has been challenging due to diagnostic complexity, poor survival, low prevalence, and heterogeneity of ascertainment methodologies.

Objectives This research aimed to estimate the rates of IPF in central and western Pennsylvania and to pilot the use of capture recapture (CR) methods to estimate the disease incidence.

Methods We identified adults ≥ 30 years old diagnosed with IPF (by ICD-9/10 coding) between 2013 to 2021 from two health systems (UPMC Health System and Penn State Health) participating in the PaTH Clinical Research Network. We extracted information on patients' sex, race, date of birth and 3-digit zip code from electronic health records (EHR). Incidence rate of IPF among Pennsylvania residents was calculated using three case definitions (broad and two restricted) and piloted the use of CR in estimating IPF incidence.

Results IPF incidence rates were 8.42, 6.95 and 4.4 per 100,000 person-years for the unrestricted (n = 3148), partially restricted (n = 2598) and fully restricted (n = 1661) samples, respectively. Low case overlap between two sites resulted in a highly inflated estimate of IPF incidence, using the CR methodology.

Conclusions The rate of IPF in central and western Pennsylvania was similar to previously published statistics. The application of CR to IPF epidemiology could be further investigated in health systems with greater overlap of patients utilizing more than one system.

Keywords Epidemiology, Interstitial lung disease

†Faina Linkov and Yue-Fang Chang Chang have equal contribution.

†Rebecca Bascom and Daniel J. Kass have equal contribution.

*Correspondence:
Faina Linkov
Linkovf@duq.edu
[1] Department of Health, Exercise & Applied Science, John G. Rangos Sr. School of Health Sciences, Duquesne University, Pittsburgh, PA, USA
[2] Department of Neurological Surgery, University of Pittsburgh, Pittsburgh, PA, USA
[3] Department of Medicine, University of Pittsburgh, Pittsburgh, PA, USA
[4] Penn State College of Medicine, Penn State Milton S. Hershey Medical Center, Hershey, PA, USA
[5] Dorothy P. and Richard P. Simmons Center for Interstitial Lung Disease, University of Pittsburgh, Pittsburgh, PA, USA
[6] Department of Biomedical Informatics, University of Pittsburgh, Pittsburgh, PA, USA

© The Author(s) 2025. **Open Access** This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by-nc-nd/4.0/.

Introduction

Idiopathic pulmonary fibrosis (IPF) is a progressive and devastating disorder characterized by unremitting progressive scarring in the lung that leads to breathlessness, respiratory failure, and death [1, 2]. The estimates of the incidence of IPF have varied widely, depending on the case definitions, ranging from less than one to greater than 90 per 100,000 per year [3]. The prevalence of IPF is increasing (especially in adults over 65, as well as in veterans) [3–5], however, it is unknown whether this is due to increased recognition [6], environmental or other risk factors, or changing definitions. In the U.S., the prevalence has been estimated at 10 to 60 cases per 100,000 [6], meeting the definition of rare disease. IPF is a highly lethal condition with an average lifespan after diagnosis of 3–5 years from the time of diagnosis [7]. There is great variation in IPF rates among various US states likely related to environmental factors [8] or differential ascrtainment.

Previous studies that have employed mortality statistics suffer from diagnostic misclassification [5]. Furthermore, IPF has undergone several changes in consensus diagnostic criteria over the last 20 years which suggest that evolving definitions could affect these numbers [9–11]. A surge in our understanding of IPF pathogenesis including the discovery of the role of genetic risk factors [12–16], the increasing recognition of pre-morbid "interstitial lung abnormalities," the role of aging [17–19], and the availability of therapies [20, 21] all point to the need to study IPF at the epidemiologic level.

IPF has been linked to exposure to certain types of dust, viral infections, genetics, and smoking [22]. Our group recently reported that exposure of IPF patients to particulate matter 2.5 μm or less in diameter ($PM_{2.5}$) was associated with worse lung function and higher mortality. And of the components of $PM_{2.5}$ sulfate, ammonium, and black carbon were associated with the greatest risk of death [23, 24], highlighting the need for reductions in human-derived sources of pollution especially in Pennsylvania. Air pollution has been very concerning in Pennsylvania for decades due to a high number of industrial plants and transportation crossroads [25].

In this study, we sought to describe the epidemiology (incidence and geographic distribution) of IPF that is specific to central and western Pennsylvania. The secondary aim of this study was to pilot the application of the Capture-Recapture (CR) methodology to estimate the incidence of IPF in Pennsylvania. CR was originally developed in the field of zoology to estimate the size of a closed wild animal population [26]. Different samples of animals are captured, counted, tagged and released. The size of the total population can then be estimated based on the prevalence of tagged animals appearing in subsequent samples. The same approach has been employed for disease monitoring. By comparing existing data from several independent sources and identifying the number of overlapping cases, it is possible to estimate the number of missing cases and generate estimates of rates for conditions of interest (in this case, IPF). Pittsburgh based investigators have used the Capture-Recapture methodology to assess the incidence of Type I diabetes in adolescent population of Allegheny County, to monitor injuries, dog bites, amputation rates, and several other conditions [26–35]. To our knowledge, CR has not been applied to the study of IPF. Our group was well equipped to undertake this project due to our long-term collaboration with the PaTH network, a Partner Network in PCORnet® which has been developed with funding from the Patient-Centered Outcomes Research Institute®. PCORnet is a national resource, funded by PCORI, in which high-quality health data, patient partnership, and research expertise are harnessed to enhance research and improve healthcare outcomes. PaTH has already been utilized by several initiatives in the area of IPF, including the evaluation of a computable IPF phenotype [36, 37].

