A historical database cohort study addressing the clinical patterns prior to idiopathic pulmonary fibrosis (IPF) diagnosis in UK primary care

David R Thickett[1], Jaco Voorham[2], Ronan Ryan[2], Rupert Jones[3], Robina Coker[4], Andrew Wilson[5], Sen Yang[2], Mandy Ow[2], Priyanka Raju[2], Isha Chaudhry[2], Antony Hardjojo[2], Victoria Carter[2], David Price[2,6,7]

[1]Birmingham Acute Research Group, University of Birmingham, Birmingham, United Kingdom, [2]Observational and Pragmatic Research Institute, Singapore, Singapore, [3]University of Plymouth, Plymouth, United Kingdom, [4]Imperial College London, London, United Kingdom, [5]Norwich Medical School, University of East Anglia, Norwich, United Kingdom. [6]Optimum Patient Care, United Kingdom; [7]Academic Primary Care, University of Aberdeen, Scotland, United Kingdom.

Corresponding author:

Prof David B Price, Academic Primary Care, Division of Applied Health Sciences, University of Aberdeen, Polwarth Building, Foresterhill, Aberdeen, UK AB25 2ZD;

Tel: +65 6962 3627; Email: dprice@opri.sg

Word count: 3871/4000 words

ABSTRACT (293/300 words)

Objective:

To explore the clinical pathways, including signs and symptoms, and symptom progression patterns preceding Idiopathic Pulmonary Fibrosis (IPF) diagnosis.

Design and setting:

A historical cohort study was conducted using primary care patient records from the Optimum Patient Care Research Database.

Participants:

Patients included were at least 30 years, had IPF diagnosis, identified via clinical-coding and free-text records, and had a consultation with a chest specialist prior to IPF diagnosis.

Outcome measures:

The signs & symptoms in the year prior to IPF diagnosis from clinical codes and free-text in primary care electronic records included: cough, dyspnoea, dry cough, weight loss, fatigue/malaise, loss of appetite, crackles and clubbed fingers. The time course of presentations of clinical features and investigations in the years prior to IPF diagnosis were mapped.

Results:

Within 462 patients identified, the majority (77.9%) had a respiratory consultation within 365 days prior to the chest specialist visit preceding the IPF diagnosis recorded in their primary care records. The most common symptoms recorded in the one-year prior to IPF diagnosis were dyspnoea (48.7%) and cough (40.9%); other signs and symptoms were rarely recorded (<5%). The majority of patients with cough (58.0%) and dyspnoea (55.0%) in the one-year before IPF diagnosis had multiple recordings of the respective symptoms. Both cough and dyspnoea were recorded in 23.4% of

patients in the year prior to diagnosis. Consultation rates for cough, dyspnoea and both, but not other signs or symptoms, began to increase 4-5 years prior diagnosis, with the sharpest increase in the last year. Cough and dyspnoea were often preceded by a reduction in measured weight over 5-years leading to IPF diagnosis.

Conclusion:

Prolonged cough and/or progressive dyspnoea, especially if accompanied with weight loss, should signal for a referral to specialist assessment at the earliest opportunity.

Keywords: Interstitial Lung Disease, Free-text Record, Pathway Features, Principal Component Analysis.

Strengths and Limitations of this Study

- This was the first study to map clinical progression patterns in the years leading to IPF diagnosis.
- We used a large primary care database with data available up to seven years prior to IPF diagnosis and records of patients' secondary care usage.
- We developed a list of relevant clinical features via initial review of patient records, allowing for identification of less common clinical features which may be used to find patterns of symptoms to potentially aid IPF diagnosis.
- A very specific definition of IPF was used in this study requiring specialist consultation and diagnosis of IPF in primary care after the consultation. This may result in the exclusion of cases where there was no primary care record indicating specialist consultation.
- Coding limitations of a database that was designed for clinical practice rather than research, are a weakness. However, we utilised both symptom coding and

free-text data to identify IPF, and clinical features indicative of IPF, increasing detection sensitivity.

Funding: This study was funded by Roche. Roche was not involved in the preparation, drafting, or editing of this manuscript but did conduct a factual accuracy check of the final manuscript; any decision to incorporate comments was made solely at the discretion of the authors.

INTRODUCTION

Idiopathic Pulmonary Fibrosis (IPF) is the most common and most lethal of the idiopathic interstitial pneumonias, estimated to affect 14-43 per 100,000 population, typically over the age of 50.[1-3] Patients with IPF have a very poor prognosis, with median years of survival of only 2.5-3.5 years and 5-year survival rate as low as 20%.[4, 5] In the UK, the mortality rate for IPF was reported to be 5.10 per 100,000 person-years in 2005-2008, increasing by around 5% per year since 1968,[6] thus signifying a significant and growing healthcare concern.

Diagnosing IPF in primary care is challenging due to the non-specific features in the early stages of the disease [7], and IPF is seldom seen by physicians in primary care.[8] In secondary care, a confident diagnosis requires expertise from a multidisciplinary team of pulmonologists, radiologists, and pathologists.[9, 10] Identification of early predictors and clinical patterns leading to subsequent IPF diagnosis may help primary care physicians identify potential IPF patients for further examination and guide appropriate referral to specialist respiratory services.

