Original Article

Cluster phenotypes in a non-idiopathic pulmonary fibrosis fibrotic interstitial lung diseases cohort in Singapore

Michelle Li Wei Kam[1], Pei Yee Tiew[1,2], Hui Zhong Chai[1], Su Ying Low[1]

[1]Department of Respiratory and Critical Care Medicine, Singapore General Hospital, Singapore; [2]Lee Kong Chian School of Medicine, Nanyang Technological University, Singapore

Contributions: (I) Conception and design: MLW Kam, SY Low; (II) Administrative support: MLW Kam; (III) Provision of study materials or patients: MLW Kam, SY Low; (IV) Collection and assembly of data: MLW Kam, HZ Chai, SY Low; (V) Data analysis and interpretation: MLW Kam, PY Tiew, SY Low; (VI) Manuscript writing: All authors; (VII) Final approval of manuscript: All authors.

Correspondence to: Michelle Li Wei Kam. Department of Respiratory and Critical Care Medicine, Singapore General Hospital, 20 College Road, Academia, 169856 Singapore, Singapore. Email: michelle.kam.l.w@singhealth.com.sg.

Background: Non-idiopathic pulmonary fibrosis fibrosing interstitial lung diseases (F-ILDs) may demonstrate a progressive disease trajectory similar to idiopathic pulmonary fibrosis (IPF). We aimed to identify novel F-ILD phenotypes in a multi-ethnic South-East Asian population.

Methods: F-ILD subjects (n=201) were analysed using unsupervised hierarchical cluster analysis and their outcomes compared against IPF (n=86).

Results: Four clusters were identified. Cluster 1 (n=53, 26.4%) comprised older Chinese males with high body mass index (BMI) and comorbidity burden, higher baseline forced vital capacity (FVC) percentage predicted and lower diffusing capacity of the lung for carbon monoxide (DLCO) percentage predicted. They had similar mortality to IPF. Cluster 2 (n=67, 33.3%) had younger female non-smokers with low comorbidity burden, groundglass changes on high-resolution chest computed tomography (HRCT) and a positive anti-nuclear antibody (ANA) titre ≥1:160. They had lower baseline FVC and higher DLCO, low mortality and slower lung function decline. Cluster 3 (n=42, 20.9%) consisted male smokers with low comorbidity burden, emphysema on HRCT and high baseline lung function. They had low mortality and slow lung function decline. Cluster 4 (n=39, 19.4%) was the highest risk and comprised of mainly Indians with high BMI. They had the highest proportion of ischemic heart disease (IHD) and previous pulmonary tuberculosis. Subjects had the lowest baseline lung function, highest mortality, and fastest lung function decline. Survival differences across clusters remained significant following adjustment for treatment.

Conclusions: We identified four distinct F-ILD clinical phenotypes with varying disease trajectories. This demonstrates heterogeneity in F-ILD and the need for complementary approaches for classification and prognostication beyond ATS/ERS guideline diagnosis.

Keywords: South-East Asia; interstitial lung disease; mortality; prognosis

Submitted Jan 10, 2022. Accepted for publication Apr 22, 2022.
doi: 10.21037/jtd-22-40
View this article at: https://dx.doi.org/10.21037/jtd-22-40

Introduction

Interstitial lung diseases (ILD) may demonstrate fibrosis and progressive deterioration, of which, idiopathic pulmonary fibrosis (IPF) is the most aggressive (1,2). Other ILD types may demonstrate similar behaviour, although disease patterns are more heterogenous (3-8). These are termed, "progressive fibrosing interstitial lung disease (PF-ILD)", however a consensus definition has yet to be established and variable trajectories within each diagnosis may limit clinical utility.

Cluster analysis is a statistical method of classifying

individuals into groups based on characteristic differences (9). It has been used in Western populations to describe ILD phenotypes (10-12). Asia is a culturally and ethnically heterogenous population but there is lack of Asian data on ILD, particularly in South-East Asia (13-16). This study aims to use cluster analysis to describe clinical phenotypes in a South-East Asian population of non-IPF fibrotic ILD (F-ILD) patients. We present the following article in accordance with the STROBE reporting checklist (available at https://jtd.amegroups.com/article/view/10.21037/jtd-22-40/rc).

Methods

Study subjects

Patients with ILD were recruited at diagnosis between 5 April 2012 to 4 April 2020 from the outpatient ILD clinic at Singapore General Hospital, a university-affiliated tertiary referral hospital (Appendix 1). Patients were followed up until death, lung transplantation, or censor date 30 June 2021. ILD diagnosis was based on prevailing ATS/ERS guidelines at time of diagnosis (2,17-19). The study was conducted in accordance with the Declaration of Helsinki (as revised in 2013) and approved by the Singhealth centralised institutional review board (CIRB Reference No. 2018/2474; Protocol No. 2012/245/C). Written informed consent was obtained from all participants.

Study design and methods

At recruitment, patients' baseline clinical data were assessed (Appendix 1). Each patient's Gender-Age-Physiology (GAP) score was calculated and assigned their respective GAP stage (20). The GAP Index was chosen over the ILD-GAP Index as some patients had ILD diagnoses which did not conform to the ILD categories in the ILD-GAP index (20,21). Primary outcome was all-cause mortality with lung transplantation as competing risk. Secondary outcomes were respiratory-related mortality and longitudinal lung function.

Statistical analysis

Both numerical and categorical data were selected for cluster analysis based on clinical relevance and previous literature (Supplementary Material Appendix 1 Methods). Subjects who were unable to perform diffusion capacity of the lung for carbon monoxide (DLCO) (n=72) were assigned a value of 0 for unsupervised hierarchical clustering and handled as missing data for other analysis of diffusion capacity. All other data fields used in cluster analysis were complete for all subjects. Statistical analysis was performed using R 3.6.1 (R Foundation for Statistical Computing, Vienna, Austria).

Non-metric multidimensional scaling (NMDS) and hierarchical cluster analysis was performed using "cluster" R package. A Gower dissimilarity matrix was calculated using "daisy" function and "sammon" function was applied on this matrix using a "k"-value of 6, which was determined by a scree plot (Figure S1). All subjects were embedded into a Euclidean space of k =6. Ward's minimum-variance unsupervised hierarchical clustering was applied on this transformed dataset using an agglomerative approach with 'hclust' function. "Nbclust" R package was used to determine the optimal number of derivation clusters, which was 4 (Figure S2).

Differences between groups were analyzed using chi-square tests or Fischer's exact tests for categorical data, Kruskal-Wallis tests for nonparametric continuous data and ANOVA for parametric continuous data. Survival analysis was performed using "survival" and "survminer" R-packages. All-cause mortality was assessed and compared using Kaplan-Meier curves, log-rank test and Cox proportional hazards regression model by cluster. Respiratory-related mortality was compared with non-respiratory causes as competing risk using "cmprsk" R-package.

Linear mixed effects modelling using "lme4" R-package, was used to describe the temporal relationship between clusters and lung function (10,22). The model was built using "lmer" function, using time and cluster as fixed effects and the random effects modelled uncorrelated random intercepts and slopes for the effect of time on each subject. Time points were set at 0, 6, 12, 18, 24, 36, 48, 60, 72, 84, 96 and 108 months. Forced vital capacity (FVC) measurements were attributed to those time points using a ±6-month window, using the nearest measurement to the specific time point for analysis (3). Lung function trends over time between clusters was compared using "afex" R-package.

Results

Subjects

There were 305 ILD patients recruited, of which 287

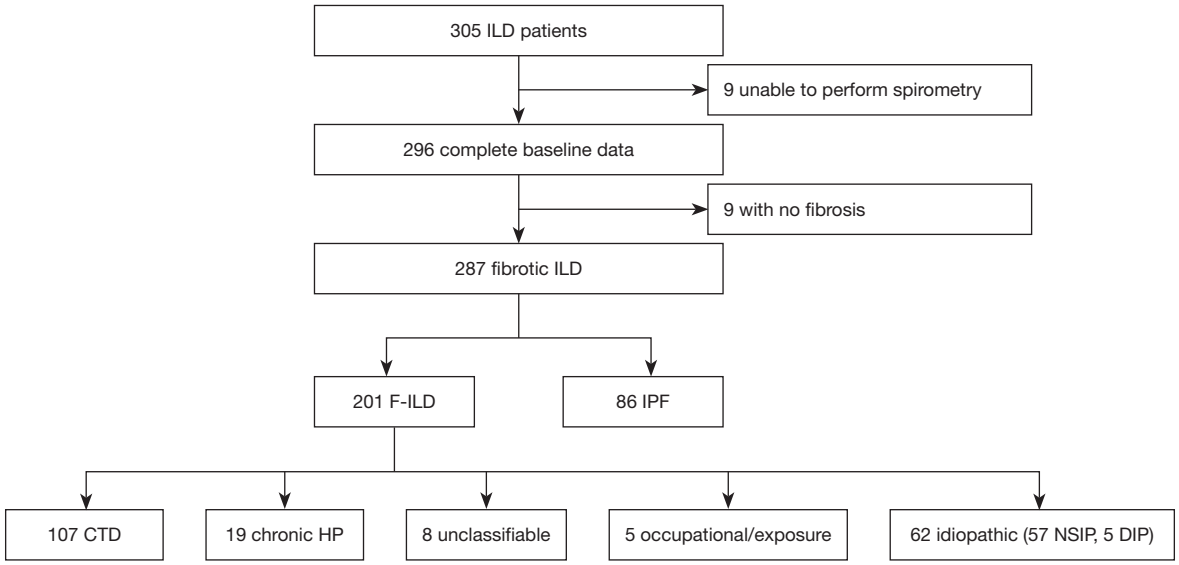


Figure 1 Diagram illustrating selection of patients diagnosed with chronic ILD between 2012 to 2020 (n=305) who were eligible to recruited for cluster analysis (n=287). F-ILD (n=201) subjects were identified and compared against IPF subjects (n=86). ILD, interstitial lung disease; F-ILD, non-idiopathic pulmonary fibrosis fibrotic interstitial lung disease; IPF, idiopathic pulmonary fibrosis; CTD, connective tissue disease; HP, hypersensitivity pneumonitis; NSIP, non-specific interstitial pneumonia; DIP, desquamative interstitial pneumonia.