Patients and methods

The University of Pittsburgh and Penn State University Institutional Review Boards designated this study exempt non-human study research (Pitt IRB STUDY19080231 and Penn State University IRB #STUDY00017636 (exempt as non-human subject research)).

PaTH network: source of data

PaTH is the name of the network that is comprised of 9 US health system and affiliated academic sites [38]. This study focuses on two PaTH academic medical centers in the mid-Atlantic region: UPMC (including 40 hospitals) and Penn State University (including 6 hospitals). Methodologically, the PaTH network was specifically developed to help characterize several targeted conditions, including IPF [39]. For this project, PaTH extracted longitudinal EHR data reflecting inpatient and outpatient settings from the two relevant health systems, UPMC and Penn State Health.. PCORnet Common Data Model instances, including patient's diagnosis (ICD-9-CM and ICD-10-CM codes), first recorded date of diagnosis recorded in the system, and encounters with healthcare providers. Additionally, patient's sex, race, date of birth and 3-digit zip code were obtained. These data elements were vetted by informatics teams at each participating site.

IPF case definition

In keeping with the question of diagnostic uncertainty with IPF as well as the shift in the field to so-called "lumping" of cases based on the shared features of a progressive phenotype [40], we have employed three definitions of IPF cases based on ICD 9/10 coding. Similar approaches have been taken in prior studies of IPF epidemiology [19, 36, 41–43] to provide a range of incidence rates. The analyses included all Pennsylvania IPF cases except those from 3-digit zip codes 180–196 since this area is considered as Eastern Pennsylvania and both healthcare systems utilized in this research do not typically serve population of eastern counties around Philadelphia metropolitan area (they are served by different healthcare systems in the area). Number of cases for zip codes 180–196 are included in Suppmental materials.

Unrestricted cases

We analyzed electronic health record data from patients undergoing care at the UPMC and Penn State Health facilities between 2013 and 2021. To improve sensitivity, we included the broader ICD-9 diagnosis codes 516.31 and its ICD-10 equivalents (J84.112) without applying any restrictions (coded as "unrestricted cases"). Record selection flow is outlined in Fig. 1.

Partially restricted cases

Additionally, we also identified "partially restricted" cases which considered a limited series of exclusionary diagnosis codes that are often ruled out during the initial workup of IPF (Appendix 1) except J84.89 (other specified interstitial pulmonary disease) and J84.9 (interstitial pulmonary disease, unspecified).

Fully restricted cases

We identified all patients who had an *International Classification of Disease* (ICD) diagnosis code for IPF (ICD-9-CM code 516.31or ICD-10-CM code J184.112) but excluded patients who also had any of the full series of exclusionary diagnosis codes (see Supplemental Material 1) as "IPF restricted cases."

Pilot investigation: capture recapture

A two-source capture-recapture method (see Fig. 2 for the methodology concept and calculation formula) was used to estimate IPF incidence after matching cases from 2 different sources: (1) UPMC/University of Pittsburgh School of Medicine (UPMC/Pitt) and (2) Penn State Hershey Medical Center/Penn State College of Medicine (Penn State Health) using date of birth, sex and 3-digit zip code. Estimated source-specific ascertainment rates

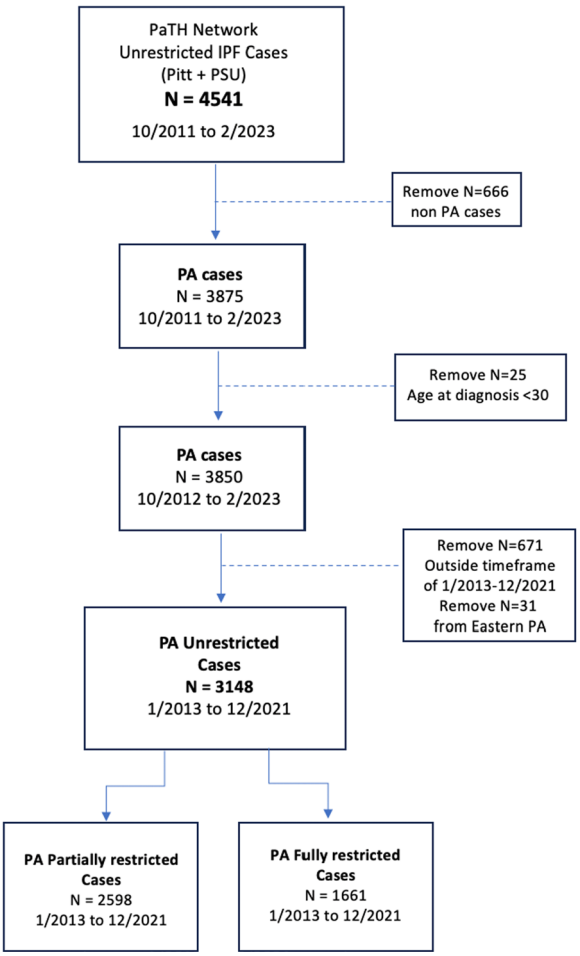

Fig. 1 Study flow

were defined as the number of IPF cases in each data source divided by the estimated number of IPF cases by capture-recapture analysis, expressed as a percentage.

