ORIGINAL ARTICLE
INTERSTITIAL LUNG DISEASE

Idiopathic pulmonary fibrosis cluster analysis highlights diagnostic delay and cardiovascular comorbidity association with outcome

Jaume Bordas-Martínez [1,2], Ricard Gavaldà[3,4], Jessica G. Shull[1], Vanesa Vicens-Zygmunt[1], Lurdes Planas-Cerezales[1], Guadalupe Bermudo-Peloche[1], Salud Santos[1,2], Neus Salord[2], Carmen Monasterio[2], Maria Molina-Molina[1] and Guillermo Suarez-Cuartin [1]

Affiliations: [1]Interstitial Lung Disease Unit, Respiratory Dept, Bellvitge University Hospital, IDIBELL, University of Barcelona, Hospitalet de Llobregat, Barcelona, Spain. [2]Sleep Unit, Respiratory Dept, Bellvitge University Hospital, IDIBELL, University of Barcelona, Hospitalet de Llobregat, Barcelona, Spain. [3]Amalfi Analytics, Barcelona, Spain. [4]Computer Science Dept, Polytechnic University of Catalonia, Barcelona, Spain.

Correspondence: Maria Molina-Molina, ILD Unit, Respiratory Dept, University Hospital of Bellvitge, IDIBELL, Hospitalet de Llobregat, 08907, Barcelona, Spain. E-mail: mariamolinamolina@hotmail.com

ABSTRACT

Introduction: Idiopathic pulmonary fibrosis (IPF) prognosis is heterogeneous despite antifibrotic treatment. Cluster analysis has proven to be a useful tool in identifying interstitial lung disease phenotypes, which has yet to be performed in IPF. The aim of this study is to identify phenotypes of IPF with different prognoses and requirements.

Methods: Observational retrospective study including 136 IPF patients receiving antifibrotic treatment between 2012 and 2018. Six patients were excluded due to follow-up in other centres. Cluster analysis of 30 variables was performed using approximate singular value-based tensor decomposition method and comparative statistical analysis.

Results: The cluster analysis identified three different groups of patients according to disease behaviour and clinical features, including mortality, lung transplant and progression-free survival time after 3-year follow-up. Cluster 1 (n=60) was significantly associated (p=0.02) with higher mortality. Diagnostic delay was the most relevant characteristic of this cluster, as 48% of patients had ⩾2 years from first respiratory symptoms to antifibrotic treatment initiation. Cluster 2 (n=22) had the longest progression-free survival time and was correlated to subclinical patients evaluated in the context of incidental findings or familial screening. Cluster 3 (n=48) showed the highest percentage of disease progression without cluster 1 mortality, with metabolic syndrome and cardiovascular comorbidities as the main characteristics.

Conclusion: This cluster analysis of IPF patients suggests that diagnostic and treatment delay are the most significant factors associated with mortality, while IPF progression was more related to metabolic syndrome and cardiovascular comorbidities.

@ERSpublications
Diagnostic delay and cardiovascular comorbidities impact IPF outcomes https://bit.ly/3lk2Z5y

Cite this article as: Bordas-Martínez J, Gavaldà R, Shull JG, *et al.* Idiopathic pulmonary fibrosis cluster analysis highlights diagnostic delay and cardiovascular comorbidity association with outcome. *ERJ Open Res* 2021; 7: 00897-2020 [https://doi.org/10.1183/23120541.00897-2020].

This article has supplementary material available from openres.ersjournals.com

Received: 2 Dec 2020 | Accepted: 7 March 2021

Introduction

Idiopathic pulmonary fibrosis (IPF) is the most frequent and lethal interstitial lung disease (ILD) [1]. The development of antifibrotic treatments have increased the expected survival of IPF patients [2]. Given the overall poor quality of life of these patients, holistic care could impact daily life expectations. In order to improve current patient approaches, it is necessary not only to understand the disease, but also to assess different aspects of individual patients. The study of comorbidities [3–6], lifestyle and psycho-emotional accompaniment for patients and family members [7] is fundamental in the comprehensive treatment of the patient. In this regard, different multivariable risk prediction models such as the gender–age–physiology (GAP) model [8] or, more recently, the TORVAN model [5], have been created to predict the risk of death using different clinical data, lung functional tests and the presence of comorbidities. However, several other patient characteristics and healthcare features may have a significant role in disease outcome and patient needs.

Due to the heterogeneity of IPF presentation and progression, different phenotypes related to disease behaviour and comorbidities have been explored [3, 4]. Some proposed phenotypes that present specific disease behaviour are rapidly progressive IPF and combined pulmonary fibrosis and emphysema (CPFE) [3, 9]. The impact of comorbidities on disease behaviour and mortality has been explored, proposing the term "comorbidome" [10]. Most IPF cases present with more than two comorbidities [10]. Cardiovascular diseases, pulmonary hypertension and lung cancer are the comorbidities with the highest impact on IPF mortality [10, 11]. However, some biological disorders may be a common trigger of different comorbidities in the same patient, such as metabolic syndrome or telomeric disorders [3, 9]. Cluster analysis has become a useful resource to identify homogeneous patients with similar clinical characteristics, prognosis and healthcare requirements [12, 13]. Additionally, the integrated study of respiratory diseases through clusters [14] helps to identify hidden and unsuspected associations between different diseases and patient features, which could generate new hypotheses to be later explored in controlled studies. Previous analysis of chronic ILDs suggested distinct phenotypes which identified some meaningful clinical outcomes independent of disease diagnosis [13]. Furthermore, the better understanding of IPF patient profiles, including the different components that could influence patient needs, such as disease behaviour, comorbidities and patient condition, would optimise patient management.

Therefore, the aim of this study is to find hidden and/or unexpected associations in clusters of IPF patients based on common disease and patient features.

Methodology

This is an observational retrospective study analysing IPF patients treated with antifibrotic therapy. 136 IPF patients were treated with pirfenidone or nintedanib at the ILD unit of Bellvitge Hospital (Barcelona, Spain) from 2012 to 2018. Of these, six were followed-up in another centre (figure 1). The diagnosis [1] and treatment [15] of the 130 included IPF patients were performed according to the American Thoracic Society/European Respiratory Society criteria [9] by the multidisciplinary committee.