Treatment of IPF had previously been limited to symptom management and palliation, but two antifibrotic drugs, Pirfenidone and Nintedanib [11, 12], have recently become available. Both treatments may improve IPF disease outcomes [13-16] and are recommended by the UK National Institute for Health and Care Excellence [17, 18] and by the ATS/ERS/JRS/ALAT clinical practice guideline.[19] To obtain optimal benefit from these novel treatment agents, an early diagnosis is required.

While efforts have been made to study the symptom pathways leading to IPF diagnosis, identifying IPF patients using diagnosis codes alone may erroneously include patients with other conditions.[20] Thus, this study was conducted to

characterise the pattern of signs, symptoms and other clinical predictors preceding IPF diagnosis using a real-life population of patients and more definitive selection criteria for IPF.

METHODS

Study design

This was a real-life historical cohort study using electronic medical records from the OPCRD, a clinical research database containing records of approximately 7 million patients from over 700 primary care centres across the UK (http://opcrd.co.uk/) [21] with linked patient-completed asthma questionnaire. Asthma outcome measures within the OPCRD have been validated against patient-reported outcomes and treatment response. [22] The study was conducted according to the quality standards suggested for observational studies,[23] including the use of an *a priori* analysis plan, study registration with a commitment to publish, and a well-maintained and monitored study database.

Case definition and inclusion criteria

The inclusion criteria for this study were: between 30-100 years of age at diagnosis date, diagnosis date between 2010 and 1st May 2017, and the diagnosis date must have occurred at least 1 year after registration at a participating general practice (GP). Patients were excluded if they had a code for sarcoidosis, allergic alveolitis, pneumoconiosis, asbestosis, or other causes of pulmonary fibrosis. The diagnosis date was defined as the date of the first IPF diagnosis or prescription for IPF-specific pharmacotherapy (Pirfenidone or Nintedanib). The primary analysis population includes patients who had a code for a consultation with a chest specialist prior to the

IPF diagnosis. Selection of patients for analyses is further detailed in the online data supplement.

Outcome assessments

The primary outcome of this study was the presence of signs and symptoms in the one-year prior to IPF diagnosis. The secondary outcomes of this study were: 1) the consultation rates for the signs and symptoms up to 7 years prior to diagnosis, 2) the proportion of patients with respiratory consultations and respiratory tests conducted within 90 and 365 days prior to the chest specialist consultation preceding IPF diagnosis, and 3) the relationship between the signs and symptoms.

Respiratory consultation was identified based on the presence of Read codes for the following clinical features: chest/respiratory infection, chest symptom, clubbed fingers, cough, crackles, dyspnoea or sputum. Respiratory tests were identified via Read codes for chest X-ray, chest CT scan, lung function test and chest examination.

The Read codes and free-text terms for signs, symptoms and clinical features investigated in this study were selected via an initial round of manual review of patient records. The methods for the selection and the list of Read codes and free-text terms are elaborated in the online data supplement.

Statistical analysis

For all data handling, statistical analyses and figures Stata SE version 14.2 and Stata MP/6 version 15.1 were used. Descriptive statistics were used for baseline characteristics. Continuous variables were summarised as mean (standard deviation) or median (interquartile range) of non-missing observations while categorical variables were presented as proportions of non-missing observations.

Prevalence of signs & symptoms identified via both Read codes and free-text in the one-year prior to IPF diagnosis were presented as number (%). Descriptive statistics were also produced for the number of patients with respiratory consultations and tests conducted within 90 days and within 365 days prior to the chest specialist consultation preceding the diagnosis date.

Principal Component Analysis (PCA) with rotation was utilised to visualise the prevalence and co-occurrence of codes for signs, symptoms and other clinical features 90 and 365 days prior to IPF diagnosis in a network plot. All signs and symptoms co-occurrence patterns in the year prior diagnosis were also presented as a table.

Consultation rates for signs and symptoms from up to seven years prior to IPF diagnosis were generated and expressed as rates per 100 patient-years with 95% confidence intervals. For each year, only patients who had a complete year of available data contributed to the frequency statistics. The number of patients with at least a certain recording frequency (recording dates/year, in the period of up to 12 years before IPF diagnosis) was also tabulated. Data are shown for, on average, once a year, once every 2, 3 and 4 years.

Time-course of signs, symptoms and weight measurement of individual patients was plotted to visualize the disease progression patterns up to 12 years leading to IPF diagnosis. Visual evaluation was used to quantify the number of patients with each symptom progression pattern.

Descriptive analysis for the mean (SD) time in years between the first recorded cough and dyspnoea until the diagnosis of IPF was conducted for patients with at least 1-year medical record history before first symptom available and had their symptoms occurring before the IPF diagnosis. The cumulative probability of IPF diagnosis every

year since the first symptoms is presented in the form of life tables and Kaplan-Meier plots.