Data analysis approaches

The analyses were based on IPF cases from Pennsylvania with diagnosis dates between 2013 to 2021 and age greater or equal to 30 to make our study sample more consistent with the typical presentation of the IPF patients across the US. For patients with multiple entries, the one with the earliest diagnosis date was used. The database had patients from October 2011 to February 2022. Incident cases were defined as patient whose first IPF diagnosis record was on or after January 1, 2013 and on or before December 31, 2021. IPF cases were identified from UPMC/Pitt and Penn State separately, and the cases from the 2 data sources were then matched by date of birth, sex and 3-digit zip code. Descriptive analyses were carried out to summarize the IPF cases. The IPF incidence rate from 2013 to 2021 was estimated as the

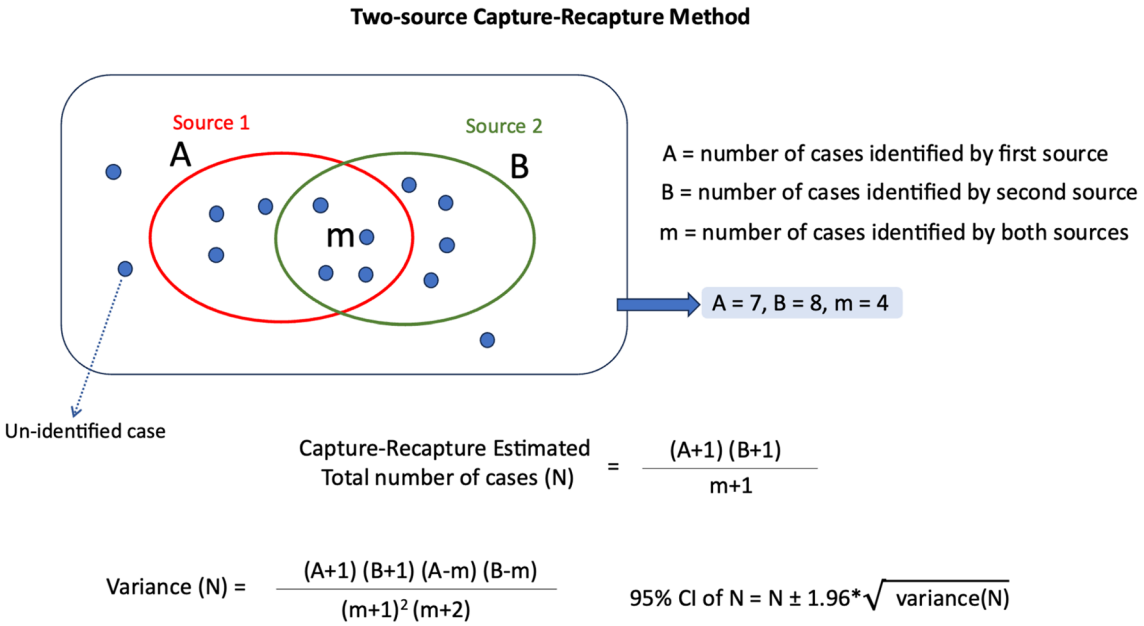

Fig. 2 Two-source capture-recapture method summary

number of IPF incidence cases in this period divided by the total number of person-years of observation, using PA census data.

Results

From 2013 to 2021, there were and 2739 incident cases of IPF using the unrestricted definition, 2279 incident cases using the partially restricted definition and 1520 incident diagnoses of IPF using the fully restricted case definition, at Pitt, and 413 incident cases of IPF using the unrestricted definition, 321 using partially restricted case definition, and 141 incident diagnoses of IPF using the fully restricted case definition, and at PSU. Four unrestricted cases and 2 partially restricted were identified by both Pitt and PSU, whereas there were no overlapping fully restricted cases. Therefore, the Pitt and PSU databases resulted in a total of 3148 unrestricted, 2598 partially restricted and 1,661 fully restricted IPF cases (Fig. 1). Using PA census data, the Pennsylvania IPF incidence rates from 2013 to 2021 were 8.42/100,000, 6.95/100,000 and 4.44/100,000 for the cases using unrestricted, partially restricted, and fully restricted definitions, respectively. Demographics were similar between the three case definition cohorts (Tables 1 and 2). There were more male and white patients, and the majority were diagnosed between the ages of 60 to 80.

Geographic distribution

The incidence rates were estimated according to the 3-digit zip codes. The geographic burden of IPF was heterogeneous across Pennsylvania, with the highest burden recorded for Allegheny County and Central Pennsylvania (zip codes 150–159, Table 3, Supplemental Table 1, Fig. 3). A similar geographic distribution was seen when both broad and restricted case definitions were used.

< Insert Table 3. IPF cases and incidence rates (per 100,000 person-years) from 2013 to 2021 by groups of 3-digit zip code. >

Trend of IPF cases

We observed a decrease trend in the IPF incidence rate from 2013 to 2021 in all 3 case definitions ($p < 0.0001$). The annual incidence rate for the unrestricted case was 9.46 per 100,000 person-years in 2013 and decreased to 7.39 per 100,000 person-years in 2021 (Fig. 4, Supplemental Table 2).

Capture recapture results

To account for the un-captured IPF cases in Western and Central Pennsylvania, we used the capture-recapture method to estimate the total number of IPF cases (Fig. 2). We observed a very small overlap of cases (4 for the unrestricted sample, 2 for the less restricted sample and 0 for the fully restricted sample). Because of the small number of overlapped IPF cases identified from both Pitt and Penn State, the capture-recapture method resulted in a highly inflated number of incident cases: 226,872 unrestricted cases; 244,720 partially restricted cases; and 215,982 fully restricted cases.