Demographic variables collected were age, gender, body mass index, previous exposures (smoking habit, occupational and environmental exposures), clinical data (dyspnoea, cough, crackles, nail clubbing), family history, comorbidities, pharmacological treatments, radiological pattern and hiatal hernia [16] on chest high-resolution computed tomography (HRCT) (hiatal hernia type II–IV with presence of air and food/fluid/air-fluid level in the oesophagus were considered moderate and severe hiatal hernia, respectively) [16], laboratory tests, sleep study (video polysomnography or respiratory polygraphy), echocardiography, telomere length and lung biopsy when required. Telomere length analysis was performed using DNA samples isolated from mouth epithelial cells (oral swabs: Isohelix, SK-2S; Cell Projects Ltd) and peripheral blood mononuclear cells (Isohelix) [17]. Telomere length was considered shortened when z-score was below the 25th percentile, and severe telomere shortening when below the 10th percentile [17]. Patients

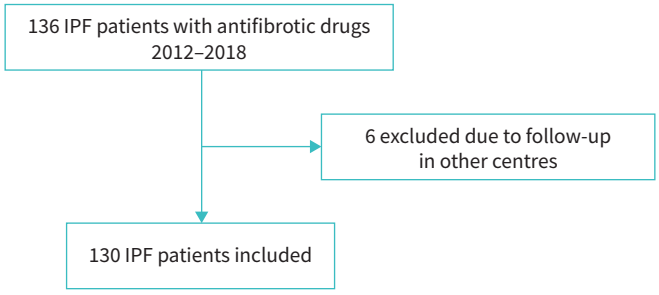

FIGURE 1 Study flow chart. IPF: idiopathic pulmonary fibrosis.

underwent pulmonary function tests including body plethysmography and spirometry, and 6-min walk tests at the time of diagnosis and thereafter every 3 months. Furthermore, forced vital capacity (FVC) and diffusing capacity of the lung for carbon monoxide (D_{LCO}) were collected before starting antifibrotic treatment. Frequent respiratory infections were defined when more than two respiratory infections with antibiotic requirement per year were present. Acute exacerbations that required hospital admission were defined following the current recommendations, regardless of the trigger [18]. Antifibrotic treatment (pirfenidone and nintedanib), adverse events and subsequent management were followed for 1 year. Family aggregation, comorbidities, treatment-related side-effects and drug compliance, and lung transplant or death due to IPF were recorded. Disease progression was defined as FVC decline ⩾10% predicted or

TABLE 1 Patient features at diagnosis

Subjects	130
Age years	69±7.8
Male	105 (80.8)
BMI kg·m^{-2}	
<18.5	2 (1.5)
18.5–24.9	16 (12.3)
25–29.9	69 (53.1)
⩾30	42 (32.3)
Smoking exposure pack-years	
<20	33 (25.4)
⩾20	60 (46.2)
Alcohol	
Active	15 (11.5)
Former	6 (4.6)
Reason for consultation	
Respiratory symptoms	85 (65.4)
Familiar study	10 (7.7)
Radiological or primary care protocol (without a respiratory clinic)	35 (26.9)
Cough	85 (65.4)
Dyspnoea (mMRC)	
0	22 (16.9)
1	55 (42.3)
2	42 (32.3)
3	11 (8.5)
Crackles	118 (90.8)
Clubbing fingers	65 (50.0)
Thorax HRCT pattern	
Definite UIP pattern	60 (46.2)
Probable UIP pattern	55 (42.3)
Pattern indeterminate for UIP	15 (11.5)
Biopsy	
Cryobiopsy	4 (3.1)
Surgical biopsy	48 (36.9)
Anatomopathological patterns	
Definite UIP	35 (26.9)
Probable UIP	10 (7.7)
Possible UIP	6 (4.6)
Nonrepresentative	1 (0.8)
Severe physiological limitation at diagnosis	
FVC <50%	2 (1.5)
D_{LCO} <30%	11 (8.4)
6MWD <350 m	20 (15.4)
Family aggregation	36 (27.7)
Telomere shortening	23 (17.7)
Antifibrotic treatment	
Pirfenidone	63 (48.5)
Nintedanib	67 (51.5)

Data are presented as n, mean±SD or n (%). BMI: body mass index; mMRC: Modified Medical Research Council; HRCT: high-resolution computed tomography; UIP: usual interstitial pneumonia; FVC: forced vital capacity; D_{LCO}: diffusing capacity for carbon monoxide; 6MWD: 6-min walk distance.

D_{LCO} ⩾15% pred in 1 year. Progression-free survival (PFS) after 3-year follow-up was defined as no progression, lung transplant or death in 3 years of follow-up.

This study was approved by the ethics committee of Bellvitge University Hospital (reference code PR413/18). The study was performed in accordance with the ethical principles of the Declaration of Helsinki, and local laws of countries in which the research was done. Informed consent was obtained from each participant by the study investigator before patient data collection was done.

Cluster and statistical analysis

Clustering was performed using the MATE tool by Amalfi Analytics. Patients were clustered using approximate singular value-based tensor decomposition (ASVTD) method described in Ruffini *et al.* [14], which takes as input a table where each row corresponds to a patient and each column corresponds to an observed variable on patients, such as a diagnostic, a clinical result, demographics such sex and age, *etc.*, plus a number of k desired clusters. This results in the description of the k clusters found, where each cluster is described by the average value of each variable in it. Each patient (in the dataset, or newly arriving patients) can then be assigned to the most-aligned cluster.

This method produces clusters based on logical weight of given attributes. Compared to distance- or similarity-based clustering methods (k-means, k-medoids or partitioning around medoids, dendograms),

TABLE 2 Patient comorbidities at diagnosis

Cardiovascular risk factor	
>1	51 (39.2)
Arterial hypertension	68 (52.3)
Dyslipidaemia	58 (44.6)
Diabetes mellitus	28 (21.5)
GORD	59 (45.4)
Hiatus hernia (CT measure)	
Mild	65 (50.0)
Moderate	33 (25.4)
Severe	7 (5.4)
Emphysema	43 (33.1)
CPFE	14 (10.8)
COPD	15 (11.5)
Heart disease	30 (23.1)
Valvular	10 (7.7)
Ischaemic	20 (15.4)
Arrhythmia	9 (6.9)
Pulmonary hypertension by echocardiography	41 (31.5)
PAH, treated	8 (6.2)
Sleep disorder	29 (22.3)
Severe OSA (AHI ⩾30 events·h^{-1})	14 (11)
Mild or moderate OSA (AHI <30 events·h^{-1})	8 (6.2)
Sleep-related hypoxaemia	13 (10)
Malignant disease	16 (12.3)
Lung	3 (2.3)
Urogenital	6 (4.6)
Digestive	5 (3.8)
Other	2 (1.6)
Depression/anxiety	12 (9.2)
Chronic kidney failure grade ⩾2	11 (8.5)
Neurological	9 (6.9)
Degenerative	2 (1.5)
Ischaemic	8 (6.2)
Peripheral vasculopathy	9 (6.9)
Liver disease	8 (6.2)
Peptic ulcer	6 (4.6)
Pulmonary embolism	3 (2.3)
Charlson's comorbidity index	4.7±1.7

Data are presented as n (%) or mean±SD. GORD: gastro-oesophageal reflux disease; CT: computed tomography; CPFE: combined pulmonary fibrosis and emphysema; PAH: pulmonary arterial hypertension; OSA: obstructive sleep apnoea; AHI: apnoea–hypopnoea index.