had fibrotic radiological changes and complete baseline data (*Figure 1*). Of the 287 subjects, 43 (15.0%) had histopathological diagnosis from surgical lung biopsy or cryobiopsy and 90 (31.4%) had bronchoalveolar lavage (BAL) cell counts. There were 86 subjects with IPF and 201 with F-ILD. Amongst non-IPF diagnosis, the most common was connective tissue disease-related ILD (CTD-ILD) (n=107, 53.2%), followed by idiopathic non-specific interstitial pneumonia (NSIP) (n=57, 28.4%), chronic hypersensitivity pneumonitis (HP) (n=19, 9.45%), unclassifiable ILD (n=8, 3.98%), desquamative interstitial pneumonia (DIP) (n=5, 2.49%) and occupational-related ILD (n=5, 2.49%) (*Figure 1*). The characteristics of IPF and F-ILD subjects are summarized in *Table 1*.

Baseline characteristics of clusters

Unsupervised hierarchical clustering of F-ILD subjects identified four clusters (*Figure 2*). The phenotypes are summarized in *Table 2*. All four clusters had subjects diagnosed with different types of ILD and radiological patterns (Table S1). Cluster 1 subjects (n=53, 26.4%) were the oldest and predominantly Chinese. They had higher body mass index (BMI), greater proportion of subjects with ischemic heart disease (IHD) and high comorbidity burden. Subjects had higher baseline FVC percentage predicted and DLCO. Those who underwent BAL had a higher proportion of neutrophilic cell counts (Table S2).

Cluster 2 subjects (n=67, 33.3%) were the youngest and mainly non-smoking females. The predominant diagnosis was CTD-ILD (Table S1). They had low comorbidity burden but higher proportion of thyroid disease. A greater proportion had groundglass changes on chest high-resolution computed tomography scan (HRCT) and antinuclear antibody (ANA) titre ≥1:160. Subjects had lower baseline FVC percentage predicted and higher DLCO percentage predicted. They had the highest proportion of immunosuppression use (*Table 3*).

Cluster 3 (n=42, 20.9%) comprised mainly males with heavy smoking history. They had low comorbidity burden, the highest proportion of subjects with emphysema on chest HRCT and the highest baseline lung function. They had the lowest immunosuppression use and highest antifibrotic use (*Table 3*); those who underwent BAL had the highest proportion of subjects with macrophagic or neutrophilic cell counts (Table S2).

Subjects in Cluster 4 (n=39, 19.4%) were mostly Indian. They had the highest BMI and highest proportion of IHD

Table 1 Characteristics of IPF and non-IPF fibrotic ILD patients

Variable	IPF (n=86)	Non-IPF (n=201)	P value
Age (years), mean ± SD	70.3±8.75	65.0±12.3	<0.001
Male, n (%)	80 (93.0)	123 (61.2)	<0.001
Ethnicity, n (%)			0.491
Chinese	63 (73.3)	129 (64.2)	0.174
Malay	7 (8.14)	21 (10.4)	0.699
Indian	15 (17.4)	46 (22.9)	0.382
Others	1 (1.16)	5 (2.59)	0.789
Smoker/ex-smoker, n (%)	57 (66.3)	55 (27.4)	<0.001
No. of pack years (IQR)	40.0 (20.0, 50.0)	25.0 (10.0, 40.0)	<0.001
Weight loss at presentation, n (%)	29 (33.7)	79 (39.3)	0.447
BMI (kg/m^2), mean ± SD	24.1±4.29	25.8±7.29	0.053
Family history of ILD, n (%)	2 (2.33)	5 (2.49)	1.000
Comorbid burden, n (%)			
Low (0–1)	26 (30.2)	75 (37.3)	0.310
Moderate (2–3)	39 (45.3)	88 (43.8)	0.908
High (≥4)	21 (24.4)	38 (18.9)	0.369
Diabetes mellitus, n (%)	35 (40.7)	50 (24.9)	0.011
Hypertension, n (%)	46 (53.5)	97 (48.3)	0.495
Hyperlipidemia, n (%)	58 (67.4)	99 (49.3)	0.007
Ischemic heart disease, n (%)	30 (34.9)	32 (15.9)	<0.001
Thyroid disease, n (%)	4 (4.65)	23 (11.4)	0.113
GERD, gastritis, peptic ulcer disease, n (%)	9 (10.5)	24 (11.9)	0.875
Cancer, n (%)	5 (5.81)	21 (10.4)	0.304
Previous history of pulmonary tuberculosis, n (%)	9 (10.5)	7 (3.48)	0.037
Genetic syndrome, n (%)	2(2.33)	0 (0.0)	0.163
Pulmonary hypertension on 2D Echo, n (%)	29 (33.7)	74 (36.8)	0.714
Radiology, n (%)			
Emphysema	15 (17.4)	11 (5.47)	0.003
UIP pattern	83 (96.5)	15 (7.46)	<0.001
FVC mean value (L), mean ± SD	2.25±0.59	1.73±0.63	<0.001
FVC percentage of predicted value (%), mean ± SD	69.1±17.8	62.1±17.4	0.002
DLCO* mean value (mmol/min/kPa), mean ± SD	4.36±2.16	4.78±2.87	0.278
DLCO* percentage of predicted value (%), mean ± SD	52.3±21.3	56.8±17.1	0.093
TLC# mean value (L), mean ± SD	3.84±0.68	3.28±0.86	<0.001
TLC# percentage of predicted value (%), mean ± SD	73.0±13.6	75.6±17.7	0.293
Immunosuppression, n (%)	8 (9.30)	154 (76.6)	<0.001
Antifibrotics, n (%)	8 (9.30)	10 (4.98)	0.016

*, 14 (16.3%) IPF patients and 58 (28.9%) non-IPF fibrotic ILD patients were unable to perform the test; #, 23 (26.7%) IPF patients and 47 (23.4%) non-IPF fibrotic ILD patients were unable to perform the test. IPF, idiopathic pulmonary fibrosis; ILD, interstitial lung disease; IQR, interquartile range; BMI, body mass index; GERD, gastroesophageal reflux disease; UIP, usual interstitial pneumonia; FVC, forced vital capacity; DLCO, diffusing capacity of the lung for carbon monoxide; TLC, total lung capacity.

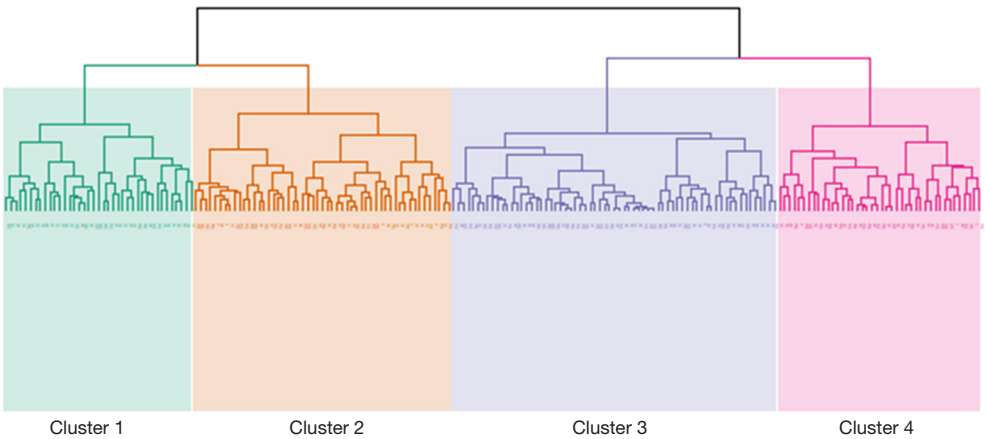


Figure 2 Dendrogram of the four clusters identified. Cluster 1 (n=53, 26.4%): older Chinese males with high BMI and comorbidity burden, and higher baseline FVC percentage predicted. Cluster 2 (n=67, 33.3%): younger female non-smokers with low comorbidity burden, groundglass changes on HRCT, positive ANA titre ≥1:160 and lower baseline FVC percentage predicted. Cluster 3 (n=42, 20.9%): male smokers with low comorbidity burden, emphysema on HRCT and high baseline lung function. Cluster 4 (n=39, 19.4%): non-Chinese ethnicity, mainly Indian, with high BMI, low baseline lung function and high proportion of IHD and previous pulmonary TB. BMI, body mass index; FVC, forced vital capacity; HRCT, high-resolution chest computed tomography; ANA, anti-nuclear antibody; IHD, ischemic heart disease.

and previous pulmonary tuberculosis (TB). They had the lowest baseline lung function and highest prednisolone use (*Table 3*). The P values for each of the cluster's characteristics compared against IPF are summarized in Table S3.