Table 1 PA IPF cases by sex, race and age at diagnosis (N(%))

		Unrestricted		Partially restricted		Fully restricted	
		N = 3148		N = 2598		N = 1661	
		n	%	n	%	n	%
Sex	*Female*	1335	42.41	1080	41.57	672	40.46
	Male	1813	57.59	1518	58.43	989	59.54
Race	*American Indian*	2	0.06	0	0.00	0	0.00
	Asian	21	0.67	17	0.65	8	0.48
	Black	157	4.99	103	3.96	67	4.03
	White	2837	90.12	2366	91.07	1510	90.91
	Other	25	0.79	20	0.77	11	0.66
	no information	106	3.37	92	3.54	65	3.91
Age at diagnosis	30- < 40	28	0.89	21	0.81	16	0.96
	40- < 50	90	2.86	54	2.08	40	2.41
	50- < 60	293	9.31	208	8.01	120	7.22
	60- < 70	802	25.48	630	24.25	375	22.58
	70- < 80	1142	36.28	960	36.95	607	36.54
	80- < 85	378	12.01	339	13.05	233	14.03
	> = 85	415	13.18	386	14.86	270	16.26

Discussion

We estimated the incidence of IPF in central and western Pennsylvania regions and found them to be similar to the the national data. This is one of the first studies to estimate the rates of IPF in Pennsylvania. A high concentration of cases encompassing southwest Pennsylvania may reflect the proximity of patients to the ILD specialty center in the Pittsburgh region.

Our group recently identified that patients who live in western Pennsylvania have a significantly increased risk of death, which is correlated with concentrations of PM2.5 [23]. This raised the question of whether or not residence in Pennsylvania also represented an increased risk for developing IPF. While the previous study employed diagnosis based on medical records and precise estimation of $PM_{2.5}$ exposure by nine digit zip code, in the present study we identified the incident cases of IPF by three digit zip code.

The reasons why we encountered regional variation based on three digit zip code are unclear. In addition to PM2.5 concentration [44], we can speculate that geographic differences in the incidence and prevalence of IPF may be explained by multiple factors, including diagnostic challenges, varied diagnostic criteria, and differences in study methodologies. Furthermore, the persistence of certain pathologic genetic variants that are associated with pulmonary fibrosis, in discrete regions in Pennsylvania, is likely a contributor [45]. Future prospective studies with targeted pollution and genetics mapping may help identify the genetic and environmental factors that lead to pulmonary fibrosis.

We calculated an incidence of IPF ranging from 4.4 to 8.4 per 100,000 person-years. This is comparable to incidence rates observed in multiple prior studies [23, 46–48]. Our incidence estimates are, however, lower than seen in other studies capturing data from early 2000's [19, 43]. The methodologies for studying the incidence of IPF have included reviews of national databases [46–48], insurance claims [19, 43], and careful but small regional analyses [42, 49].

National database samples suggested estimates ranging from 4.6 to 8.7 per 100,000 [3], while the analysis of insurance claims, emphasizing how the numbers can vary based on case definitions identified 6.8 (narrow) to 16.3 (broad) per 100,000 person-years [3, 19]. Taken together, these data suggest that the incidence of IPF in central and western Pennsylvania is comparable to what has been observed elsewhere.

Consistent with previously published research, we observed a higher incidence of IPF in men and in individuals older than 60. Coultas et.al. estimated an incidence of IPF to be 31% higher in men [49]. Large database studies may be associated with significant misclassification, while local studies may suffer from small numbers and lack of generalizability [5].

As our study utilized three different diagnostic critieria, it is important to comment on which diagnostic classification is more practical for identifying incidence

Table 2 IPF cases and incidence rates (per 100,000 person-years) from 2013 to 2021 by age at diagnosis and sex*

(a). Unrestricted cases

Age	All		Female		Male	
	# IPF	*Rate/100,000*	*# IPF*	*Rate/100,000*	*# IPF*	*Rate/100,000*
30-<40	*28*	*0.41*	*16*	*0.48*	*12*	*0.35*
40-<50	90	1.08	51	1.23	39	0.94
50-<60	293	3.28	130	2.88	163	3.69
60-<70	802	12.67	325	9.87	477	15.70
70-<80	1142	28.95	453	20.57	689	39.56
80-<85	378	24.08	163	17.03	215	35.08
>=85	415	27.31	197	18.94	218	45.46
total	3148	8.42	1335	6.84	1813	10.15

(b). Partially restricted cases

Age	All		Female		Male	
	# IPF	*Rate/100,000*	*# IPF*	*Rate/100,000*	*# IPF*	*Rate/100,000*
30-<40	21	0.31	12	0.36	9	0.26
40-<50	54	0.65	30	0.72	24	0.58
50-<60	208	2.33	88	1.95	120	2.72
60-<70	630	9.95	255	7.74	375	12.34
70-<80	960	24.34	368	16.71	592	33.99
80-<85	339	21.59	147	15.36	192	31.33
≥85	386	25.40	180	17.30	206	42.96
Total	2598	6.95	1080	5.54	1518	8.49

(c). Fully restricted cases

Age	All		Female		Male	
	# IPF	*Rate/100,000*	*# IPF*	*Rate/100,000*	*# IPF*	*Rate/100,000*
30-<40	16	0.24	9	0.27	7	0.20
40-<50	40	0.48	21	0.51	19	0.46
50-<60	120	1.35	48	1.06	72	1.63
60-<70	375	5.92	150	4.55	225	7.41
70-<80	607	15.39	220	9.99	387	22.22
80-<85	233	14.84	99	10.34	134	21.87
≥85	270	17.77	125	12.02	145	30.24
Total	1661	4.44	672	3.44	989	5.53

* One patients without age at diagnosis

Table 3 IPF cases and incidence rates (per 100,000 person-years) from 2013 to 2021 by groups of 3-digit zip code