In a sensitivity analysis, all analyses were repeated in an overall IPF patients group consisting of patients with Read codes or free-text mention for IPF diagnosis and/or medication (detailed in the online supplementary data). For this group, the proportions of patients with respiratory consultations and tests were analysed prior to the date of IPF diagnosis or chest consultation (whichever came first).

Ethical approval

The OPCRD is approved by the Health Research Authority of the UK NHS for clinical research use (REC reference: 15/EM/0150), and the protocol for this study was approved by the Anonymised Data Ethics Protocols and Transparency (ADEPT) committee, the independent scientific advisory committee for the OPCRD. The study protocol was registered with the European Network of Centres for Pharmacoepidemiology and Pharmacovigilance (ENCePP, registration number EUPAS20994).[24]

Data availability

The authors do not have permission to give public access to the study dataset; researchers may request access to OPCRD data for their own purposes. Access to OPCRD can be made via the OPCRD website (https://opcrd.co.uk/our-database/data-requests/) or via the enquiries email info@opcrd.co.uk.

Patient and Public Involvement

This study does not involve patients. The public was not invited in the design of the study nor the writing of this document.

RESULTS

Patients

The patient selection flow is depicted in Figure 1. A total of 1,166 patients were identified to have IPF diagnosis or IPF-related medication following inclusion and exclusion criteria (overall IPF patient group). Of these, 462 patients had a consultation with a chest specialist prior to their IPF diagnosis (primary analysis group).

Of the 462 patients, 58.9% were male with a mean age at the time of diagnosis was 75 years (Table 1). The majority (63.7%) of patients had a history of smoking, with 9.8% being current smokers. A total of 88 (19.0%) patients had spirometry recorded within 90 days and 160 (36.4%) within 365 days prior to IPF diagnosis. Baseline demographics for the overall IPF patient group (n=1,166) were similar (Supplementary Table E1).

Table 1. Baseline demographic patient characteristics and procedures (n=462).

Variable	Frequency*
Age (years) • mean (SD) • Median (IQR)	 74.6 (9.6) 75.0 (69.0; 81.0)
Male gender	272 (58.9)
BMI • n (% non-missing) • <18.5 • 18.5 - <25 • 25 - <30 • ≥30	 442 (95.7) 10 (2.3) 137 (31.0) 163 (39.8) 132 (29.9)
Smoking status • n (% non-missing) • Current smoker • Ex-smoker • Never smoked	 438 (94.8) 43 (9.8) 236 (53.9) 159 (36.3)

Asthma diagnosis†	24 (5.2)
COPD diagnosis†	19 (4.1)
Respiratory tract cancer	0 (0.0)
Other chronic respiratory diseases (excl. cancer)¶	4 (0.9)
Lung function test conducted • 90 days prior diagnosis • 365 days prior diagnosis	 88 (19.0) 160 (34.6)
*Numbers are presented as n (%) unless specified. †First diagnostic code recorded ever prior and up to diagnosis date. SD: Standard deviation, IQR: Interquartile range, COPD: chronic obstructive pulmonary disease. ¶Listed in the online data supplement.	

186

187 **Primary outcome – Signs and symptoms in the one-year prior to IPF diagnosis**

188 Of the investigated signs and symptoms, cough and dyspnoea were the most common
189 symptoms recorded in the one-year prior to IPF diagnosis: cough in 189 (40.9%) and
190 dyspnoea in 225 (48.7%) patients (Table 2). The majority of the patients with cough
191 (58.0%) and dyspnoea (55.0%) had more than one recording of the respective
192 symptoms in the one-year period prior diagnosis. Other signs and symptoms were
193 rarely observed as the first recorded signs and symptoms (<5%). Cough and dyspnoea
194 co-occurred on 23.4% of the patients. The expected rate for both symptoms to co-
195 occur, based on the individual rates, was 20.0% (95% confidence interval 18.1% -
196 21.9%). This was not significantly different from the observed co-occurrence rate
197 (23.4% [21.4% - 25.4%]), suggesting both symptoms occurred independently of each
198 other.

199 Detailed results including signs and symptoms identified via Read codes only, as well
200 as signs & symptoms in the overall IPF patients, are presented in Supplementary
201 Table E2. Compared to using Read codes only, the addition of free-text data increased
202 the number of patients with cough by 8.6% and dyspnoea by 13.6%.

Table 2. Signs & symptoms in the one-year prior to IPF diagnosis (n=462)	
Signs and symptom	**n (%)**

Dyspnoea	225 (48.7)
Cough*	189 (40.9)
Fatigue or malaise	22 (4.8)
Weight loss	19 (4.1)
Crackles	14 (3.0)
Loss of appetite	13 (2.8)
Clubbed fingers	2 (0.4)
Symptom combinations	**n (%)**
Cough & Dyspnoea	108 (23.4)
Dyspnoea without cough	117 (25.3)
Cough without dyspnoea	81 (17.5)
*Of which 25 were “dry cough”.	