Survival

Kaplan-Meier curves demonstrated survival differences for all-cause mortality. Cluster 4 had the highest mortality across all time points and the shortest median survival time at 30 months (*Figure 3A*, Table S4) (log-rank test, P<0.001). When adjusted for treatment with immunosuppression and antifibrotics, survival differences between clusters remained significant (*Figure 3B*) (log-rank test, P<0.001). Cluster 4 had higher mortality risk than IPF, hazards ratio (HR) for mortality: 1.974 (HR, 1.974; 95% CI: 1.202–3.240; P=0.007). Clusters 2 and 3 had lower mortality risk than IPF, with HR for mortality: 0.105 (HR, 0.105; 95% CI: 0.037–0.295; P<0.001) and 0.413 (HR, 0.413; 95% CI: 0.198–0.858; P=0.018), respectively (Table S5).

Respiratory-related mortality differences were similar to that for all-cause mortality (*Figure 3C*). When survival was analysed by diagnosis, CTD-ILD had lower mortality risk than IPF with HR for mortality: 0.298 (HR, 0.298; 95% CI: 0.167–0.531; P<0.001) (Table S5 and Figure S3). A higher GAP stage correlated well with mortality (Table S5 and Figure S4A) (log-rank test, P<0.001) and within each GAP stage there were significant differences in survival by cluster (Figure S4B-S4D).

Lung function trajectory

There were significant differences in lung function trajectories between clusters (*Figure 4A*). Cluster 4 had the lowest baseline FVC percentage predicted value and greatest FVC decline from baseline [rate of FVC decline from baseline: 55.4 (±3.88) mL/year] (*Figure 4B*, Table S4). This was followed by Cluster 1 [47.0 (±9.64) mL/year], whilst Cluster 3 had the slowest rate of FVC decline from baseline at 4.22 (±2.88) mL/year (*Figure 4B*, Table S4). Post-hoc analysis showed that there was no significant difference in the rate of FVC decline in subjects who received anti-fibrotic therapy and those who did not, for both Clusters 3 and 4 (Table S6). When analysed for a composite outcome of decline in FVC ≥5% of the predicted value from baseline or death at 12 months, there was no significant difference between clusters (Figure S5A) (log-rank test, P=0.09). There was also no significant difference between clusters when analysed for a composite outcome of decline in FVC ≥10% of the predicted value from baseline or death at 12 months (Figure S5B) (log-rank test, P=0.1).

Table 2 Characteristics of clusters

Variable	Cluster 1 (n=53)	Cluster 2 (n=67)	Cluster 3 (n=42)	Cluster 4 (n=39)	P value
Age (years), mean ± SD	71.5±7.91	58.1±13.1	63.7±11.2	69.6±10.5	<0.001
Male, n (%)	18 (34.0)	2 (2.99)	38 (90.5)	20 (51.3)	<0.001
Ethnicity, n (%)					
Chinese	51 (96.2)	45 (67.2)	33 (78.6)	0 (0.0)	<0.001
Malay	0 (0.0)	10 (14.9)	5 (11.9)	6 (15.4)	0.047
Indian	0 (0.0)	12 (17.9)	4 (9.52)	30 (76.9)	<0.001
Others	2 (3.77)	0 (0.0)	0 (0.0)	3 (7.69)	0.052
Smoker/ex-smoker, n (%)	14 (26.4)	3 (4.48)	30 (71.4)	8 (20.5)	<0.001
No. of pack years (IQR)	20 (7.5, 55)	10 (10, 15)	32.5 (13.5, 40)	20 (12.5, 60)	<0.001
BMI (kg/m^2), mean ± SD	25.5±7.14	24.7±5.33	25.0±3.75	28.7±11.6	0.008
Family history of ILD, n (%)	0 (0.0)	2 (2.99)	1 (2.38)	2 (5.13)	0.627
Comorbid burden, n (%)					
Low (0–1)	3 (5.66)	39 (58.2)	26 (61.9)	7 (17.9)	<0.001
Moderate (2–3)	33 (62.3)	23 (34.3)	11 (26.1)	21 (53.8)	0.002
High (4–6)	17 (32.1)	5 (7.46)	5 (11.9)	11 (28.2)	0.004
Diabetes mellitus, n (%)	23 (43.4)	6 (8.96)	5 (11.9)	16 (41.0)	<0.001
Hypertension, n (%)	42 (79.2)	12 (17.9)	14 (33.3)	29 (74.4)	<0.001
Hyperlipidemia, n (%)	44 (83.0)	17 (25.4)	12 (28.6)	26 (66.7)	<0.001
Ischemic heart disease, n (%)	14 (26.4)	1 (1.49)	5 (11.9)	12 (30.8)	<0.001
Thyroid disease, n (%)	5 (9.43)	11 (16.4)	0 (0.0)	7 (17.9)	0.008
Asthma, n (%)	0 (0.0)	2 (2.99)	0 (0.0)	3 (7.69)	0.070
Cancer, n (%)	5 (9.43)	5 (7.46)	7 (16.7)	4 (10.3)	0.363
GERD, gastritis, peptic ulcer disease, n (%)	11 (20.8)	5 (7.46)	5 (11.9)	3 (7.69)	0.186
Previous history of pulmonary tuberculosis, n (%)	0 (0.0)	1 (1.49)	1 (2.38)	5 (12.8)	0.021
Pulmonary hypertension on 2D echo, n (%)	19 (35.8)	23 (34.3)	17 (40.5)	15 (38.5)	0.946
Groundglass, n (%)	42 (79.2)	57 (85.1)	31 (73.8)	33 (84.6)	<0.001
Emphysema, n (%)	1 (1.89)	0 (0.0)	9 (21.4)	1 (2.56)	<0.001
UIP pattern, n (%)	4 (7.55)	4 (5.97)	4 (9.52)	4 (10.3)	1.000
Positive ANA ≥1:160	12 (22.6)	49 (73.1)	15 (35.7)	10 (25.6)	<0.001
FVC percentage of predicted value (%), mean ± SD	65.7±16.7	61.0±16.6	69.0±16.3	51.6±16.4	<0.001
DLCO percentage of predicted value (%)*, mean ± SD	54.3±18.6	59.2±15.6	62.4±16.1	46.2±14.5	<0.001

*, 15 patients from Cluster 1, 25 patients from Cluster 2, 1 patient from Cluster 3 and 17 patients from Cluster 4 were unable to perform the test. IQR, interquartile range; BMI, body mass index; ILD, interstitial lung disease; GERD, gastroesophageal reflux disease; UIP, usual interstitial pneumonia; ANA, antinuclear antibody; FVC, forced vital capacity; DLCO, diffusing capacity of the lung for carbon monoxide.

Table 3 Treatment received by cluster

Treatment, n (%)	Cluster 1 (n=53)	Cluster 2 (n=67)	Cluster 3 (n=42)	Cluster 4 (n=39)	P value
Immunosuppression	43 (81.1)	57 (85.1)	22 (52.4)	32 (82.1)	<0.001
Prednisolone	43 (81.1)	54 (80.6)	27 (64.3)	33 (84.6)	0.134
Azathioprine	9 (17.0)	10 (14.9)	5 (11.9)	7 (17.9)	0.899
Mycophenolate mofetil	14 (26.4)	28 (41.8)	7 (16.7)	8 (20.5)	0.044
Cyclophosphamide	2 (3.77)	8 (11.9)	2 (4.76)	1 (2.56)	0.218
Rituximab	2 (3.77)	3 (4.48)	2 (4.76)	0 (0.0)	0.787
Antifibrotics	0 (0.0)	1 (1.49)	6 (14.3)	3 (7.69)	0.034