Zip code	*Unrestricted cases*		*Partially restricted cases*		*Fully restricted cases*	
	# IPF cases	Incidence rate *per 100,000*	# IPF cases	Incidence rate *per 100,000*	# IPF cases	Incidence rate *per 100,000*
150–159	1733	11.53	1412	9.40	943	6.28
160–169	900	10.76	768	9.18	501	5.99
170–179	515	3.57	418	2.89	217	1.50

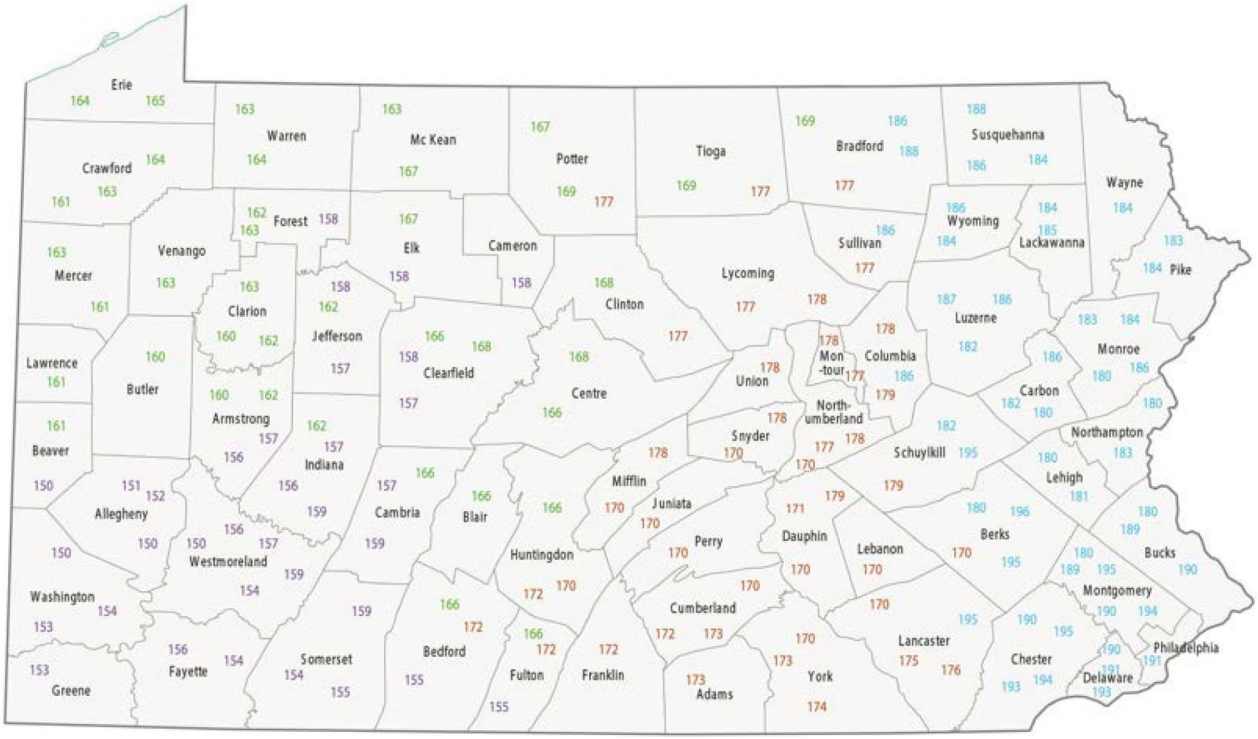

Fig. 3 Pennsylvania county and 3-digit zip code map >

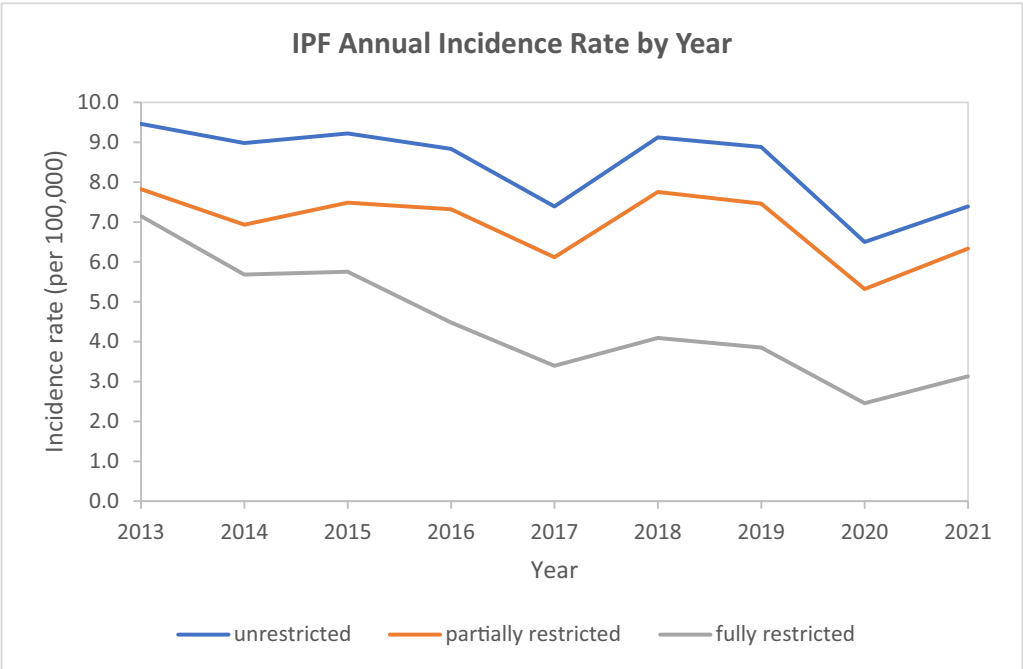

Fig. 4 Incidence trends

and prevalence of IPF moving forward. The pulmonary fibrosis field is moving towards less restrictive criteria for diagnosis which acknowledges that a diagnosis of IPF is a diagnosis of exclusion and with relatively poor interobserver agreement [50]. The use of the term "progressive fibrosing interstitial lung disease" (PF-ILD) recognizes the shared physiology of patients with pulmonary fibrosis of diverse etiologies. The authors of this paper see value in the restrictive diagnosis because certain etiologies, if properly identified, may be more amenable to treatment. However, to better align with the trends in the field, we recommend using the partially restricted case definition. This is consistent with the overall direction of the field towards less restrictive definition highlighted in the article by Raghu et. Al [51].