Secondary outcome – History of respiratory consultation and respiratory test prior to chest specialist consultation

Around three-quarters of (77.9%) patients had a respiratory primary care consultation recorded in the 365 days before their chest specialist consultation (Table 3). Respiratory tests were recorded in 61.3% patients in the 365 days prior to the specialist consultation, and more than a third of that (38.1%) was within a 90 days period. Prevalence of the components of respiratory consultation and respiratory tests are presented in Supplementary Table E3

Table 3. History of respiratory consultation and tests before chest specialist consultation prior to IPF diagnosis.*	
Respiratory consultation † • within 90 days • within 365 days	 238 (51.5) 360 (77.9)
Respiratory tests conducted‡ • within 90 days • within 365 days	 176 (38.1) 283 (61.3)
*Frequency expressed as n (%). †Codes for chest/respiratory infection, chest symptoms, clubbed fingers, cough, crackles, dyspnoea, or sputum or wheeze. ‡Codes for chest X-ray, chest CT scan, lung function test and chest examination.	

The proportions of overall IPF patients with respiratory consultations and tests are presented in Supplementary Table E4.

Secondary outcome – Relationship between signs and symptoms

The network plots in Figure 2 depicts the relationship between signs and symptoms recorded in the 90 days and in the 365 days prior to IPF diagnosis. Strong associations were observed between cough and dyspnoea, and among weight loss, loss of appetite, and fatigue/malaise. These two clusters of symptoms were more pronounced in the 90 days period compared to in the 365 days period. Within the extended list of codes which includes clinical features, chest infection was identified to be associated with the cough and dyspnoea cluster. However, chest symptoms, a category containing unspecified chest symptoms, chest examinations, and chest CT and X-rays, were more closely associated with the fatigue/malaise cluster. Similar results were observed within the overall IPF patient population (Supplementary Figure E1).

The prevalence of every symptom combination within one-year prior to IPF diagnosis is shown in Table 4. No symptoms were recorded in 30.7% of the patients. The most common symptom pattern was dyspnoea alone (22.9%), cough with dyspnoea (16.5%) and cough alone (13.9%). Other symptom combinations were rarely observed (≤2% patients). Symptom combinations in the overall IPF patients are shown in Supplementary Table E5.

Table 4. Co-occurrence of signs and symptoms in the one-year period up to IPF diagnosis

Clubbed fingers	Cough	Crackles	Dry cough	Dyspnoea	Fatigue or malaise	Loss of appetite	Weight loss	Patients	%	Cumulative %
								142	30.7	30.7
				■				106	22.9	53.7
	■			■				76	16.5	70.1
	■							64	13.9	84.0
	■		■					9	1.9	85.9
	■		■	■				9	1.9	87.9
							■	8	1.7	89.6
	■			■	■			5	1.1	90.7
		■						4	0.9	91.6
	■	■		■				4	0.9	92.4
				■	■			3	0.6	93.1
	■				■			3	0.6	93.7
				■		■		2	0.4	94.2
	■		■	■	■			2	0.4	94.6
	■						■	2	0.4	95.0
	■	■						2	0.4	95.5
	■			■		■		2	0.4	95.9
	■			■	■	■	■	1	0.2	96.1
■				■				1	0.2	96.3
						■	■	1	0.2	96.5
■	■			■				1	0.2	96.8
	■		■			■		1	0.2	97.0
				■		■	■	1	0.2	97.2
				■			■	1	0.2	97.4
	■	■		■	■			1	0.2	97.6
	■		■	■	■	■		1	0.2	97.8
				■	■	■	■	1	0.2	98.1
	■		■	■	■	■	■	1	0.2	98.3
					■			1	0.2	98.5
		■		■				1	0.2	98.7
		■		■	■			1	0.2	98.9
	■		■	■		■		1	0.2	99.1
	■			■	■	■		1	0.2	99.4
	■	■		■			■	1	0.2	99.6
	■		■	■	■		■	1	0.2	99.8
	■			■			■	1	0.2	100.0

233

234 **Secondary outcome – Consultation rate in the years leading to IPF diagnosis**

The frequencies of codes for signs and symptoms up to 7 years leading to the IPF diagnosis are shown in Figure 3. A rapid rise in the consultation rate for cough and dyspnoea occurred in the year prior to IPF diagnosis, but the increase in cough and dyspnoea started 4-5 years before diagnosis. A similar pattern was observed for consultation for concurrent cough and dyspnoea. This rise was not observed in the other, less common signs and symptoms. The frequencies of codes for signs and symptoms in the overall IPF group are shown in Supplementary Figure E2.

The number of patients with at least a certain average symptom frequency in the period before IPF diagnosis (up to 12 years) are presented in Supplementary Table E6.

Symptom progression patterns preceding IPF diagnosis

Several typical patterns of patient pathways were identified from visual assessment of individual patient timelines; cough tended to precede dyspnoea. Weight loss, observed via weight measurements over time, commonly followed recordings of cough and dyspnoea. 244 of the 462 (52.8%) patients had ≥4 records of weight in the previous five years, and of these, 116 (47.5%) recorded weight loss of more than five kilograms. The weight loss often occurred over at least 5 years period (example from 2 patients in Figure 4), however, acute weight loss was also observed (Supplementary Figure E3).