MATE is known to work better in the presence of irrelevant or noisy attributes, and does not require the definition of an *a priori* "similarity" function to be used (such as Euclidean distance) [14]. The task of choosing a final number of clusters is left to the user, combining the intuitive meaning of each cluster plus the usual requirement to have a small number of clusters. Additional information about cluster analysis is available in the supplementary material.

SPSS for Windows 25.0 (IBM, USA) was used for noncluster statistical analysis. For descriptive analysis, frequency and percentage were used for the categorical variables, and mean±SD or median (interquartile range) for continuous variables, when appropriate. For comparative analysis of categorical variables, Chi-squared test or Fisher's exact test were used when required. For continuous variables, ANOVA or the corresponding nonparametrical test were used when appropriate. Time-to-event data (time to lung transplant and/or death) were analysed using Kaplan–Meier survival analysis. A p-value <0.05 was considered statistically significant. Strengthening the Reporting of Observational Studies in Epidemiology initiative recommendations were followed [19].

Results

Patient features

Baseline characteristics of the 130 patients enrolled are shown in table 1. The mean±SD age was 69±7.8 years, and 81% were male. Regarding toxic habits, 72% had smoking exposure, of whom 46% had a cumulative dose associated with an IPF risk factor (⩾20 pack-years) [20] and 12% had a history of alcohol abuse (three or more standard drinks per day). 33% of cases were obese and 1.5% were underweight. 65% of patients were referred because of respiratory symptoms. Exertional dyspnoea was present in 83% of patients at diagnosis, and 65% referred dry cough. Velcro crackles on chest auscultation and clubbing finger were present in 91% and 50% of patients, respectively. 46% of patients showed a consistent usual interstitial pneumonia (UIP) pattern on chest HRCT, 42% showed a probable UIP pattern and 12% indeterminate pattern for UIP. Lung biopsy was performed in 52 cases; 48 surgical biopsy and four cryobiopsies (table 1). Telomere length analysis had been performed on 79 patients with family aggregation or some telomeric clinical sign. Familial aggregation was identified in 28% of cases and telomere shortening was recognised in 18% of patients.

The main comorbidities at diagnosis are shown in table 2. Charlson's comorbidity index was 4.7±1.7. Cardiovascular risk factors were prevalent, and 39% of cases had at least two factors: arterial hypertension (52%), dyslipidaemia (45%) or diabetes mellitus (22%). Symptomatic gastro-oesophageal reflux disease

TABLE 3 Patient follow-up and outcomes

	0 years	1 year	2 years	3 years
Patients	130	113	80	50
Pulmonary function tests				
FVC mL	2742±821	2738±807	2745±854	2658±860
FVC %	82.6±17.5	85.2±19.3	85.8±21.2	81.7±19.5
TLC mL	4744±1183	4526±1212	4397±1179	4399±1152
TLC %	78.9±15.7	76.3±15.6	74.8±14.7	73.1±15.5
D_{LCO} %	50.8±16.8	52.9±17.8	53.6±17.4	52.0±15.9
K_{CO} %	75.4±20.7	78.7±23.4	84.0±27.6	82.9±20.8
6MWD m	429±88.7	427±104.1	433±94.1	437±94.2
Respiratory infection pattern (⩾2 respiratory infections per year in ⩾2 years of the 3 years of follow-up)				23 (17.7)
Acute exacerbation (hospital admission)				28 (21.5)
Antifibrotic stop or switch due to adverse effects				44 (33.8)
Antifibrotic stop or switch due to IPF progression				6 (4.6)
Progression-free survival				55 (42.3)
FVC progression (decrease ⩾10%)				42 (32.3)
D_{LCO} progression (decrease ⩾15%)				20 (15.4)
Lung transplant by IPF progression				12 (9.2)
Death by IPF progression				25 (19.2)

Data are presented as n, mean±SD or n (%). FVC: forced vital capacity; TLC: total lung capacity; D_{LCO}: diffusing capacity of the lung for carbon monoxide; K_{CO}: transfer coefficient of the lung for carbon monoxide; 6MWD: 6-min walk distance; IPF: idiopathic pulmonary fibrosis.

(GORD) was referred by 45% of patients, while hiatal hernia measured by HRCT was 5% severe and 25% moderate. Emphysema was detected in 33% of patients, but only 11% satisfied the CPFE diagnostic criteria [21]. Heart disease was found in 23% of the participants, most of them in the form of ischaemic cardiomyopathy (15%). Pulmonary arterial hypertension (PAH) was suspected upon echocardiography in 32% of patients, but only 6% had PAH by right catheterisation and received specific treatment. Sleep studies were performed on 29 patients who presented clinical symptoms of obstructive sleep apnoea (OSA), of whom 14 were OSA under continuous positive airway pressure (CPAP) treatment and 13 diagnosed with sleep-related hypoxaemia (peripheral oxygen saturation ⩽88% for ⩾5 min) according to International Classification of Sleep Disorders criteria [22].

Patient follow-up and outcomes are depicted in table 3. At the initiation of antifibrotic treatment, most patients presented preserved or mildly decreased FVC, but severe D_{LCO} deterioration. After 3-year follow-up, 18% of subjects had at least two respiratory infections per year in a minimum of 2 years without requiring hospital admission; 22% suffered an acute exacerbation requiring hospital admission. 34% stopped or switched antifibrotic drug due to adverse effects and 5% altered protocol due to IPF progression. 42% of patients didn't show disease progression after 3 years; 32% showed a decline of FVC ⩾10% pred and 15% a decline in D_{LCO} ⩾15% pred in 1 year. Lung transplant or death related to IPF progression was observed in 28% of cases (9% and 19%, respectively).

IPF clustering

The cluster analysis identified three different types of patient groups, aggregating 60, 22 and 48 cases in each group. This clustering grouped the cases by similar disease behaviour and patient features, including death, lung transplant and PFS after 3-year follow-up. The characteristics of each cluster are shown in figure 2. Furthermore, values and significance of each variable is exposed in table 4.