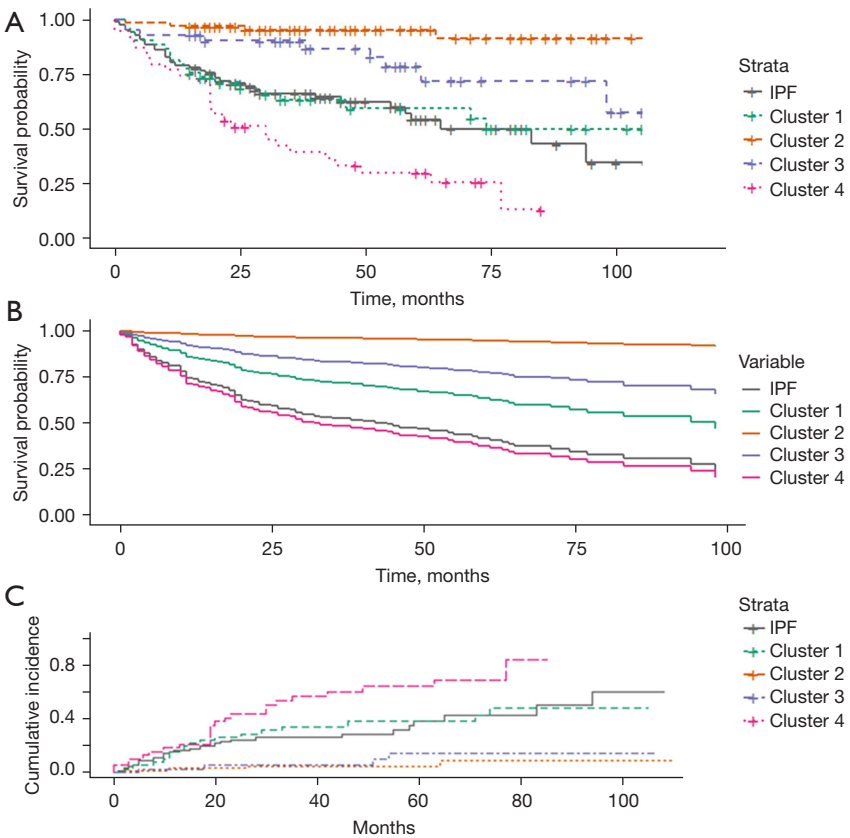
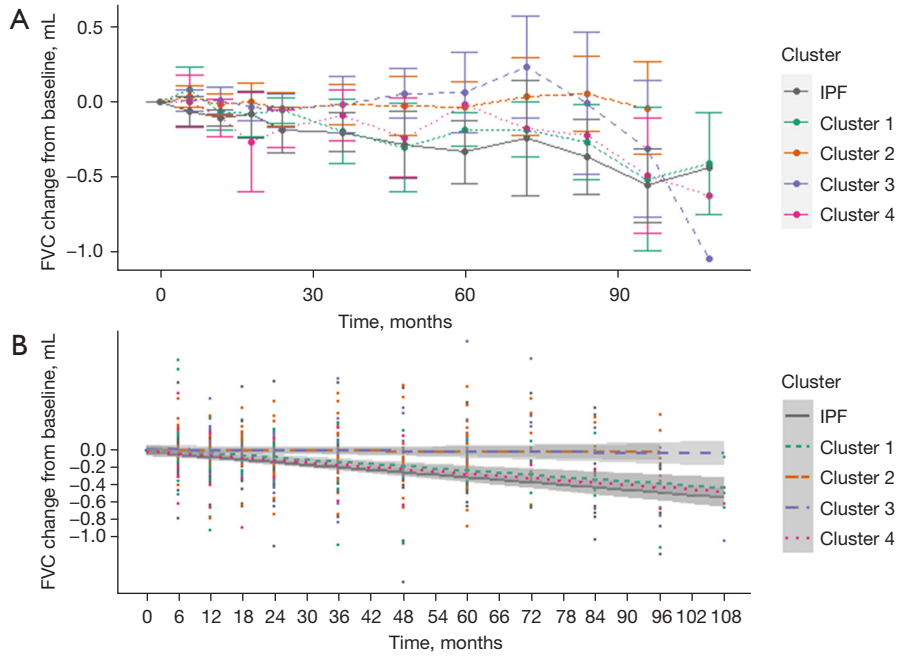


Figure 3 Amongst the clusters, Cluster 4 has the lowest survival, Cluster 1 has similar survival to IPF, Clusters 2 and 3 have high survival as illustrated in (A) Kaplan-Meier curve comparing survival differences between the 4 clusters against IPF for all-cause mortality. Survival differences described between the clusters remained significant after adjustment for treatment with immunosuppression and antifibrotics as shown in (B) Cox-regression survival curve between the clusters. (C) Clusters 1 and 4 are high risk for respiratory-related mortality as shown in (C) cumulative incidence curves for respiratory-related mortality across the four clusters against IPF. Clusters are coded according to the legend. IPF, idiopathic pulmonary fibrosis.



Time	0	6	12	18	24	36	48	60	72	84	96	108
Overall	287	173	138	111	97	81	60	43	32	24	21	6
IPF	86	49	39	26	23	22	17	10	9	9	8	1
Cluster 1	53	31	24	22	17	11	8	7	6	5	4	3
Cluster 2	67	34	32	28	27	24	19	14	9	5	5	0
Cluster 3	42	36	29	21	21	16	11	11	7	4	3	1
Cluster 4	39	23	14	14	9	8	5	1	1	1	1	1

Figure 4 Clusters 1 and 4 have a faster rate of decline compared to Clusters 2 and 3. (A) Graph of raw FVC mean change from baseline (mL) and 95% CI at calculated time points by cluster compared against IPF. (B) Graph of linear mixed model estimation and 95% CI for FVC (mL) at calculated time points. Clusters are coded according to the legend. FVC, forced vital capacity; IPF, idiopathic pulmonary fibrosis.

Discussion

This is a real-world, prospective cohort study involving South-East Asian patients with F-ILD. Unsupervised hierarchal clustering classified patients into four distinct phenotypes with different outcomes and trajectories. We identified two high-risk clusters, namely Clusters 1 and 4. Cluster 1 comprised older Chinese males with high BMI, comorbidity burden, higher baseline FVC but lower DLCO, and similar mortality to IPF. Cluster 4 had the highest mortality and comprised mainly of Indians with high BMI, low baseline lung function and a higher proportion of IHD and previous pulmonary TB. Clusters 2 and 3 both had low mortality. Cluster 2 subjects were mainly younger female non-smokers with low comorbidity burden, groundglass changes on HRCT, positive ANA titre ≥1:160, lower baseline FVC but higher DLCO. Cluster 3 subjects were male smokers with low comorbidity burden, emphysema on HRCT and high baseline lung function.

Some of our findings are similar to published literature. Increasing age, male sex and poorer lung function are well-described baseline predictors of increased mortality and were demonstrated in our clusters (20,21). Previous cluster analysis also showed that younger females with positive ANA had improved survival, similar to that in Cluster 2 (10). Unclassifiable ILD and chronic HP have been identified as ILD diagnoses associated with poorer outcomes in PF-ILD cohorts (3,15,23). However, our study and other cluster analysis demonstrate that there is heterogeneity of disease behaviour within a diagnosis and overlap in disease trajectories across different diagnoses, highlighting the importance of classifying different ILDs by disease behaviour, as previously proposed in ATS/ERS guidelines

(10-12,18).

Some of our findings are unique to the South-East Asian region. We found that the two high-risk clusters, Clusters 1 and 4, comprised mainly Chinese and Indian patients respectively, with Cluster 4 demonstrating a more aggressive course. In 2019, Singapore's population distribution comprised 74.4% Chinese, 13.4% Malays, 9.0% Indians and 3.2% other ethnicities (24). In contrast, our F-ILD cohort had a lower proportion of Chinese (64.2%) and Malays (10.4%), and a higher proportion of Indians (22.9%). Differences between Asian and Caucasian ILD have been described, such as higher rates of exacerbations in East-Asians and a high proportion of chronic HP in the Indian registry (14,25). The differences in disease behaviour and prevalence by geography and ethnicity requires further research to identify potential factors such as genetic polymorphisms, environmental exposure and lifestyle practices which may account for this.

TB is endemic in Singapore and TB prevalence is high in South-East Asia (26). We found that the high-risk Cluster 4 had a high proportion of subjects with prior pulmonary TB, which is unique and of significance to TB-endemic regions. Prior pulmonary TB has been associated with poorer disease outcomes in chronic obstructive pulmonary disease (COPD) (27). Although TB incidence in ILD is 4.5 times higher than that of the general population in Israel and 15.5% of the Indian ILD registry had prior pulmonary TB, the effects of prior pulmonary TB on ILD outcomes have not been delineated (14,28). Furthermore, ethnic differences have been found to result in variations in inflammatory profiles and clinical phenotypes in TB (29,30). Given that Cluster 4 was predominantly Indian, implications of prior infection and ethnic differences on ILD outcomes require further study.

Our study also highlights the importance of identifying and managing comorbidities in ILD. We found that Clusters 1 and 4, which had high comorbidity burden and high proportion of IHD respectively, had high mortality. Increasing number of comorbidities, particularly cardiovascular risk factors and untreated cardiac disease, increases ILD mortality (11,31,32). We found that all the clusters had a mean BMI above the Asian cut-off of 23.0 kg/m^2 (33). This could be related to corticosteroid therapy which 78.1% received. The effect of BMI on ILD outcomes requires further study. Low BMI and decreasing BMI trend correlates with increased mortality, however high BMI increases cardiovascular risk and is associated with increased exacerbations (34,35). Our study highlights the importance of defining the ILD comorbidome for targeted screening and early treatment of comorbidities (11,12,31,36).

Our findings demonstrate the challenges in identifying high-risk F-ILD at diagnosis (23). Current ILD risk scores emphasize lung function and diagnosis, with most developed for IPF (20,21). PF-ILD is characterised by lung function decline and increased mortality risk if untreated (3,5,8,23). Currently, there is no consensus on the PF-ILD definition and different criteria have been used in trials (5,7,8). Although we did not examine for PF-ILD based on existing trial criteria, we found that 45.8% of our F-ILD cohort were high-risk with mortality and disease progression similar to, or more aggressive than IPF. This is higher than the reported PF-ILD incidence of 13–40% (3,4,6). The higher proportion of F-ILD in our cohort with aggressive disease behaviour requires further study.

Some of the limitations are the small size and lack of validation cohort. Furthermore, only subjects with complete baseline data were included for cluster analysis, thus patients such as those unable to perform spirometry due to more advanced disease were excluded. Patients with sarcoidosis and rarer ILD types like pleuroparenchymal fibroelastosis were also not studied; thus, the applicability and generalisability to other ILD types requires further study. The extent of fibrosis radiologically was not quantified and hence identifying subjects that would benefit from antifibrotic therapy was also limited.

Obtaining supporting histopathological evidence to establish an ILD diagnosis is an important component of multidisciplinary diagnosis. However, surgical lung biopsy for ILD has an in-hospital mortality of 2%, which doubles by 90-day, and 19.1% will experience at least one surgical-related complication (37-39). Globally, only 10% of ILD patients undergo surgical lung biopsy due to advanced disease or lack of access to services (40). The utility of histopathology in establishing a multidisciplinary ILD diagnosis thus varies around the world, and is often dependent on local practices and resource availability which may restrict the applicability of ATS/ERS guidelines (13,40). Our cohort's biopsy rate of 15.0%, is reflective of local clinical practices and patient preferences. Furthermore, as our unit is a tertiary referral centre, some patients are only referred in advanced stages and thus unable to undergo biopsy due to high risks.