Other examples of identified patterns are presented in Supplementary Figure E4. A minority, 69 out of 462 (14.9%), received a diagnosis without any symptom codes in the preceding years. Many patients had no clear pattern of respiratory symptoms or weight loss.

Probability of IPF diagnosis from the first recording of symptoms

Analysis of probability for IPF diagnosis since the first recorded symptom of cough or dyspnoea included 463 patients (322 and 293 patients with cough and dyspnoea respectively). The mean (SD) time since the first cough was longer (6.3 [5.5] years) compared to since the first dyspnoea (4.3 [4.3] years). Cumulative probability of IPF diagnosis since the first recording of symptoms is illustrated as a life table (Supplementary Table E7) and a Kaplan-Meier plot (Supplementary Fig E5). These show that 50% of the patients are diagnosed within 5 years from their first recorded cough, and within 3 years from their first recorded dyspnoea. However, it took respectively 13 and 10 years since their first symptoms for 90% of the patients to receive an IPF diagnosis.

DISCUSSION

Main findings

This is a real-life historical cohort study of IPF patients to identify the symptoms, signs, and clinical features recorded before the diagnosis of IPF. Cough and dyspnoea were observed to be the most common recorded symptoms in the one-year prior to IPF diagnosis, occurring in above 40% of the patients. Other symptoms, such as weight loss and fatigue, or signs such as crackles and clubbed fingers were rarely observed (<5% patients). The majority of patients (77.9%) had a history of at least one primary care respiratory consultation within a year prior to their chest specialist consultation preceding the diagnosis of IPF.

A retrospective review of records from up to 7 years prior to diagnosis demonstrated an increased frequency of codes for cough and dyspnoea starting from 4-5 years prior to IPF diagnosis, which further rose sharply nearer to the IPF diagnosis. Thus,

repeated consultations for prolonged cough and dyspnoea are likely to be the characteristic symptoms of IPF.

The symptom combination of cough and dyspnoea in the year before IPF diagnosis was observed to be one of the most common symptom co-occurrence patterns. Using Principal Component Analysis, we have shown the close association between cough and dyspnoea in the preceding 90 and 365 days within patients with IPF. Rates for consultations for both symptoms also increased 4-5 years prior diagnosis, similar to the coding of individual symptoms. Thus, prolonged history or presentation of both symptoms is likely to be a “red flag” to consider IPF and to refer to specialist respiratory services especially when other conditions such as asthma or COPD have been ruled out. A second cluster between weight loss, loss of appetite and fatigue or malaise was identified. However, these signs and symptoms were too infrequent and unlikely to be helpful for the early identification of IPF in community settings.

Surprisingly, 30.7% of our patients were not recorded as having any signs or symptoms in the one-year prior to IPF diagnosis (Table 4). Analysis of symptom progression patterns also showed that 14.9% of the patients did not have any symptom codes in the years leading to IPF diagnosis. This may be due to a tendency for primary care physicians to code for the final diagnosis instead of the presenting symptoms during consultations. Alternatively, IPF may be diagnosed coincidentally by investigations for other conditions, such as cardiac CT scanning, without records of respiratory signs and symptoms.

A novelty of the current study is the investigation of symptom progression patterns in the years leading to IPF diagnosis. Cough symptoms tended to precede dyspnoea. However, many of these cough records may have been unrelated to IPF, for example,

viral infection. Weight loss was observed to be a common sequel to cough and dyspnoea, occurring over at least 5 years, suggesting the possibility of a disease process preceding the diagnosis by several years. Many patient pathways had no clear pattern of respiratory symptoms or weight loss.

Strengths & Limitations

A strength of this study is the conduct of initial reviews of individual primary care case notes to create the list of Read codes and free-text terms. This enabled less commonly IPF-related signs, symptoms and clinical features to be included within this study, which may have been otherwise missed if a pre-determined list of features and codes was used instead.

This study used information from free-text primary care records in addition to diagnostic Read codes. The usage of information from free-text records increased detection sensitivity for signs & symptoms, contributing an additional 8.6% patients with symptoms of cough and 13.6% with dyspnoea in the one-year prior to IPF diagnosis. The recording of respiratory symptoms such as breathlessness has previously been found to occur in free-text well before specific codes are entered in for people with COPD and asthma.[25]

This study utilised every eligible patient from a primary care records database, with the data availability extending to 7 years prior to the diagnosis of IPF. The real-life design of this study provides high generalisability of the results to primary care patients managed in actual primary care practice. Additionally, the long observation period and the availability of free-text clinical records demonstrate the particular strengths of the OPCRD database for longitudinal studies on rare diseases. Clinical records within

participating GPs also contain records of patients' secondary care usage, allowing identification of patients who had previous consultation with a specialist.

In this study, we selected patients who had a consultation with a chest specialist prior to their diagnosis for analysis. This group of patients was selected out of concern that coding for IPF diagnosis may have been entered as a diagnostic query or mistake instead of a definitive diagnosis. Indeed, higher rates of cough and dyspnoea were observed from both Read codes and free-text in patients with specialist consultation prior to IPF diagnosis (Supplementary Table E2). Potential bias of GP diagnosis was also addressed by the additional requirement for prior specialist assessment.