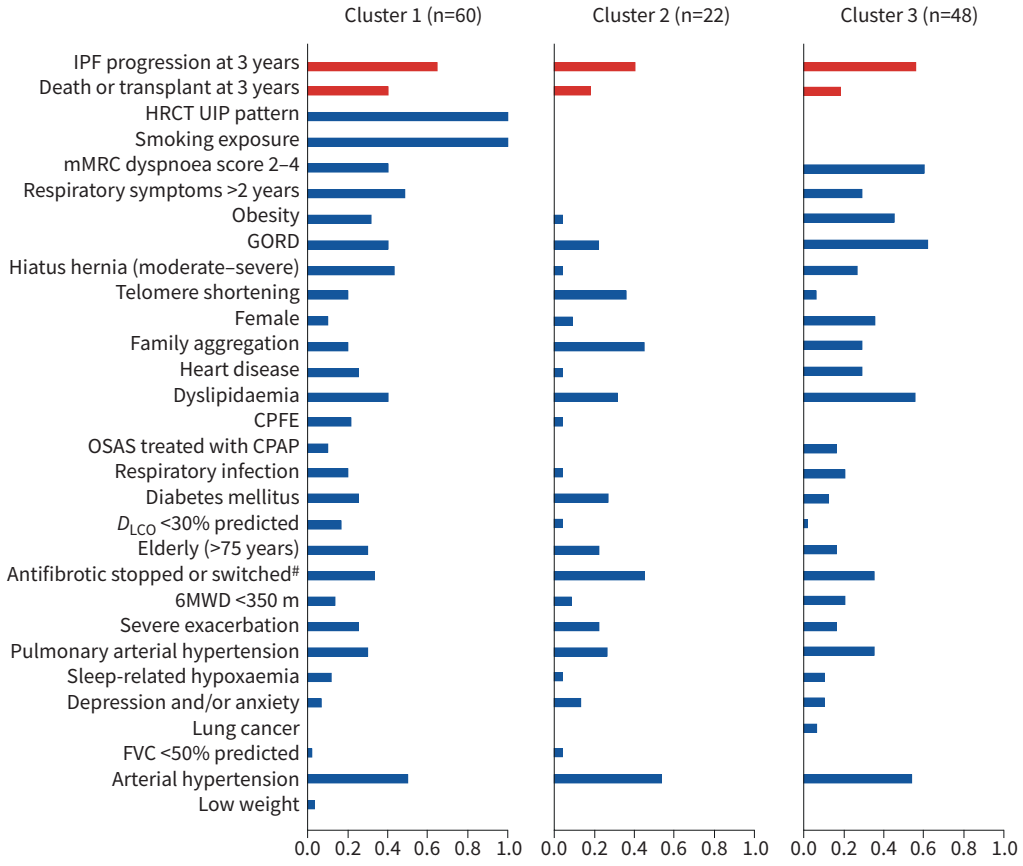

FIGURE 2 Cluster analysis. IPF: idiopathic pulmonary fibrosis; HRCT: high-resolution computed tomography; UIP: usual interstitial pneumonia; mMRC: modified Medical Research Council; GORD: gastro-oesophageal reflux disease; CPFE: combined pulmonary fibrosis and emphysema; OSAS: obstructive sleep apnoea syndrome; CPAP: continuous positive airway pressure; D_{LCO}: diffusing capacity of the lung for carbon monoxide; 6MWD: 6-min walk distance; FVC: forced vital capacity. #: due to both adverse effect of antifibrotic drug treatment and IPF progression.

TABLE 4 Cluster differences analysis

	Cluster 1 "delayed treatment"	Cluster 2 "early diagnosis"	Cluster 3 "cardiovascular comorbidity"	p-value
Patients n	60	22	48	
Age years	68.71±9.32	70.34±6.40	69.47±6.28	0.691
Age >75 years	18 (30)	5 (22.7)	8 (16.7)	0.268
Male	54 (90)	20 (90.9)	31 (64.6)	0.002*
Death or transplant at 3 years	24 (40)	4 (18.2)	9 (18.8)	0.026*
Survival time weeks, median (IQR)	113.0 (108.8)	160.5 (132.8)	134.0 (103.3)	0.084
Time from symptoms to diagnosis weeks, median (IQR)	104.00 (118)	48.00 (18.0)	54.50 (86.0)	0.007*
Respiratory symptoms >2 years	29 (48.3)	0	14 (29.17)	<0.001*
Progression at 3 years	39 (65)	9 (40.9)	27 (56.3)	0.142
⩾2 respiratory infections	12 (20)	1 (4.6)	10 (20.8)	0.219
⩾1 severe exacerbation	15 (25)	5 (22.7)	8 (16.7)	0.557
Charlson index, median (IQR)	4.50 (3.0)	4.00 (1.0)	5.00 (2.0)	0.668
FVC <50% predicted	1 (1.7)	1 (4.5)	0	0.445
D_{LCO} <30% predicted	9 (15.3)	1 (4.5)	1 (2.1)	0.034*
6MWD <350 m	8 (13.3)	2 (10)	10 (20.8)	0.426
Definite HRCT UIP pattern	60 (100)	0	0	<0.001*
Probable HRCT UIP pattern	0	14 (63.6)	41 (85.4)	<0.001*
CPFE	13 (21.7)	1 (4.54)	0	<0.001*
Familial pulmonary fibrosis	12 (20)	10 (45.5)	14 (29.2)	0.071
Telomere shortening	12 (20)	8 (36.4)	3 (6.3)	0.006*
Obesity	19 (31.7)	1 (4.5)	22 (45.8)	0.002*
Low weight	2 (3.3)	0	0	0.656
mMRC dyspnoea score 2–3	24 (40)	0	29 (60.4)	<0.001*
Smoking exposure (⩾20 pack-years)	60 (100)	0	0	<0.001*
Smoking exposure (<20 pack-years)	0	8 (36.4)	25 (52.1)	<0.001*
OSAS treated with CPAP	6 (10)	0	8 (16.7)	0.096
Nocturnal hypoxaemia	7 (11.7)	1 (4.5)	5 (10.4)	0.741
HRCT moderate and severe hiatal hernia	26 (43.3)	1 (4.5)	13 (27.1)	0.002*
Cardiopathy	15 (25)	1 (4.5)	14 (29.2)	0.067
Pulmonary arterial hypertension	18 (30)	6 (27.3)	17 (35.4)	0.746
Antifibrotic treatment stopped or switched	20 (33.3)	10 (45.5)	17 (35.4)	0.593
Antifibrotic treatment stopped	18 (30)	3 (13.6)	13 (27.1)	0.322
Antifibrotic treatment length weeks, median (IQR)	90 (99.3)	153 (41.5)	115 (85.8)	0.009*
Radiological findings without respiratory symptoms	14 (23.3)	11 (50)	10 (20.8)	0.026*

Data are presented as n, mean±SD or n (%), unless otherwise stated. % is in relation to the cluster concerned. IQR: interquartile range; FVC: forced vital capacity; D_{LCO}: diffusing capacity of the lung for carbon monoxide; 6MWD: 6-min walk distance; HRCT: high-resolution computed tomography; UIP: usual interstitial pneumonia; CPFE: combined pulmonary fibrosis and emphysema; mMRC: modified Medical Research Council; OSAS: obstructive sleep apnoea syndrome; CPAP: continuous positive airway pressure. *: p<0.05.