Key strengths of our study were the utilization of reliable databases for data collecting. We ensured that the resulting cohort of patients included in the analysis was consistent with the typical presentation of the IPF patients across the US in terms of age and diagnostic criteria. Employing data obtained through the PaTH network facilitated identical data queries at both Pitt and Penn State. The use of CR is predicated on the idea that a subject can seek care and be counted in either dataset with an equal probability. To our surprise, there was very little overlap of patients between UPMC and Penn State Health. This lack of overlap resulted in an inflated estimate of the IPF incidence estimate by CR in Pennsylvania. In general, greater overlap of the cases between data sources suggests more accurate estimation. Was the lack of overlap the consequence of insurance plans that limit the opportunities to cross health care systems? Regions of Pennsylvania are highly rural, and this imposes multiple burdens on the identification and care of IPF [52]. We can speculate that the highly rural population of Pennsylvania has limited access to consultations across health care systems for second opinions [53]. Furthermore, since patients with IPF are typically older, a trip of several hundred miles to a specialty center may not be feasible. Rural status has been associated with lower pulmonary function and increased dyspnea in patients with IPF [54]. Despite the limitation of implementing the CR method in our study, with the proper application, CR offers the potential to reduce the costs associated with running disease registries and limit bias in the estimation of incidence [55]. It may be possible to employ the same approach across a broader range of PCORnet sites to determine if CR may reflect a more realistic estimate of the epidemiology of IPF. One possible way to improve CR estimate is to incorporate other data sources, such as insurance records and/or death certificates. Also, our future studies will further investigate the factors that contributed to the poor overlap among patient groups.

Our study has a number of limitations. While we attempted to obtain a range of incidence by introducing several ICD-9/10 based case definitions of IPF, our approach may be limited by the misclassification of IPF cases. Furthermore, as noted above, without nine-digit zip codes, we were unable to determine if regions of poor air quality were associated with an increased incidence of IPF. Our study was also limited by lack of access to data on pulmonary function or on mortality. Such studies will be quite important in the coming years as clinicians develop enough experience with the FDA-approved therapies (pirfenidone and nintedanib). These therapies have been shown to slow the loss of forced vital capacity in patients with IPF [20, 21]. It is unknown, however, if the introduction of these therapies will impact mortality rates at a population level. Again, such studies may be possible through common data extraction methods employed by PCORnet sites.

In conclusion, while regional differences may exist, we report that the incidence of IPF in western and central Pennsylvania is similar to that reported previously in the literature. We modeled the application of CR to estimate the incidence of IPF. While the CR approach in this study had limited utility, we suggest that the methodology is worth considering to estimate the incidence of IPF in the entire PATH netowork and also for other interstitial lung diseases. IPF epidemiology in Pennsylvania need to be further investigated, especially in identifying to IPF "hot spots" which may be related to poor air quality.

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s12931-025-03164-2.

Supplementary material 1.

Acknowledgements
Ronald LaPorte, PhD (1949-2021), was a faculty member and an IPF patient who initiated this investigation. Ron's legacy lives on among his students and colleagues and in the inspiration to pursue this study.

Disclaimers
This was an independent, investigator initiated study supported by Boehringer Ingelheim Pharmaceuticals, Inc. (BIPI). BIPI had no role in the design, analysis or interpretation of the results in this study; BIPI was given the opportunity to review the manuscript for medical and scientific accuracy as it relates to BIPI substances, as well as intellectual property consideration.

ATS subject category
Epidemiology and Pulmonary fibrosis/fibroblast biology.

Online supplement
This article has an online supplement, which is accessible from this issue's table of contents online at www.atsjournals.org

Author contributions
Faina Linkov- Substantial contributions to the conception or design of the work, securing funding, the acquisition, analysis, or interpretation of data for the work, drafting the work, critically reviewing the paper for intellectual

content, final approval of the work, and agreeing to be accountable for all aspects of the work. Yue-Fang Chang- Substantial contributions to the conception or design of the work, the acquisition, analysis, or interpretation of data for the work, drafting the work, critically reviewing the paper for intellectual content, final approval of the work, and agreeing to be accountable for all aspects of the work. Harshitha Ramanan- The acquisition, analysis, or interpretation of data for the work, drafting the work, final approval of the work, and agreeing to be accountable for all aspects of the work. Richard S. Morgan—The acquisition, analysis, or interpretation of data for the work, critically reviewing the paper for intellectual content, final approval of the work, and agreeing to be accountable for all aspects of the work. Kathleen M McTigue- Substantial contributions to the conception or design of the work, the acquisition, analysis, or interpretation of data for the work, critically reviewing the paper for intellectual content, final approval of the work, and agreeing to be accountable for all aspects of the work. Anne EF Dimmock- The acquisition, analysis, or interpretation of data for the work, critically reviewing the paper for intellectual content, final approval of the work, and agreeing to be accountable for all aspects of the work. Rebecca Bascom- The acquisition, analysis, or interpretation of data for the work, critically reviewing the paper for intellectual content, final approval of the work, and agreeing to be accountable for all aspects of the work. Daniel J Kass- Substantial contributions to the conception or design of the work, securing funding the acquisition, analysis, or interpretation of data for the work, drafting the work, final approval of the work, and agreeing to be accountable for all aspects of the work.