The main weakness of this study is our lack of gold standard diagnostic criteria in many patients as we were unable to review the contents of letters from secondary care or hospital records directly. Due to this, it is not possible to directly confirm that the diagnoses of IPF were made in secondary care, despite the presence of preceding consultation with chest specialists. Our additional selection criterion, requiring prior chest specialist consultation, also resulted in a reduced sample size. Regardless, the pattern of previous respiratory examinations and the outcomes observed seem to support the diagnosis of IPF within the patients. A further weakness is that the primary care records in the OPCRD were initially collected for clinical instead of research purposes, thus the coding of symptoms may be inconsistent. However, the use of free text searches in this study may mitigate this issue.

We did not exclude patients with asthma or COPD, which may also cause symptoms of cough and dyspnoea. We felt that asthma and COPD are both common disorders and misdiagnosis of IPF as either symptom are likely. Regardless, the numbers of patients with asthma (n=24) and COPD (n=19) are too small to conduct separate

analyses and are unlikely to relevantly change our conclusions. The small number of patients with concomitant COPD is likely due to the unique requirement of the UK primary care system since 2002 in which a diagnosis of COPD requires confirmation by spirometry. Consequentially, the UK may have less misdiagnosis of IPF as COPD than other health care systems.

Placing results in the context of published studies

The observed increase in frequency for cough and/or dyspnoea from 5 years prior IPF diagnosis suggests a delay between recognition of symptoms and diagnosis. Such a delay in the referral and diagnosis of IPF since the initial symptom presentation has been observed in previous studies.[26, 27] A multicentre cohort study was recently conducted to describe the factors responsible for the delay in IPF diagnosis.[28] The study reported a mean delay of 2.1 years and that the delay can be mainly attributed to the patients, general practitioners and community hospitals. Due to the importance of early diagnosis of IPF,[29] there is a need to further understand and rectify the causes for delays in diagnosis and referral.

A previous study investigated the prevalence of symptoms from up to 5 years before IPF diagnosis using another primary care record database in the UK, The Health Improvement Network (THIN).[20] Similar to our study, the study also reported breathlessness and cough, identified via Read codes, to be the most common symptoms increasing sharply at one year prior to IPF diagnosis. However, the study relied solely on Read codes (for idiopathic fibrosing alveolitis [H563.00], Hamman-Rich syndrome [H563.11], cryptogenic fibrosing alveolitis, diffuse pulmonary fibrosis, and idiopathic fibrosing alveolitis NOS) to identify patients with IPF. The authors acknowledged that there was a possibility of miscoding leading to the inclusion of

patients with other fibrotic lung disorders [6] and thus labelled cases within their study as “IPF-clinical syndrome” instead of definitive IPF patients. Our current study confirmed their finding using a more conservative IPF definition (requiring chest specialist consultation prior to the diagnosis) which was expected to be more selective for patients with actual IPF. The current study also extends the previous study by investigating the relationship between the signs and symptoms, demonstrating a common co-occurrence between cough and breathlessness. Furthermore, our study has the additional strength of a longer observational period prior to IPF diagnosis (7 years) and the utilisation of free-text clinical records to identify signs and symptoms which would have been missed by using only Read codes.

Conclusions

This is the first study to analyse the presence, association and progression patterns of symptoms and clinical features leading to IPF diagnosis within a more definitive group of patients with IPF, utilising both coded diagnosis and free-text primary care records. Cough and dyspnoea represented the most common symptoms in the one-year prior to the diagnosis of IPF. Both symptoms were found to be closely associated and were likely to co-occur. While the consultation rates for cough and dyspnoea rose sharply one-year preceding the diagnosis, the increases in rates were already observable from up to 5 years prior to IPF diagnosis. Lastly, the majority of the patients had records of respiratory consultations and tests conducted prior to their IPF diagnosis.

Taken together, general practitioners should further assess the possibility of IPF in patients who have increasing consultations for prolonged cough and/or progressive dyspnoea, especially if accompanied with weight loss.

Further work will investigate the value of chest x-ray and spirometry in predicting subsequent IPF diagnosis. Linking data within electronic medical records and registries such as BLUETEQ database register of UK specialist drug prescribing [30] and BTS ILD registries for higher quality research in IPF is also needed to provide deeper insight into the patterns of disease progression identified in this study. Further research comparing the clinical pathway of IPF patients with a control group of patients, as well as investigating whether spirometry led to a timelier referral to specialists may also be warranted.

Acknowledgements

The authors would like to extend their acknowledgement to Dr Dermot Ryan for his clinical and scientific input.