Cluster 1 was significantly associated (p=0.02) with higher mortality, as shown in the Kaplan–Meier PFS curve (figure 3). 40% of patients in this cluster died or underwent lung transplantation after 3-year follow-up. Median (interquartile range) survival time was 113 (109) weeks. It is remarkable that 48% of the cluster presented a delay of >2 years from the first symptom to beginning antifibrotic treatment. Interestingly, the whole cluster presented tobacco exposure of ⩾20 pack-years and UIP pattern on chest HRCT at diagnosis. Additionally, it included nearly all CPFEs (22% of the cluster) and more severe D_{LCO} decrease at diagnosis (15%). The highest percentage of moderate–severe hiatal hernia measured by HRCT (43%) was included in this cluster. Finally, this was the only group with low weight (two patients).

Cluster 2 had the longest PFS and it was predominantly characterised by having <2 years delay from the symptoms to the beginning of the antifibrotic treatment, no smoking history and no clear factor for comorbidity. This cluster has the highest percentage of ILD suspicion due to incidental findings in a radiological study by a nonrespiratory cause (50%) or screening in the context of subclinical family aggregation (45%). These patients did not present consistent UIP pattern on chest HRCT and only a minority of cases had two or more respiratory infections per year.

Cluster 3 showed the highest percentage of disease progression, but with lower mortality than cluster 1. Cluster 3 was characterised by a high rate of metabolic syndrome, including dyslipidaemia (56%), obesity (46%) and arterial hypertension (54%). Severe OSA with CPAP treatment was present in 17% of cases. Cardiomyopathy was observed in 29% of cases. Although mean FVC and D_{LCO} were not severely decreased and no consistent UIP pattern was present at diagnosis, moderate or severe dyspnoea was referred in 60% of cases and 21% showed a relevant limitation for exercise capacity (6-min walk distance <350 m). GORD was present in most cases (60%).

Discussion

The cluster analysis of this IPF cohort identifies three different types of patients with similar clinical features and disease behaviour. Cluster 1 presents the worst 3-year survival rate and involves patients with diagnostic and treatment delay, consistent UIP pattern, smoking history and emphysema. Although clusters 2 and 3 present a similar prognosis, patient features are different. Cluster 3 includes predominantly patients with obesity and other associated comorbidities such as cardiovascular diseases and OSA syndrome. Different clinical features and comorbidities may impact on quality of life and prognosis of IPF patients [10]. Inclusion of comorbidities in a new multivariable model for predicting the risk of progression (TORVAN) [5] is an improvement over previous models such as GAP [8]. Furthermore, distinct phenotypes with different clinical outcomes and healthcare requirements have been suggested using clustering analysis of chronic ILDs [13]. Therefore, clustering IPF by patient features and disease behaviour may also help predict outcome and identify patients' needs. This model elucidates associations between clinical data, comorbidities and evolution by clinical clusters.

Diagnostic and treatment delay are the most outstanding factors of cluster 1; 48% of patients had an average wait time of >2 years from the first respiratory symptoms to antifibrotic treatment initiation. This diagnostic delay has been described by LAMAS *et al.* [23] as an independent variable of mortality, and notes the highest percentage of former smokers in the group with a 2–4-year delay. Interestingly, all patients in cluster 1 were ex-smokers. Thus, tobacco is a risk factor in pulmonary fibrosis and emphysema [1], but it could also be a confounding factor that causes a delay in the assessment of respiratory symptoms. A recent study found the use of inhaled therapy as the most important risk factor for delayed IPF diagnosis [24]. Although cluster 1 had 30% of patients aged >75 years, and it is possible that the age may impact on the time to refer patient symptoms, no significant differences in age between clusters were observed. A high rate of consistent UIP pattern in the chest HRCT at diagnosis has been associated with the delay in IPF diagnosis [25]. Similar to HOYER *et al.* [24], our study shows a predominance of consistent UIP pattern and lower FVC and D_{LCO} at diagnosis in these patients. Another factor that has reported a poor survival rate is the presence of CPFE [26, 27], which is present in 22% of cases clustered in this group. Furthermore, hiatal hernia is more frequently observed in this group. An increased incidence of hiatal

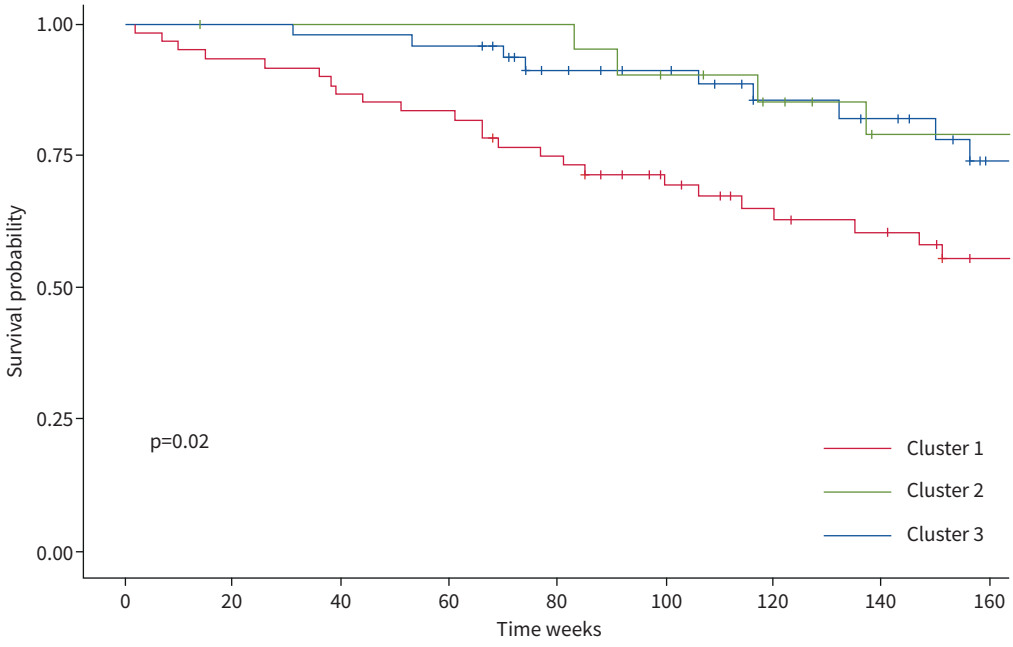

FIGURE 3 Kaplan–Meier progression-free survival curve at 3 years by cluster.

hernia measured by HRCT in IPF [28] and its association to a worse prognosis has been described previously [16]. Regarding the results, the diagnostic delay may also associate low patient weight at diagnosis, which has been identified as a poor prognostic factor [25].