Funding
Boehringer Ingelheim grant: Epidemiology of Idiopathic Pulmonary Fibrosis, Incidence, Prevalence and Mortality: Use of Capture Recapture Methods Advance Research (R2019-01). PI: Daniel Kass Dorothy P. and Richard P. Simmons Center for Interstitial Lung Disease, University of Pittsburgh.

Availability of data and materials
The data are stored at the University of Pittsburgh. It is not currently available in open access due to issues involving confidentiality of research participants.

Declarations

Ethics approval and consent to participate
University of Pittsburgh IRB # STUDY19080231 (exempt as non-human subject research). Penn State University IRB #STUDY00017636 (exempt as non-human subject research).

Consent for publication
Not applicable.

Competing interests
This was an independent, investigator initiated study supported by Boehringer Ingelheim Pharmaceuticals, Inc. (BIPI). BIPI had no role in the design, analysis or interpretation of the results in this study; BIPI was given the opportunity to review the manuscript for medical and scientific accuracy as it relates to BIPI substances, as well as intellectual property consideration.

Received: 6 August 2024 Accepted: 22 February 2025
Published online: 10 March 2025

References:

1. Raghu G, et al. An official ATS/ERS/JRS/ALAT statement: idiopathic pulmonary fibrosis: evidence-based guidelines for diagnosis and management. Am J Respir Crit Care Med. 2011;183(6):788–824.
2. Ley B, Collard HR. Epidemiology of idiopathic pulmonary fibrosis. Clin Epidemiol. 2013;5:483–92.
3. Hutchinson J, et al. Global incidence and mortality of idiopathic pulmonary fibrosis: a systematic review. Eur Respir J. 2015;46(3):795–806.
4. Kaul B, et al. Epidemiology of idiopathic pulmonary fibrosis among U.S. Veterans, 2010–2019. Ann Am Thorac Soc. 2022;19(2):196–203.
5. Samet JM, Coultas D, Raghu G. Idiopathic pulmonary fibrosis: tracking the true occurrence is challenging. Eur Respir J. 2015;46(3):604–6.
6. Lederer DJ, Martinez FJ. Idiopathic pulmonary fibrosis. N Engl J Med. 2018;378(19):1811–23.
7. *IPF - Idiopathic Pulmonary Fibrosis.* Breathe (Sheff), 2019; **15**(2): p. 153–160.
8. Jeganathan N, Sathananthan M. Mortality differences in pulmonary fibrosis among the most populated states in the United States. Respir Med. 2021;187:106565.
9. Raghu G, et al. An official ATS/ERS/JRS/ALAT clinical practice guideline: treatment of idiopathic pulmonary fibrosis. An update of the 2011 clinical practice guideline. Am J Respir Crit Care Med. 2015;192(2):e3-19.
10. Travis WD, et al. An official American Thoracic Society/European respiratory society statement: update of the international multidisciplinary classification of the idiopathic interstitial pneumonias. Am J Respir Crit Care Med. 2013;188(6):733–48.
11. American Thoracic Society. Idiopathic pulmonary fibrosis: diagnosis and treatment. International consensus statement. American Thoracic Society (ATS), and the European Respiratory Society (ERS). Am J Respir Crit Care Med. 2000;161(21):646–64.
12. Tsakiri KD, et al. Adult-onset pulmonary fibrosis caused by mutations in telomerase. Proc Natl Acad Sci U S A. 2007;104(18):7552–7.
13. Cronkhite JT, et al. Telomere shortening in familial and sporadic pulmonary fibrosis. Am J Respir Crit Care Med. 2008;178(7):729–37.
14. Armanios MY, et al. Telomerase mutations in families with idiopathic pulmonary fibrosis. N Engl J Med. 2007;356(13):1317–26.
15. Seibold MA, et al. A common MUC5B promoter polymorphism and pulmonary fibrosis. N Engl J Med. 2011;364(16):1503–12.
16. Zhang Y, et al. A variant in the promoter of MUC5B and idiopathic pulmonary fibrosis. N Engl J Med. 2011;364(16):1576–7.
17. Araki T, et al. Development and Progression of Interstitial Lung Abnormalities in the Framingham Heart Study. Am J Respir Crit Care Med. 2016;194(12):1514–22.
18. Podolanczuk AJ, et al. High attenuation areas on chest computed tomography in community-dwelling adults: the MESA study. Eur Respir J. 2016;48(5):1442–52.
19. Raghu G, et al. Incidence and prevalence of idiopathic pulmonary fibrosis. Am J Respir Crit Care Med. 2006;174(7):810–6.
20. King TE Jr, et al. A phase 3 trial of pirfenidone in patients with idiopathic pulmonary fibrosis. N Engl J Med. 2014;370(22):2083–92.
21. Richeldi L, et al. Efficacy and safety of nintedanib in idiopathic pulmonary fibrosis. N Engl J Med. 2014;370(22):2071–82.
22. NHS *Overview: Idiotpathic Pulmonary Fibrosis*. 2022.
23. Goobie GC, et al. Association of particulate matter exposure with lung function and mortality among patients with fibrotic interstitial lung disease. JAMA Intern Med. 2022;182(12):1248–59.
24. *Error in Figure 2.* JAMA Internal Med, 2022. **182**(12): 1331–1331.
25. Jacobs ET, Burgess JL, Abbott MB. The donora smog revisited: 70 years after the event that inspired the clean air act. Am J Public Health. 2018;108(S2):S85-s88.
26. Corrao G, et al. Capture-recapture methods to size alcohol related problems in a population. J Epidemiol Community Health. 2000;54(8):603–10.
27. Chang YF, et al. Dog bite incidence in the city of Pittsburgh: a capture-recapture approach. Am J Public Health. 1997;87(10):1703–5.
28. Spichler ER, et al. Capture-recapture method to estimate lower extremity amputation rates in Rio de Janeiro Brazil. Rev Panam Salud Pub. 2001;10(5):334–40.
29. Aaron DJ, et al. Estimating the lesbian population: a capture-recapture approach. J Epidemiol Community Health. 2003;57(3):207–9.
30. Bernillon P, et al. Record-linkage between two anonymous databases for a capture-recapture estimation of underreporting of AIDS cases: France 1990–1993. The Clinical Epidemiology Group from Centres d'Information et de Soins de l'Immunodeficience Humaine. Int J Epidemiol. 2000;29(1):168–74.
31. Lange JH, Chang YF, LaPorte RE. Use of the capture-recapture method for epidemiological studies in determining prevalence. Acta Neurol Scand. 2004;109(1):79–80.
32. Lange JH, LaPorte RE. Severe acute respiratory syndrome: capture-recapture method should be used to count how many cases of SARS really exist. BMJ. 2003;326(7403):1396.