Declaration of interests

David Price has board membership with Amgen, AstraZeneca, Boehringer Ingelheim, Chiesi, Circassia, Mylan, Mundipharma, Napp, Novartis, Regeneron Pharmaceuticals, Sanofi Genzyme, Teva Pharmaceuticals; consultancy agreements with Amgen, AstraZeneca, Boehringer Ingelheim, Chiesi, GlaxoSmithKline, Mylan, Mundipharma, Napp, Novartis, Pfizer, Teva Pharmaceuticals, Theravance; grants and unrestricted funding for investigator-initiated studies (conducted through Observational and Pragmatic Research Institute Pte Ltd) from AKL Research and Development Ltd, AstraZeneca, Boehringer Ingelheim, British Lung Foundation, Chiesi, Circassia, Mylan, Mundipharma, Napp, Novartis, Pfizer, Regeneron Pharmaceuticals, Respiratory Effectiveness Group, Sanofi Genzyme, Teva Pharmaceuticals, Theravance, UK National Health Service, Zentiva (Sanofi Generics); payment for lectures/speaking engagements from AstraZeneca,

Boehringer Ingelheim, Chiesi, Cipla, GlaxoSmithKline, Kyorin, Mylan, Merck, Mundipharma, Novartis, Pfizer, Regeneron Pharmaceuticals, Sanofi Genzyme, Teva Pharmaceuticals; payment for manuscript preparation from Mundipharma, Teva Pharmaceuticals; payment for the development of educational materials from Mundipharma, Novartis; payment for travel/accommodation/meeting expenses from AstraZeneca, Boehringer Ingelheim, Circassia, Mundipharma, Napp, Novartis, Teva Pharmaceuticals; funding for patient enrolment or completion of research from Chiesi, Novartis, Teva Pharmaceuticals, Zentiva (Sanofi Generics); stock/stock options from AKL Research and Development Ltd which produces phytopharmaceuticals; owns 74% of the social enterprise Optimum Patient Care Ltd (Australia and UK) and 74% of Observational and Pragmatic Research Institute Pte Ltd (Singapore); and is peer reviewer for grant committees of the Efficacy and Mechanism Evaluation programme, and Health Technology Assessment.

Rupert Jones

Rupert Jones declares grants from AstraZeneca, GlaxoSmithKline, and Novartis; and personal fees for consultancy, speakers fees, or travel support from AstraZeneca, Boehringer Ingelheim, Chiesi, GlaxoSmithKline, Nutricia, OPRI, and Pfizer.

.

David R Thickett, Robina Coker, and Andrew Wilson have no interest to declare.

Jaco Voorham, Ronan Ryan, Sen Yang, Mandy Ow, Priyanka Raju, Isha Chaudhry, Victoria Carter and **Antony Hardjojo** are employees of Observational and Pragmatic Research Institute Pte Ltd, which has conducted paid research in respiratory disease on behalf of the following organizations in the past 5 years: Anaxys, AstraZeneca, Boehringer Ingelheim, British Lung Foundation, Chiesi, Circassia (formerly Aerocrine), GlaxoSmithKline, Harvey Walsh, Mapi, Morningside

Healthcare, Mundipharma, Mylan (formerly Meda), Napp, Novartis, Orion, Plymouth University, Regeneron, Respiratory Effectiveness Group, Roche, Sanofi, Takeda, Teva, University of East Anglia, Zentiva (a Sanofi company).

Author Contribution:

All authors contributed to the planning, conduct, and reporting of this study. DT, RC, AW, SY, VC, and DP conducted the design and planning of this study. JV, RR, RJ, MO, PR, IC and AH contributed to the data analysis and interpretation. The overall conduct of this study was supervised by DP. All authors were involved in the drafting of this manuscript and approved the submission of this manuscript.