Cluster 2 included a low number of patients with greater survival time (160 weeks on average), longer antifibrotic drug treatment time (median 153 weeks) and a low rate of disease progression (41%), which may be related to early diagnosis as the main characteristic of this group. The mean time from the onset of respiratory symptoms to the antifibrotic treatment initiation was 48 weeks and none exceeded 2 years. This could be due to the rate of subclinical patients evaluated in the context of incidental findings or familial screening that have been clustered in this group. The familial study could explain the increased significant telomere shortening in this cluster (36%). Although a minority of new diagnosed cases, these patients could be better managed with preventive measures and comprehensive therapeutic approaches [25, 29].

Metabolic syndrome and cardiovascular comorbidities were the main features of cluster 3, which associates a high rate of disease progression. The association between obesity, dyslipidaemia, cardiomyopathy, reduced physical capacity and exertional dyspnoea has been well documented [30, 31]. Obesity and the high prevalence of severe OSA could explain the higher rates of cardiovascular comorbidities [31, 32]. Cardiovascular risk factors have also been associated with menopause [33]. This cluster included the majority of women in our cohort. The higher prevalence of severe OSA under CPAP treatment could be explained by the predominance of obesity [34]. In this cluster, 72% of sleep study subjects were diagnosed with OSA, a disproportionate prevalence, as it is similar to morbid obesity series [35]. It would suggest a possible underdiagnosis of these disorders and the potential need for systematic screening in these types of IPF patients [36–38]. It is likely that obesity plays a major role in the higher incidence of GORD, as described previously [39, 40]. At the same time, GORD can be another risk factor for disease progression and acute exacerbations [41].

The number of patients included in the cluster analysis from a single centre and the retrospective nature are the main limitations of this pilot study. Another limitation is the inclusion of patients from a broad period of time, which may have had an impact on time to referral, patient management and clinical outcomes. However, only 13 patients were included between 2012 and 2013, when access to antifibrotic treatment was limited and awareness of the disease lower. Cluster analysis should ideally be performed on large multinational cohorts of >1000 patients to identify as many patient profiles as possible [12, 13]. However, the highlighted disease and patient features at diagnosis associated with disease outcome by using this methodology that integrates all potential risk factors have revealed at least two major points: diagnostic delay and cardiovascular–metabolic comorbidities. These results should be validated and better explored in prospective multicentre studies.

In conclusion, this cluster study helps analyse IPF patients, a population which consistently presents a complex variability of features at diagnosis related to the disease, comorbidities and other patient-related conditions, and automatically clusters them depending on similar features and disease behaviour. With further work, cluster studies could identify intricate associations invisible without analysis. Therefore, the cluster analysis at diagnosis could identify different groups of IPF patients that would benefit from a better personalised management and therapeutic approach, which would be useful for anticipating patient needs and required resources.

Conflict of interest: J. Bordas-Martínez reports FIS-ISCIII grant PI18/00367 (cofunded by the European Regional Development Fund (ERDF)), Spanish Society of Pneumology and Thoracic Surgery (SEPAR) grants 631/2018 and 685/2018, Emerging ILD Group of SEPAR grant 005 (Boehringer–Roche), Pneumology Foundation of Catalonia grant 2019, Spanish Sleep Society grant 2019, institutional support of the CERCA Programme/Generalitat de Catalunya, and research support BRN-Fundació Ramon Pla Armengol from ISCIII grant CM20/00093 (cofunded by European Regional Development Fund), during the conduct of the study. R. Gavaldà reports FIS-ISCIII grant PI18/00367 (cofunded by the European Regional Development Fund (ERDF)), Spanish Society of Pneumology and Thoracic Surgery (SEPAR) grants 631/2018 and 685/2018, Emerging ILD Group of SEPAR grant 005 (Boehringer–Roche), Pneumology Foundation of Catalonia grant 2019, Spanish Sleep Society grant 2019, institutional support of the CERCA Programme/Generalitat de Catalunya, and research support BRN-Fundació Ramon Pla Armengol from ISCIII grant CM20/00093 (cofunded by European Regional Development Fund), during the conduct of the study. J.G. Shull reports FIS-ISCIII grant PI18/00367 (cofunded by the European Regional Development Fund (ERDF)), Spanish Society of Pneumology and Thoracic Surgery (SEPAR) grants 631/2018 and 685/2018, Emerging ILD Group of SEPAR grant 005 (Boehringer–Roche), Pneumology Foundation of Catalonia grant 2019, Spanish Sleep Society grant 2019, institutional support of the CERCA Programme/Generalitat de Catalunya, and research support BRN-Fundació Ramon Pla Armengol from ISCIII grant CM20/00093 (cofunded by European Regional Development Fund), during the conduct of the study. V. Vicens-Zygmunt reports FIS-ISCIII grant PI18/00367 (cofunded by the European Regional Development Fund (ERDF)), Spanish Society of Pneumology and Thoracic Surgery (SEPAR) grants 631/2018 and 685/2018, Emerging ILD Group of SEPAR grant 005 (Boehringer–Roche), Pneumology Foundation of Catalonia grant 2019, Spanish Sleep Society grant 2019, institutional support of the CERCA Programme/Generalitat de Catalunya, and research support BRN-Fundació Ramon Pla Armengol from ISCIII grant CM20/00093 (cofunded by European Regional Development Fund), during the conduct of