Although this was a single centre study, our centre receives ILD referrals across Singapore and thus our cohort is representative of local ILD patients. Patients were

diagnosed strictly according to ATS/ERS guidelines and longitudinal disease trends were characterised. However, referral bias from clinicians at referring centres may have resulted in patients assessed to be too advanced or too early in the disease to require tertiary specialist care, and hence not referred. Some patients may have received treatment prior to consultation at our institution, altering the clinical phenotype and disease trajectory; however, we found that survival differences across clusters remained significant after adjusting for treatment. In addition, due to cost limitations, such as for the use of antifibrotics, some subjects may have declined treatment and hence deteriorate more rapidly. More research is needed to identify factors which may account for differences amongst clusters.

Our study characterises the disease trajectories of South-East Asian F-ILD into four distinct clinical phenotypes. Although further validation is needed in larger multicentre cohorts, this has broader clinical utility in the management of ILD and sheds light on future areas of study with regards to disease behaviour in different South-East Asian ethnicities and comorbidities. There is currently a void in the understanding of ILD in South-East Asia and our findings are helpful in prognostication for such patients and highlight the need for further research in South-East Asia to identify high-risk groups.

Acknowledgments

Funding: This work was supported by Singapore General Hospital (grant No. SRG-NIG-04-2021 to MLW Kam).

Footnote

Reporting Checklist: The authors have completed the STROBE reporting checklist. Available at https://jtd.amegroups.com/article/view/10.21037/jtd-22-40/rc

Data Sharing Statement: Available at https://jtd.amegroups.com/article/view/10.21037/jtd-22-40/dss

Peer Review File: Available at https://jtd.amegroups.com/article/view/10.21037/jtd-22-40/prf

Conflicts of Interest: All authors have completed the ICMJE uniform disclosure form (available at https://jtd.amegroups.com/article/view/10.21037/jtd-22-40/coif). MLWK has received research funding from Singapore General Hospital (SGH). The other authors have no conflicts of interest to declare.

Ethical Statement: The authors are accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved. The study was conducted in accordance with the Declaration of Helsinki (as revised in 2013) and approved by the Singhealth centralised institutional review board (CIRB Reference No. 2018/2474; Protocol No. 2012/245/C). Written informed consent was obtained from all participants.

References

1. Park JH, Kim DS, Park IN, et al. Prognosis of fibrotic interstitial pneumonia: idiopathic versus collagen vascular disease-related subtypes. Am J Respir Crit Care Med 2007;175:705-11.
2. 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:e44-68.
3. Nasser M, Larrieu S, Si-Mohamed S, et al. Progressive fibrosing interstitial lung disease: a clinical cohort (the PROGRESS study). Eur Respir J 2021;57:2002718.
4. Wijsenbeek M, Kreuter M, Olson A, et al. Progressive fibrosing interstitial lung diseases: current practice in diagnosis and management. Curr Med Res Opin 2019;35:2015-24.
5. Flaherty KR, Wells AU, Cottin V, et al. Nintedanib in Progressive Fibrosing Interstitial Lung Diseases. N Engl J Med 2019;381:1718-27.
6. Brown KK, Martinez FJ, Walsh SLF, et al. The natural history of progressive fibrosing interstitial lung diseases. Eur Respir J 2020;55:2000085.
7. Maher TM, Corte TJ, Fischer A, et al. Pirfenidone in patients with unclassifiable progressive fibrosing interstitial lung disease: a double-blind, randomised,

placebo-controlled, phase 2 trial. Lancet Respir Med 2020;8:147-57.
8. Behr J, Prasse A, Kreuter M, et al. Pirfenidone in patients with progressive fibrotic interstitial lung diseases other than idiopathic pulmonary fibrosis (RELIEF): a double-blind, randomised, placebo-controlled, phase 2b trial. Lancet Respir Med 2021;9:476-86.
9. McLachlan GJ. Cluster analysis and related techniques in medical research. Stat Methods Med Res 1992;1:27-48.
10. Adegunsoye A, Oldham JM, Chung JH, et al. Phenotypic Clusters Predict Outcomes in a Longitudinal Interstitial Lung Disease Cohort. Chest 2018;153:349-60.
11. Wong AW, Lee TY, Johannson KA, et al. A cluster-based analysis evaluating the impact of comorbidities in fibrotic interstitial lung disease. Respir Res 2020;21:322.
12. Prior TS, Hoyer N, Hilberg O, et al. Clusters of comorbidities in idiopathic pulmonary fibrosis. Respir Med 2021;185:106490.
13. Rivera-Ortega P, Molina-Molina M. Interstitial Lung Diseases in Developing Countries. Ann Glob Health 2019;85:4.
14. Singh S, Collins BF, Sharma BB, et al. Interstitial Lung Disease in India. Results of a Prospective Registry. Am J Respir Crit Care Med 2017;195:801-13.
15. Kwon BS, Choe J, Chae EJ, et al. Progressive fibrosing interstitial lung disease: prevalence and clinical outcome. Respir Res 2021;22:282.
16. Zhan X, Yan W, Wang Y, et al. Clinical features of anti-synthetase syndrome associated interstitial lung disease: a retrospective cohort in China. BMC Pulm Med 2021;21:57.
17. American Thoracic Society; European Respiratory Society. American Thoracic Society/European Respiratory Society International Multidisciplinary Consensus Classification of the Idiopathic Interstitial Pneumonias. This joint statement of the American Thoracic Society (ATS), and the European Respiratory Society (ERS) was adopted by the ATS board of directors, June 2001 and by the ERS Executive Committee, June 2001. Am J Respir Crit Care Med 2002;165:277-304.
18. Travis WD, Costabel U, Hansell DM, 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;188:733-48.
19. 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.
20. Ley B, Ryerson CJ, Vittinghoff E, et al. A multidimensional index and staging system for idiopathic pulmonary fibrosis. Ann Intern Med 2012;156:684-91.
21. Ryerson CJ, Vittinghoff E, Ley B, et al. Predicting survival across chronic interstitial lung disease: the ILD-GAP model. Chest 2014;145:723-8.
22. Edwards LJ. Modern statistical techniques for the analysis of longitudinal data in biomedical research. Pediatr Pulmonol 2000;30:330-44.
23. Goos T, De Sadeleer LJ, Yserbyt J, et al. Defining and predicting progression in non-IPF interstitial lung disease. Respir Med 2021;189:106626.
24. Singapore Department of Statistics/Ministry of Trade and Industry/Republic of Singapore. Population Trends, 2019. Available Online: https://www.singstat.gov.sg/-/media/files/publications/population/population2019.pdf
25. Natsuizaka M, Chiba H, Kuronuma K, et al. Epidemiologic survey of Japanese patients with idiopathic pulmonary fibrosis and investigation of ethnic differences. Am J Respir Crit Care Med 2014;190:773-9.
26. Global tuberculosis report 2018. Geneva: World Health Organization, 2018.
27. Park HJ, Byun MK, Kim HJ, et al. History of pulmonary tuberculosis affects the severity and clinical outcomes of COPD. Respirology 2018;23:100-6.
28. Shachor Y, Schindler D, Siegal A, et al. Increased incidence of pulmonary tuberculosis in chronic interstitial lung disease. Thorax 1989;44:151-3.
29. Coussens AK, Wilkinson RJ, Nikolayevskyy V, et al. Ethnic variation in inflammatory profile in tuberculosis. PLoS Pathog 2013;9:e1003468.
30. Pareek M, Evans J, Innes J, et al. Ethnicity and mycobacterial lineage as determinants of tuberculosis disease phenotype. Thorax 2013;68:221-9.
31. Schwarzkopf L, Witt S, Waelscher J, et al. Associations between comorbidities, their treatment and survival in patients with interstitial lung diseases - a claims data analysis. Respir Res 2018;19:73.
32. Carter P, Lagan J, Fortune C, et al. Association of Cardiovascular Disease With Respiratory Disease. J Am Coll Cardiol 2019;73:2166-77.
33. WHO Expert Consultation. Appropriate body-mass index for Asian populations and its implications for policy and intervention strategies. Lancet 2004;363:157-63.
34. Alakhras M, Decker PA, Nadrous HF, et al. Body mass index and mortality in patients with idiopathic pulmonary fibrosis. Chest 2007;131:1448-53.

35. Kondoh Y, Taniguchi H, Katsuta T, et al. Risk factors of acute exacerbation of idiopathic pulmonary fibrosis. Sarcoidosis Vasc Diffuse Lung Dis 2010;27:103-10.
36. Yagyu H, Murohashi K, Hara Y, et al. Clinical utility of a composite scoring system including Charlson Comorbidity Index score in patients with interstitial lung disease. J Thorac Dis 2020;12:5774-82.
37. Hutchinson JP, McKeever TM, Fogarty AW, et al. Surgical lung biopsy for the diagnosis of interstitial lung disease in England: 1997-2008. Eur Respir J 2016;48:1453-61.
38. Kreider ME, Hansen-Flaschen J, Ahmad NN, et al. Complications of video-assisted thoracoscopic lung biopsy in patients with interstitial lung disease. Ann Thorac Surg 2007;83:1140-4.
39. Luo Q, Han Q, Chen X, et al. The diagnosis efficacy and safety of video-assisted thoracoscopy surgery (VATS) in undefined interstitial lung diseases: a retrospective study. J Thorac Dis 2013;5:283-8.
40. Richeldi L, Launders N, Martinez F, et al. The characterisation of interstitial lung disease multidisciplinary team meetings: a global study. ERJ Open Res 2019;5:e00209-2018.