33. Lange JH, et al. Capture-recapture method: the gold standard for incidence and prevalence. N Z Med J. 2003;116(1176):U488.
34. Lange JH, et al. Use of the capture-recapture method for determining the prevalence of neurological parasitic diseases. Neuroepidemiology. 2004;23(1–2):99.
35. Nishimura R, et al. Mortality trends in type 1 diabetes. The Allegheny County (Pennsylvania) Registry 1965–1999. Diabetes Care. 2001;24(5):823–7.
36. Dimmock AE. Evaluation of a Computable Phenotype for Idiopathic Pulmonary Fibrosis, in Public Health Sciences. 2016, Penn State University.
37. Dimmock AEF, et al. Shrinking the haystack: an approach to identifying idiopathic pulmonary fibrosis in the electronic health record using a computable phenotype. Res Sq. 2022;36:1416.
38. Network P. About the PaTH Network. 2016. https://www.pathnetwork.org/about/. Accessed 26 Feb 2025
39. Amin W, et al. PaTH: towards a learning health system in the Mid-Atlantic region. J Am Med Inform Assoc. 2014;21(4):633–6.
40. Cottin V, et al. Fibrosing interstitial lung diseases: knowns and unknowns. Eur Respir Rev. 2019;28(151):180100.
41. Lai CC, et al. Idiopathic pulmonary fibrosis in Taiwan—a population-based study. Respir Med. 2012;106(11):1566–74.
42. Fernandez Perez ER, et al. Incidence, prevalence, and clinical course of idiopathic pulmonary fibrosis: a population-based study. Chest. 2010;137(1):129–37.
43. Raghu G, et al. Idiopathic pulmonary fibrosis in US Medicare beneficiaries aged 65 years and older: incidence, prevalence, and survival, 2001–11. Lancet Respir Med. 2014;2(7):566–72.
44. Shull JG, et al. Mapping IPF helps identify geographic regions at higher risk for disease development and potential triggers. Respirology. 2021;26(4):352–9.
45. Alder JK, et al. Lung transplantation for idiopathic pulmonary fibrosis enriches for individuals with telomere-mediated disease. J Heart Lung Transpl. 2022;41(5):654–63.
46. Gribbin J, et al. Incidence and mortality of idiopathic pulmonary fibrosis and sarcoidosis in the UK. Thorax. 2006;61(11):980–5.
47. Kornum JB, et al. The incidence of interstitial lung disease 1995–2005: a Danish nationwide population-based study. BMC Pulm Med. 2008;8:24.
48. Navaratnam V, et al. The rising incidence of idiopathic pulmonary fibrosis in the UK. Thorax. 2011;66(6):462–7.
49. Coultas DB, et al. The epidemiology of interstitial lung diseases. Am J Respir Crit Care Med. 1994;150(4):967–72.
50. Thomeer M, et al. Multidisciplinary interobserver agreement in the diagnosis of idiopathic pulmonary fibrosis. Eur Respir J. 2008;31(3):585–91.
51. Raghu G, et al. Idiopathic pulmonary fibrosis (an update) and progressive pulmonary fibrosis in adults: an official ATS/ERS/JRS/ALAT clinical practice guideline. Am J Respir Crit Care Med. 2022;205(9):e18–47.
52. Foundation PF. Rural Outreach in Pulmonary Fibrosis: Provider Statement. 2024. https://www.pulmonaryfibrosis.org/researchers-healthcare-providers/clinical-resources/position-statements/rural-outreach-in-pulmonary-fibrosis--provider-statement. Accessed 26 Feb 2025
53. Greenfield G, et al. Patient-initiated second medical consultations-patient characteristics and motivating factors, impact on care and satisfaction: a systematic review. BMJ Open. 2021;11(9): e044033.
54. DeDent AM, Collard HR, Thakur N. Disparities in rural populations with idiopathic pulmonary fibrosis. Chest. 2022;162(3):630–4.
55. LaPorte RE, et al. Efficiency and accuracy of disease monitoring systems: application of capture-recapture methods to injury monitoring. Am J Epidemiol. 1995;142(10):1069–77.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.