the study. L. Planas-Cerezales reports FIS-ISCIII grant PI18/00367 (cofunded by the European Regional Development Fund (ERDF)), Spanish Society of Pneumology and Thoracic Surgery (SEPAR) grants 631/2018 and 685/2018, Emerging ILD Group of SEPAR grant 005 (Boehringer–Roche), Pneumology Foundation of Catalonia grant 2019, Spanish Sleep Society grant 2019, institutional support of the CERCA Programme/Generalitat de Catalunya, and research support BRN-Fundació Ramon Pla Armengol from ISCIII grant CM20/00093 (cofunded by European Regional Development Fund), during the conduct of the study. G. Bermudo-Peloche reports FIS-ISCIII grant PI18/00367 (cofunded by the European Regional Development Fund (ERDF)), Spanish Society of Pneumology and Thoracic Surgery (SEPAR) grants 631/2018 and 685/2018, Emerging ILD Group of SEPAR grant 005 (Boehringer–Roche), Pneumology Foundation of Catalonia grant 2019, Spanish Sleep Society grant 2019, institutional support of the CERCA Programme/Generalitat de Catalunya, and research support BRN-Fundació Ramon Pla Armengol from ISCIII grant CM20/00093 (cofunded by European Regional Development Fund), during the conduct of the study. S. Santos reports FIS-ISCIII grant PI18/00367 (cofunded by the European Regional Development Fund (ERDF)), Spanish Society of Pneumology and Thoracic Surgery (SEPAR) grants 631/2018 and 685/2018, Emerging ILD Group of SEPAR grant 005 (Boehringer–Roche), Pneumology Foundation of Catalonia grant 2019, Spanish Sleep Society grant 2019, institutional support of the CERCA Programme/Generalitat de Catalunya, and research support BRN-Fundació Ramon Pla Armengol from ISCIII grant CM20/00093 (cofunded by European Regional Development Fund), during the conduct of the study. N. Salord reports FIS-ISCIII grant PI18/00367 (cofunded by the European Regional Development Fund (ERDF)), Spanish Society of Pneumology and Thoracic Surgery (SEPAR) grants 631/2018 and 685/2018, Emerging ILD Group of SEPAR grant 005 (Boehringer–Roche), Pneumology Foundation of Catalonia grant 2019, Spanish Sleep Society grant 2019, institutional support of the CERCA Programme/Generalitat de Catalunya, and research support BRN-Fundació Ramon Pla Armengol from ISCIII grant CM20/00093 (cofunded by European Regional Development Fund), during the conduct of the study. C. Monasterio reports FIS-ISCIII grant PI18/00367 (cofunded by the European Regional Development Fund (ERDF)), Spanish Society of Pneumology and Thoracic Surgery (SEPAR) grants 631/2018 and 685/2018, Emerging ILD Group of SEPAR grant 005 (Boehringer–Roche), Pneumology Foundation of Catalonia grant 2019, Spanish Sleep Society grant 2019, institutional support of the CERCA Programme/Generalitat de Catalunya, and research support BRN-Fundació Ramon Pla Armengol from ISCIII grant CM20/00093 (cofunded by European Regional Development Fund), during the conduct of the study. M. Molina-Molina reports FIS-ISCIII grant PI18/00367 (cofunded by the European Regional Development Fund (ERDF)), Spanish Society of Pneumology and Thoracic Surgery (SEPAR) grants 631/2018 and 685/2018, Emerging ILD Group of SEPAR grant 005 (Boehringer–Roche), Pneumology Foundation of Catalonia grant 2019, Spanish Sleep Society grant 2019, institutional support of the CERCA Programme/Generalitat de Catalunya, and research support BRN-Fundació Ramon Pla Armengol from ISCIII grant CM20/00093 (cofunded by European Regional Development Fund), during the conduct of the study. G. Suarez-Cuartin reports FIS-ISCIII grant PI18/00367 (cofunded by the European Regional Development Fund (ERDF)), Spanish Society of Pneumology and Thoracic Surgery (SEPAR) grants 631/2018 and 685/2018, Emerging ILD Group of SEPAR grant 005 (Boehringer–Roche), Pneumology Foundation of Catalonia grant 2019, Spanish Sleep Society grant 2019, institutional support of the CERCA Programme/Generalitat de Catalunya, and research support BRN-Fundació Ramon Pla Armengol from ISCIII grant CM20/00093 (cofunded by European Regional Development Fund), during the conduct of the study.

Support statement: This study was funded by Instituto de Salud Carlos III through grants CM20/00093 (cofunded by the European Social Fund) and PI18/00367 (cofunded by European Regional Development Fund), Spanish Society of Pneumology and Thoracic Surgery (SEPAR) grants 631/2018 and 685/2018, Emerging ILD Group of SEPAR grant 005 (Boehringer–Roche), Pneumology Foundation of Catalonia (FUCAP) grant 2019, Spanish Sleep Society (SES) grant 2019, and Investigation Support BRN-Fundació Ramon Pla Armengol. We thank CERCA Programme/Generalitat de Catalunya for institutional support. Funding information for this article has been deposited with the Crossref Funder Registry.

References

1 Travis WD, King TE, Bateman ED, *et al.* American Thoracic Society/European Respiratory Society international multidisciplinary consensus classification of the idiopathic interstitial pneumonias. *Am J Respir Crit Care Med* 2002; 165; 277–304.
2 Fisher M, Nathan SD, Hill C, *et al.* Predicting life expectancy for pirfenidone in idiopathic pulmonary fibrosis. *J Manag Care Spec Pharm* 2017; 23: S17–S24.
3 Sauleda J, Núñez B, Sala E, *et al.* Idiopathic pulmonary fibrosis: epidemiology, natural history, phenotypes. *Med Sci* 2018; 6: 110.
4 Buendía-Roldán I, Mejía M, Navarro C, *et al.* Idiopathic pulmonary fibrosis: clinical behaviour and aging associated comorbidities. *Respir Med* 2017; 129: 46–52.
5 Torrisi SE, Ley B, Kreuter M, *et al.* The added value of comorbidities in predicting survival in idiopathic pulmonary fibrosis: a multicentre observational study. *Eur Respir J* 2019; 53: 1801587.
6 Caminati A, Lonati C, Cassandro R, *et al.* Comorbidities in idiopathic pulmonary fibrosis: an underestimated issue. *Eur Respir Rev* 2019; 28: 190044.
7 Barratt SL, Morales M, Spiers T, *et al.* Specialist palliative care, psychology, interstitial lung disease (ILD) multidisciplinary team meeting: a novel model to address palliative care needs. *BMJ Open Respir Res* 2018; 5: e000360.
8 Ley B, Ryerson CJ, Vittinghoff E, *et al.* A multidimensional index and staging system for idiopathic pulmonary fibrosis. *Ann Intern Med* 2012; 156: 684–695.
9 Raghu G, Remy-Jardin M, Myers JL, *et al.* Diagnosis of idiopathic pulmonary fibrosis. An official ATS/ERS/JRS/ALAT clinical practice guideline. *Am J Respir Crit Care Med* 2018; 198: e44–e68.
10 Kreuter M, Ehlers-Tenenbaum S, Palmowski K, *et al.* Impact of comorbidities on mortality in patients with idiopathic pulmonary fibrosis. *PLoS One* 2016; 11: e0151425.
11 Pedraza-Serrano F, Jiménez-García R, López-de-Andrés A, *et al.* Comorbidities and risk of mortality among hospitalized patients with idiopathic pulmonary fibrosis in Spain from 2002 to 2014. *Respir Med* 2018; 138: 137–143.