Cite this article as: Kam MLW, Tiew PY, Chai HZ, Low SY. Cluster phenotypes in a non-idiopathic pulmonary fibrosis fibrotic interstitial lung diseases cohort in Singapore. J Thorac Dis 2022;14(7):2481-2492. doi: 10.21037/jtd-22-40

Appendix 1

Methods

Inclusion and exclusion criteria

The study inclusion criteria were as follows: patients on follow up with the Singapore General Hospital (SGH) outpatient Interstitial Lung Diseases (ILD) clinic, with a confirmed diagnosis of ILD based on the latest ATS/ERS guidelines at the time of diagnosis and at least 21 years old at the time of recruitment. Unclassifiable ILD was defined as patients without a specific ILD diagnosis following multidisciplinary review of clinical, radiological and pathological data (18,21,41).

Exclusion criteria was any patient with ILD who did not provide informed written consent or did not fulfil all of the inclusion criteria.

Data collection

Clinical data was obtained for all subjects at recruitment and was categorised into the following domains: demographics, exposures, comorbidities, symptoms, serology, radiology and lung function.

Demographics entailed sex, age and ethnicity. Ethnicity was classified according to Chinese, Indian, Malay or "Others" as stated on patients' passport or national identification card which is required for hospital registration. Exposures comprised of smoking history and pack year exposure where relevant, environmental exposures from either occupation, hobbies or daily life as elicited on history taking by the clinician. A family history of interstitial lung disease was defined as a reported history from the patient of at least one first degree relative with interstitial lung disease.

All comorbidities were recorded by the recruiting clinician at the point of study through patient electronic medical records. This was corroborated by reviewing confirmatory tests, physician reports and patient's medication lists. Comorbidities recorded were as follow: hypertension, hyperlipidaemia, diabetes mellitus, stroke, ischemic heart disease, stroke; chronic kidney disease (due to any aetiology), chronic liver disease (liver cirrhosis and steatohepatitis); asthma; connective tissue disease [rheumatoid arthritis, systemic sclerosis, Sjogren's Syndrome, systemic lupus erythematosus (SLE), undifferentiated connective tissue disease, dermatomyositis, polymyositis and inflammatory myositis]; cancer; previous tuberculosis; gastritis, gastroesophageal reflux, gastric or duodenal ulcers (peptic ulcer disease);thyroid disease (hyperthyroidism or hypothyroidism of any aetiology), psychiatric disease (anxiety, depression and schizophrenia).

Recorded symptoms were those reported by patients at the point of recruitment and consisted of dyspnoea, cough and weight loss. Serological studies recorded were rheumatoid factor, antinuclear antibody (ANA), extractable nuclear antigen (ENA) antibody panel which comprised of Ro, La, Sm, Scl-70, Jo-1 and RNP. A result was considered positive in accordance with standard laboratory reporting cut-offs. The median ANA titre amongst subjects who had a positive ANA titre was 160.

Radiological findings as evaluated by a reporting thoracic radiologist and clinician during multi-disciplinary discussion were determined to be fibrotic or non-fibrotic. Subjects who had no evidence of any fibrotic change were excluded from the study. They were then categorised according to UIP pattern, possible UIP pattern and inconsistent for UIP pattern up to 2018 (19), when the following classification was adopted in accordance with ATS/ERS guideline revisions: UIP pattern, probable UIP pattern, indeterminate for UIP pattern and alternative diagnosis (2). Lung function parameters recorded were the absolute and percentage predicted forced vital capacity and diffusion capacity.

Comorbidities definitions

Diabetes mellitus was defined as either fasting blood glucose levels of ≥7.0 mmol/L, blood glucose levels of ≥11.1 mmol/L two-hour post oral glucose tolerance test, random blood glucose level ≥11.1 mmol/L with hyperglycaemia symptoms, or HbA1c ≥6.5% (42).

Hypertension was defined as systolic blood pressure ≥130 mmHg or diastolic blood pressure ≥85 mmHg (43).

Hyperlipidaemia was defined as either LDL Cholesterol ≥3.4 mmol/L or triglyceride ≥1.7 mmol/L (44).

Cardiovascular risk factors were defined as the presence of at least one of the following: history of smoking, diabetes mellitus, hypertension, hyperlipidaemia.

Ischaemic heart disease was defined as the presence of coronary artery disease based on functional testing, radiological imaging and/or coronary angiography with the presence of ischemic symptoms or prior history of acute coronary syndrome (45).

Stroke was defined as a central nervous system infarction or haemorrhage (46).

Chronic kidney disease was defined as an estimated glomerular filtration rate (eGFR) <60 mL/min (47).

Liver cirrhosis and steatohepatitis were defined by the

https://dx.doi.org/10.21037/jtd-22-40

presence of clinical, biochemical and radiological features consistent with the diagnosis, with or without liver biopsy (48,49).

Asthma was defined as clinical features consistent with asthma and the presence of variable expiratory airflow limitation (50).

Connective tissue disease, namely, rheumatoid arthritis, systemic sclerosis, Sjogren's Syndrome, systemic lupus erythematosus (SLE), undifferentiated connective tissue disease, dermatomyositis, polymyositis and inflammatory myositis were defined according to their respective guideline criteria (51-55).

Cancers were diagnosed by radiological imaging and/or histological confirmation.

Previous pulmonary tuberculosis was defined as a prior history of documented tuberculosis with positive sputum analysis for mycobacteria tuberculosis including positive acid-fast bacilli smear, culture or nucleic acid amplification, radiological features (on chest radiography or computed tomography) and/or prior pharmacological treatment for pulmonary tuberculosis (56,57).

Gastroesophageal reflux disease, esophagitis, gastritis and peptic ulcer disease were diagnosed according to guideline criteria with endoscopy for esophagitis, gastritis and peptic ulcer disease (58,59).

Hypothyroidism was defined as a subnormal assessment of serum free T4 with either elevated or normal serum TSH. Hyperthyroidism was defined as a subnormal serum TSH with or without the presence of an elevated or normal free T4 or elevated free T3. Subclinical hyperthyroidism was defined as a normal serum-free T4 estimate and normal total T3 or free T3 estimate, with subnormal serum TSH concentration (60,61).

Anxiety, depression and schizophrenia were defined according to their respective DSM-V criteria (62).

Outcome definitions

All-cause mortality was defined as death due to any cause.

Respiratory-related mortality was defined as primary cause of death due to one of the following: pneumonia, pulmonary embolism, exacerbation of ILD, end-stage/advanced ILD, and respiratory failure.

Exacerbation-related mortality was defined as primary cause of death due to acute exacerbation of ILD. An acute exacerbation was defined as an acute worsening or development of dyspnoea over less than one month duration not due to cardiac failure or fluid overload, in a patient with a known or new diagnosis of ILD, associated with new groundglass opacity changes and/or consolidation on CT thorax, superimposed on a background pattern consistent with a diagnosis of ILD (63).

Survival time was defined as time from date of first consultation at the ILD clinic to death. Survival time was censored on 30 June 2021, or when a subject underwent lung transplantation or was lost to follow-up.

References

41. Ryerson CJ, Urbania TH, Richeldi L, et al. Prevalence and prognosis of unclassifiable interstitial lung disease. Eur Respir J 2013;42:750-7.
42. Ministry of Health Singapore. Diabetes Mellitus: MOH Clinical Practice Guidelines 1/2014. Available online: https://www.moh.gov.sg/content/dam/moh_web/HPP/Doctors/cpg_medical/current/2014/diabetes_mellitus/cpg_Diabetes Mellitus Booklet - Jul 2014.pdf
43. Ministry of Health Singapore. MOH clinical practice guidelines 1/2017: Hypertension. Available online: https://www.moh.gov.sg/content/dam/moh_web/HPP/Doctors/cpg_medical/current/2015/anxiety_disorders/cpg_Anxiety Disorders Apr 2015 - Summary Card.pdf
44. Ministry of Health Singapore. MOH Clinical Practice Guidelines for Lipids, 2016: 30-84.
45. Fihn SD, Blankenship JC, Alexander KP, et al. 2014 ACC/AHA/AATS/PCNA/SCAI/STS focused update of the guideline for the diagnosis and management of patients with stable ischemic heart disease: a report of the American College of Cardiology/American Heart Association Task Force on Practice Guidelines, and the American Association for Thoracic Surgery, Preventive Cardiovascular Nurses Association, Society for Cardiovascular Angiography and Interventions, and Society of Thoracic Surgeons. J Am Coll Cardiol 2014;64:1929-49.
46. Sacco RL, Kasner SE, Broderick JP, et al. An updated definition of stroke for the 21st century: a statement for healthcare professionals from the American Heart Association/American Stroke Association. Stroke 2013;44:2064-89.
47. KDIGO CKD Work Group. KDIGO 2012 Clinical Practice Guideline for the Evaluation and Management of Chronic Kidney Disease. Kidney Int Suppl 2013;3:1-150.
48. European Association for the Study of the Liver. Electronic address: easloffice@easloffice.eu; European Association for the Study of the Liver. EASL Clinical Practice Guidelines for the management of patients with decompensated cirrhosis. J Hepatol 2018;69:406-60.