REFERENCES

1. Spagnolo P, Sverzellati N, Rossi G, et al. Idiopathic pulmonary fibrosis: an update. Ann Med. 2015;47(1):15-27.
2. Raghu G, Chen SY, Hou Q, et al. Incidence and prevalence of idiopathic pulmonary fibrosis in US adults 18-64 years old. The European respiratory journal. 2016;48(1):179-86.
3. Agabiti N, Porretta MA, Bauleo L, et al. Idiopathic Pulmonary Fibrosis (IPF) incidence and prevalence in Italy. Sarcoidosis, vasculitis, and diffuse lung diseases : official journal of WASOG. 2014;31(3):191-7.
4. Ley B, Collard HR, King TE, Jr. Clinical course and prediction of survival in idiopathic pulmonary fibrosis. American journal of respiratory and critical care medicine. 2011;183(4):431-40.
5. Strongman H, Kausar I, Maher TM. Incidence, Prevalence, and Survival of Patients with Idiopathic Pulmonary Fibrosis in the UK. Advances in therapy. 2018;35(5):724-36.
6. Navaratnam V, Fleming KM, West J, et al. The rising incidence of idiopathic pulmonary fibrosis in the U.K. Thorax. 2011;66(6):462-7.
7. Zibrak JD, Price D. Interstitial lung disease: raising the index of suspicion in primary care. NPJ Prim Care Respir Med. 2014;24:14054.
8. Thickett DR, Kendall C, Spencer LG, et al. Improving care for patients with idiopathic pulmonary fibrosis (IPF) in the UK: a round table discussion. Thorax. 2014;69(12):1136-40.
9. Wells AU. The revised ATS/ERS/JRS/ALAT diagnostic criteria for idiopathic pulmonary fibrosis (IPF)--practical implications. Respir Res. 2013;14 Suppl 1:S2.
10. Raghu G, Collard HR, Egan JJ, et al. An official ATS/ERS/JRS/ALAT statement: idiopathic pulmonary fibrosis: evidence-based guidelines for diagnosis and management. American journal of respiratory and critical care medicine. 2011;183(6):788-824.
11. George PM, Wells AU. Pirfenidone for the treatment of idiopathic pulmonary fibrosis. Expert review of clinical pharmacology. 2017;10(5):483-91.
12. Mazzei ME, Richeldi L, Collard HR. Nintedanib in the treatment of idiopathic pulmonary fibrosis. Therapeutic advances in respiratory disease. 2015;9(3):121-9.
13. Meyer KC. Pulmonary fibrosis, part I: epidemiology, pathogenesis, and diagnosis. Expert Rev Respir Med. 2017;11(5):343-59.
14. Meyer KC. Pulmonary fibrosis, part II: state-of-the-art patient management. Expert Rev Respir Med. 2017;11(5):361-76.
15. Vianello A, Molena B, Turato C, et al. Pirfenidone improves the survival of patients with idiopathic pulmonary fibrosis hospitalized for acute exacerbation. Current medical research and opinion. 2019:1.
16. Nakano A, Ohkubo H, Fukumitsu K, et al. Remarkable Improvement in a Patient with Idiopathic Pulmonary Fibrosis after Treatment with Nintedanib. Internal medicine (Tokyo, Japan). 2018.
17. National Institute for Health and Care Excellence. Nintedanib for treating idiopathic pulmonary fibrosis. Available at: https://www.nice.org.uk/guidance/ta379. Accessed: 8th October 2018.
18. National Institute for Health and Care Excellence. Pirfenidone for treating idiopathic pulmonary fibrosis. Available at: https://www.nice.org.uk/guidance/ta504/chapter/1-Recommendations. Accessed: 8th October 2018.
19. Raghu G, Rochwerg B, Zhang Y, et al. An Official ATS/ERS/JRS/ALAT Clinical Practice Guideline: Treatment of Idiopathic Pulmonary Fibrosis. An Update of the 2011 Clinical Practice Guideline. American journal of respiratory and critical care medicine. 2015;192(2):e3-19.
20. Hewson T, McKeever TM, Gibson JE, et al. Timing of onset of symptoms in people with idiopathic pulmonary fibrosis. Thorax. 2017.
21. OPCRD. The Optimum Patient Care Research Database (OPCRD). https://opcrd.co.uk/. Accessed: June 2018.

22. Colice G, Chisholm A, Dima AL, et al. Performance of database-derived severe exacerbations and asthma control measures in asthma: responsiveness and predictive utility in a UK primary care database with linked questionnaire data. Pragmatic and observational research. 2018;9:29-42.
23. Roche N, Reddel H, Martin R, et al. Quality standards for real-world research. Focus on observational database studies of comparative effectiveness. Annals of the American Thoracic Society. 2014;11 Suppl 2:S99-104.
24. Electronic Register of Studies, European Network of Centres for Pharmacoepidemiology and Pharmacovigilance (ENCePP). [cited 2015 February 20]. Available from: http://www.encepp.eu/encepp/studiesDatabase.jsp.
25. Hayward RA, Chen Y, Croft P, et al. Presentation of respiratory symptoms prior to diagnosis in general practice: a case-control study examining free text and morbidity codes. BMJ open. 2015;5(6):e007355.
26. Collard HR, Tino G, Noble PW, et al. Patient experiences with pulmonary fibrosis. Respiratory medicine. 2007;101(6):1350-4.
27. Purokivi M, Hodgson U, Myllarniemi M, et al. Are physicians in primary health care able to recognize pulmonary fibrosis? Eur Clin Respir J. 2017;4(1):1290339.
28. Hoyer N, Prior TS, Bendstrup E, et al. Risk factors for diagnostic delay in idiopathic pulmonary fibrosis. Respir Res. 2019;20(1):103.
29. Cottin V, Richeldi L. Neglected evidence in idiopathic pulmonary fibrosis and the importance of early diagnosis and treatment. Eur Respir Rev. 2014;23(131):106-10.
30. blueteq. Commissioner High Cost Drugs System. http://www.blueteq.com/commhcd. Accessed: 7 February 2019.

FIGURE LEGENDS

Figure 1. Flowchart of patient selection. †code for a consultation to a chest specialist followed by an IPF diagnosis (code for consultation allowed till up to 60 days after diagnosis). *Patients with registration date before date of birth.

Figure 2. Principal component analysis (PCA) based network chart for the association between codes within 90 and within 365 days prior to IPF diagnosis. Bubble size indicates prevalence, and thickness of lines indicate the degree of association between signs & symptoms.

Figure 3. Frequency of codes for signs and symptoms from 7 years prior to IPF diagnosis. 3A: All signs and symptoms, 3B: Co-occurring dyspnoea and cough, 3C: Excluding cough and dyspnoea.

Figure 4. Time course of respiratory symptoms and weight prior to the diagnosis of IPF in two patients showing a long duration of weight loss.