12 Badagliacca R, Rischard F, Papa S, *et al.* Clinical implications of idiopathic pulmonary arterial hypertension phenotypes defined by cluster analysis. *J Heart Lung Transplant* 2020; 39: 310–320.
13 Adegunsoye A, Oldham JM, Chung JH, *et al.* Phenotypic clusters predict outcomes in a longitudinal interstitial lung disease cohort. *Chest* 2018; 153: 349–360.
14 Ruffini M, Gavaldà R, Limon E. Clustering patients with tensor decomposition. *Proc Mach Learn Res* 2017; 68: 126–146.
15 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. *Am J Respir Crit Care Med* 2015; 192: e3–e19.
16 Tossier C, Dupin C, Plantier L, *et al.* Hiatal hernia on thoracic computed tomography in pulmonary fibrosis. *Eur Respir J* 2016; 48: 833–842.
17 Planas-Cerezales L, Arias-Salgado EG, Buendia-Roldán I, *et al.* Predictive factors and prognostic effect of telomere shortening in pulmonary fibrosis. *Respirology* 2019; 24: 146–153.
18 Collard HR, Ryerson CJ, Corte TJ, *et al.* Acute exacerbation of idiopathic pulmonary fibrosis. An international working group report. *Am J Respir Crit Care Med* 2016; 194: 265–275.
19 Chai K-X, Chen Y-Q, Fan P-L, *et al.* STROBE: the correlation of Cyr61, CTGF, and VEGF with polymyositis/dermatomyositis. *Medicine* 2018; 97: e11775.
20 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. *Am J Respir Crit Care Med* 2011; 183: 788–824.
21 Cottin V, Nunes H, Brillet PY, *et al.* Combined pulmonary fibrosis and emphysema: a distinct underrecognised entity. *Eur Respir J* 2005; 26: 586–593.
22 Sateia MJ. International Classification of Sleep Disorders – third edition: highlights and modifications. *Chest* 2014; 146: 1387–1394.
23 Lamas DJ, Kawut SM, Bagiella E, *et al.* Delayed access and survival in idiopathic pulmonary fibrosis: a cohort study. *Am J Respir Crit Care Med* 2011; 184: 842–847.
24 Hoyer N, Prior TS, Bendstrup E, *et al.* Risk factors for diagnostic delay in idiopathic pulmonary fibrosis. *Respir Res* 2019; 20: 103.
25 Molina-Molina Maria, Wijsenbeek Marlies. Comprehensive care in pulmonary fibrosis. *BRN Rev* 2019; 5: 35–47.
26 Cottin V. The impact of emphysema in pulmonary fibrosis. *Eur Respir Rev* 2013; 22: 153–157.
27 Mejía M, Carrillo G, Rojas-Serrano J, *et al.* Idiopathic pulmonary fibrosis and emphysema: decreased survival associated with severe pulmonary arterial hypertension. *Chest* 2009; 136: 10–15.
28 Noth I, Zangan SM, Soares RV, *et al.* Prevalence of hiatal hernia by blinded multidetector CT in patients with idiopathic pulmonary fibrosis. *Eur Respir J* 2012; 39: 344–351.
29 Jouneau S, Kerjouan M, Rousseau C, *et al.* What are the best indicators to assess malnutrition in idiopathic pulmonary fibrosis patients? A cross-sectional study in a referral center. *Nutrition* 2019; 62: 115–121.
30 Molina-Molina M, Aburto M, Acosta O, *et al.* Importance of early diagnosis and treatment in idiopathic pulmonary fibrosis. *Expert Rev Respir Med* 2018; 12: 537–539.
31 Csige I, Ujvárosy D, Szabó Z, *et al.* The impact of obesity on the cardiovascular system. *J Diabetes Res* 2018; 2018: 3407306.
32 Poirier P, Giles TD, Bray GA, *et al.* Obesity and cardiovascular disease: pathophysiology, evaluation, and effect of weight loss: an update of the 1997 American Heart Association Scientific Statement on Obesity and Heart Disease from the Obesity Committee of the Council on Nutrition, Physical Activity, and Metabolism. *Circulation* 2006; 113: 898–918.
33 Collins P, Webb CM, de Villiers TJ, *et al.* Cardiovascular risk assessment in women – an update. *Climacteric* 2016; 19: 329–336.
34 Hudgel DW, Patel SR, Ahasic AM, *et al.* The role of weight management in the treatment of adult obstructive sleep apnea. An official American Thoracic Society clinical practice guideline. *Am J Respir Crit Care Med* 2018; 198: e70–e87.
35 Gasa M, Salord N, Fortuna AM, *et al.* Obstructive sleep apnoea and metabolic impairment in severe obesity. *Eur Respir J* 2011; 38: 1089–1097.
36 Bosi M, Milioli G, Fanfulla F, *et al.* OSA and prolonged oxygen desaturation during sleep are strong predictors of poor outcome in IPF. *Lung* 2017; 195: 643–651.
37 Lancaster LH, Mason WR, Parnell JA, *et al.* Obstructive sleep apnea is common in idiopathic pulmonary fibrosis. *Chest* 2009; 136: 772–778.
38 Pihtili A, Bingol Z, Kiyan E, *et al.* Obstructive sleep apnea is common in patients with interstitial lung disease. *Sleep Breath* 2013; 17: 1281–1288.
39 El-Serag HB, Graham DY, Satia JA, *et al.* Obesity is an independent risk factor for GERD symptoms and erosive esophagitis. *Am J Gastroenterol* 2005; 100: 1243–1250.
40 Jacobson B, Somers S, Fuchs C, *et al.* Body-mass index and gastroesophageal reflux symptoms in women. *N Engl J Med* 2006; 354: 2340–2348.
41 Wang Z, Bonella F, Li W, *et al.* Gastroesophageal reflux disease in idiopathic pulmonary fibrosis: uncertainties and controversies. *Respiration* 2018; 96: 571–587.