https://dx.doi.org/10.21037/jtd-22-40

49. Chalasani N, Younossi Z, Lavine JE, et al. The diagnosis and management of nonalcoholic fatty liver disease: Practice guidance from the American Association for the Study of Liver Diseases. Hepatology 2018;67:328-57.
50. Global Initiative for Asthma. Global Strategy for Asthma Management and Prevention, 2018. Available online: www.ginasthma.org
51. Aletaha D, Neogi T, Silman AJ, et al. 2010 Rheumatoid arthritis classification criteria: an American College of Rheumatology/European League Against Rheumatism collaborative initiative. Arthritis Rheum 2010;62:2569-81.
52. van den Hoogen F, Khanna D, Fransen J, et al. 2013 classification criteria for systemic sclerosis: an American College of Rheumatology/European League against Rheumatism collaborative initiative. Arthritis Rheum 2013;65:2737-47.
53. Shiboski SC, Shiboski CH, Criswell L, et al. American College of Rheumatology classification criteria for Sjögren's syndrome: a data-driven, expert consensus approach in the Sjögren's International Collaborative Clinical Alliance cohort. Arthritis Care Res (Hoboken) 2012;64:475-87.
54. Aringer M, Costenbader K, Daikh D, et al. 2019 European League Against Rheumatism/American College of Rheumatology Classification Criteria for Systemic Lupus Erythematosus. Arthritis Rheumatol 2019;71:1400-12.
55. Lundberg IE, Tjärnlund A, Bottai M, et al. 2017 European League Against Rheumatism/American College of Rheumatology Classification Criteria for Adult and Juvenile Idiopathic Inflammatory Myopathies and Their Major Subgroups. Arthritis Rheumatol 2017;69:2271-82.
56. Ministry of Health Singapore. Prevention, Diagnosis and Management of Tuberculosis. MOH Clinical Practical Guidelines 1/2016. Available online: http://www.moh.gov.sg/cpg
57. Lewinsohn DM, Leonard MK, LoBue PA, et al. Official American Thoracic Society/Infectious Diseases Society of America/Centers for Disease Control and Prevention Clinical Practice Guidelines: Diagnosis of Tuberculosis in Adults and Children. Clin Infect Dis 2017;64:e1-e33.
58. Katz PO, Gerson LB, Vela MF. Guidelines for the diagnosis and management of gastroesophageal reflux disease. Am J Gastroenterol 2013;108:308-28; quiz 329.
59. Satoh K, Yoshino J, Akamatsu T, et al. Evidence-based clinical practice guidelines for peptic ulcer disease 2015. J Gastroenterol 2016;51:177-94.
60. Garber JR, Cobin RH, Gharib H, et al. Clinical practice guidelines for hypothyroidism in adults: cosponsored by the American Association of Clinical Endocrinologists and the American Thyroid Association. Endocr Pract 2012;18:988-1028.
61. Bahn RS, Burch HB, Cooper DS, et al. Hyperthyroidism and other causes of thyrotoxicosis: management guidelines of the American Thyroid Association and American Association of Clinical Endocrinologists. Endocr Pract 2011;17:456-520.
62. Regier DA, Kuhl EA, Kupfer DJ. The DSM-5: Classification and criteria changes. World Psychiatry 2013;12:92-8.
63. 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;194:265-75.

https://dx.doi.org/10.21037/jtd-22-40

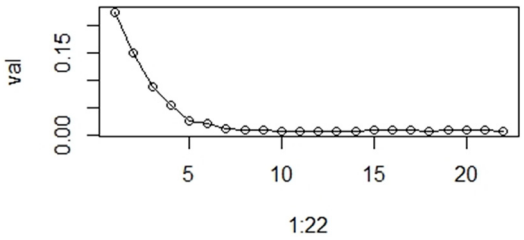


Figure S1 Scree plot to determine "k" value. The scree plot shows that the curve levels off at k=6.

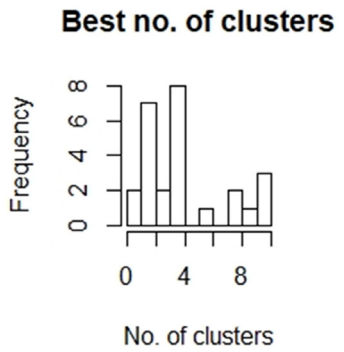


Figure S2 Histogram showing the best number of clusters. The histogram shows that the best number of clusters is 4.

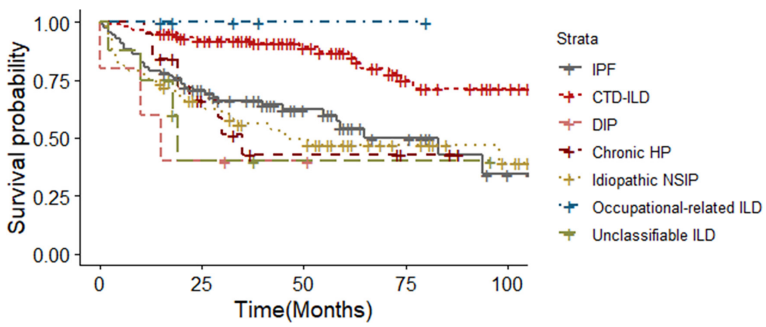


Figure S3 Kaplan-Meier curve comparing survival differences between ILD diagnosis for all-cause mortality. Diagnoses are coded according to the legend. CTD-ILD, connective tissue disease related interstitial lung disease; DIP, desquamative interstitial pneumonia; HP, hypersensitivity pneumonitis; ILD, interstitial lung disease; IPF, idiopathic pulmonary fibrosis; NSIP, idiopathic non-specific interstitial pneumonia.

https://dx.doi.org/10.21037/jtd-22-40

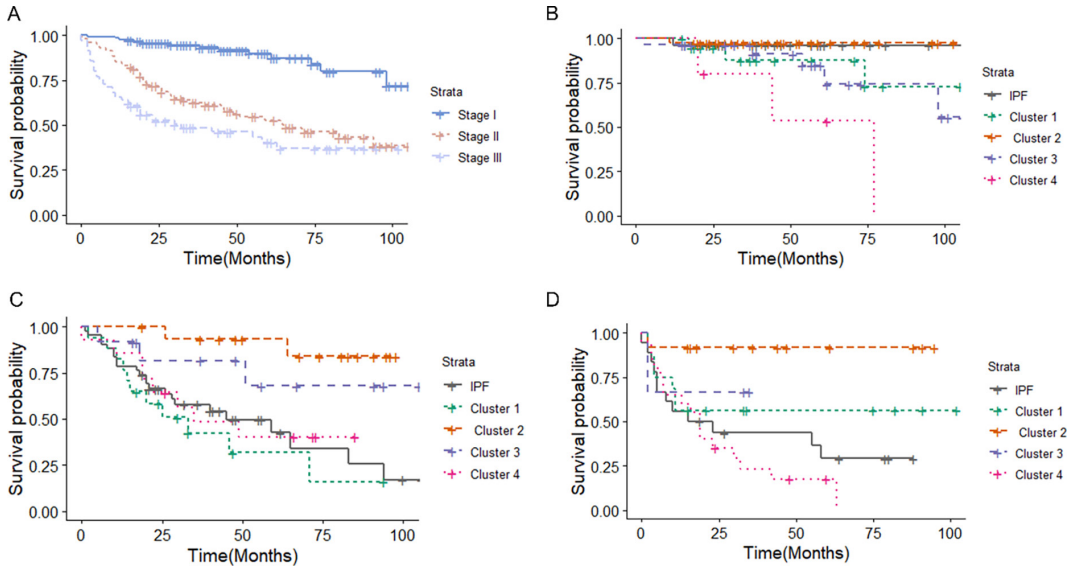


Figure S4 Kaplan-Meier curves comparing survival differences between GAP Stage for all-cause mortality (A); between clusters against IPF for all-cause mortality with GAP Stage I (B); between clusters against IPF for all-cause mortality with GAP Stage II (C); between clusters against IPF for all-cause mortality with GAP Stage III (D). GAP Stages are colour coded according to the legend. Clusters are coded according to the legend. IPF, idiopathic pulmonary fibrosis.

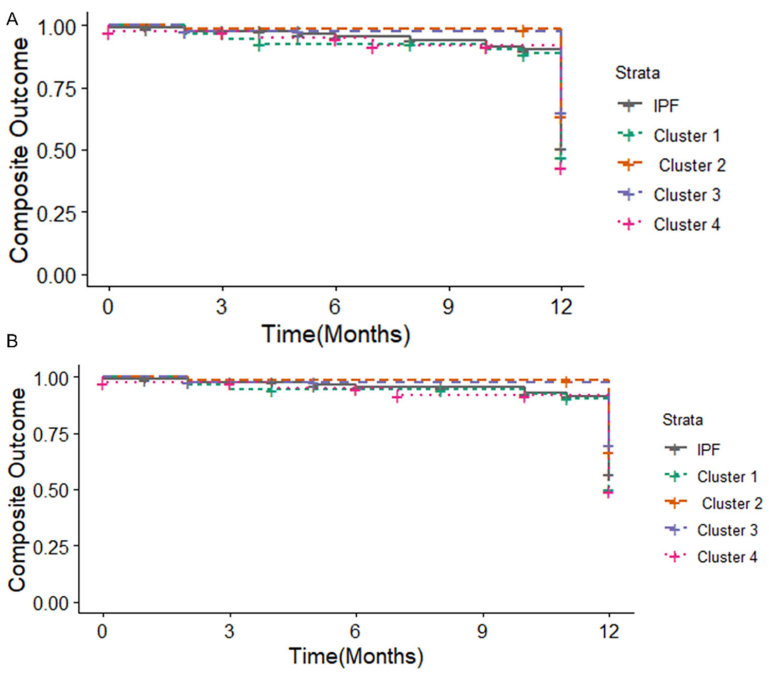


Figure S5 Kaplan-Meier curves comparing clusters against IPF for the composite outcome of decline in FVC of at least 5% of the predicted value from baseline or death at 12 months (A) and comparing clusters against IPF for the composite outcome of decline in FVC of at least 10% of the predicted value from baseline or death at 12 months (B). Clusters are coded according to the legend. IPF, idiopathic pulmonary fibrosis.

https://dx.doi.org/10.21037/jtd-22-40

Table S1 ILD diagnosis according to cluster

Variable	Cluster 1 (n=53)	Cluster 2 (n=67)	Cluster 3 (n=42)	Cluster 4 (n=39)	P value
Diagnosis, n (%)					
CTD-ILD	23 (43.4)	55 (82.1)	18 (42.9)	11 (28.2)	<0.001
Rheumatoid arthritis	9 (17.0)	3 (4.48)	4 (9.52)	3 (7.69)	
Antisynthetase syndrome	7 (13.2)	17 (25.4)	4 (9.62)	1 (2.56)	
Systemic sclerosis	2 (3.77)	19 (28.4)	3 (7.14)	1 (2.56)	
Sjogren syndrome	2 (3.77)	2 (2.99)	3 (7.14)	1 (2.56)	
Systemic lupus erythematosus	1 (1.89)	1 (1.49)	1 (2.38)	0 (0)	
Mixed connective tissue disease	1 (1.89)	9 (13.4)	1 (2.38)	0 (0)	
Undifferentiated connective tissue disease	1 (1.89)	4 (5.97)	2 (4.76)	5 (12.8)	
Chronic HP	7 (13.2)	3 (4.48)	2 (4.76)	7 (17.9)	0.066
Occupational-related ILD	4 (7.55)	0 (0)	1 (2.38)	0 (0)	0.040
Idiopathic NSIP	17 (32.1)	7 (10.4)	15 (35.7)	18 (46.2)	<0.001
DIP	2 (3.77)	1 (1.49)	1 (2.38)	1 (2.56)	0.887
Unclassifiable ILD	0 (0)	1 (1.49)	5 (11.9)	2 (5.12)	0.016
Radiology, n (%)[a]					
UIP	4 (7.55)	4 (5.97)	4 (9.52)	3 (7.69)	0.924
Possible UIP	11 (20.8)	28 (41.8)	10 (23.8)	13 (33.3)	0.061
Probable UIP	4 (7.55)	4 (5.97)	1 (2.38)	0 (0)	0.287
Indeterminate UIP	3 (5.66)	4 (5.97)	6 (14.3)	1 (2.56)	0.181
Inconsistent/alternative diagnosis	31 (58.5)	27 (40.3)	21 (50.0)	22 (56.4)	0.718

[a], for cases diagnosed prior to 2018, radiology description is based on the 2011 American Thoracic Society Idiopathic Pulmonary Fibrosis guidelines, and from 2018 onwards, based on the 2018 American Thoracic Society Idiopathic Pulmonary Fibrosis guideline. CTD-ILD, connective tissue disease related interstitial lung disease; HP, hypersensitivity pneumonitis; ILD, interstitial lung disease; NSIP, non-specific interstitial pneumonia; DIP, desquamative interstitial pneumonia.

Table S2 Bronchoalveolar lavage cell count patterns by cluster

Bronchoalveolar lavage cell count pattern	Cluster 1 (n=29)[a], n (%)	Cluster 2 (n=22)[b], n (%)	Cluster 3 (n=23)[c], n (%)	Cluster 4 (n=8)[d], n (%)	P value
Eosinophilic	3 (10.3)	1 (4.55)	2 (8.70)	2 (25.0)	0.208
Lymphocytic	2 (6.90)	1 (4.55)	1 (4.35)	1 (12.5)	0.688
Neutrophilic	8 (27.6)	4 (18.2)	7 (30.4)	2 (25.0)	0.014
Macrophagic	4 (13.8)	5 (22.7)	6 (26.1)	0 (0)	0.004
Eosinophilic and neutrophilic	5 (17.2)	6 (27.3)	4 (17.4)	1 (12.5)	0.374
Lymphocytic and neutrophilic	1 (3.45)	1 (4.55)	2 (8.70)	0 (0)	0.255
Neutrophilic, eosinophilic, and lymphocytic	6 (20.7)	4 (18.2)	1 (4.35)	2 (25.0)	0.737

[a], 29/53 patients underwent BAL; [b], 22/67 patients underwent BAL; [c], 23/42 patients underwent BAL; [d], 8/39 patients underwent BAL.

https://dx.doi.org/10.21037/jtd-22-40

Table S3 Summary of P values for characteristics of each cluster compared against IPF

Variable	Cluster 1, P value	Cluster 2, P value	Cluster 3, P value	Cluster 4, P value
Age	0.415	<0.001	<0.001	0.718
Sex	<0.001	<0.001	0.878	<0.001
Ethnicity	<0.001	0.468	0.529	<0.001
Chinese	0.001	0.521	0.664	<0.001
Malay	0.083	0.287	0.716	0.361
Indian	0.003	1	0.359	<0.001
Others	0.669	1	1	0.17
Smoker/ex-smoker	<0.001	<0.001	0.701	<0.001
No. of pack years	<0.001	<0.001	0.716	<0.001
Weight loss at presentation	0.335	1	0.285	0.973
BMI	0.158	0.439	0.243	0.002
Family history of ILD	0.7	1	1	0.782
Comorbid burden				
Low (0–1)	0.001	<0.001	0.001	0.221
Moderate (2–3)	0.078	0.226	0.058	0.492
High (4–6)	0.431	0.011	0.156	0.819
Diabetes mellitus	0.892	<0.001	0.002	1
Hypertension	0.004	<0.001	0.05	0.044
Hyperlipidaemia	0.069	<0.001	<0.001	1
Ischemic heart disease	0.393	<0.001	0.011	0.805
Thyroid disease	0.448	0.031	0.379	0.037
GERD, gastritis, peptic ulcer disease	0.153	0.722	1	0.873
Asthma	1	0.827	1	0.17
Cancer	0.642	0.937	0.098	0.605
Previous history of pulmonary tuberculosis	0.037	0.486	0.071	0.873
Pulmonary hypertension on 2D echo	0.942	1	0.581	0.755
Groundglass	<0.001	<0.001	<0.001	<0.001
Emphysema	0.012	<0.001	0.763	0.044
UIP pattern	<0.001	<0.001	<0.001	<0.001
Positive ANA >1:160	1	<0.001	0.203	0.95
FVC percentage predicted	0.271	0.005	0.972	0.972
DLCO percentage predicted	0.623	0.069	0.009	0.212
Antifibrotics	0.056	0.091	0.585	1

IPF, idiopathic pulmonary fibrosis; BMI, body mass index; GERD, gastroesophageal reflux disease; UIP, usual interstitial pneumonia; ANA, antinuclear antibody; FVC, forced vital capacity; DLCO, diffusion capacity for carbon monoxide.

https://dx.doi.org/10.21037/jtd-22-40

Table S4 Summary of outcomes against IPF according to clusters

Outcome	IPF	Cluster 1	Cluster 2	Cluster 3	Cluster 4	P value
Survival						<0.001
12-month	79.1 (70.9, 88.2)	79.2 (69.0, 91.0)	97.0 (93.0, 100)	92.9 (85.4, 100)	74.4 (61.8, 89.4)	
24-month	70.6 (61.6, 81.0)	68.6 (56.8, 82.8)	95.3 (90.1, 100)	90.2 (81.5, 99.8)	51.2 (37.6, 69.6)	
60-month	54.1 (42.6, 68.7)	54.7 (40.6, 73.7)	91.4 (83.0, 100)	71.9 (55.7, 92.7)	25.5 (14.1, 46.2)	
Median survival time (months)	65	74	>108	>108	30	
Annual change in FVC from baseline (mL)	−58.5±8.52	−47.0±9.64	−12.2±4.24	−4.22±2.88	−55.4±3.88	<0.001

FVC, forced vital capacity; IPF, idiopathic pulmonary fibrosis.

https://dx.doi.org/10.21037/jtd-22-40

Table S5 Comparison of mortality prediction between clusters, diagnosis and GAP stage

Variable	HR	95% CI	P value
Cluster (IPF as reference)			<0.001
Cluster 1	0.961	0.560, 1.648	0.885
Cluster 2	0.105	0.037, 0.295	<0.001
Cluster 3	0.413	0.198, 0.858	0.018
Cluster 4	1.974	1.202, 3.240	0.007
Diagnosis (IPF as reference)			<0.001
CTD-ILD	0.298	0.167, 0.531	<0.001
Chronic HP	1.155	0.556, 2.402	0.699
Occupational-related ILD	0.000	0.000, ∞	0.995
Idiopathic NSIP	1.191	0.729, 1.946	0.484
DIP	2.263	0.695, 7.368	0.175
Unclassifiable ILD	1.580	0.561, 4.451	0.387
GAP Stage (Stage I as reference)			
Stage II	4.867	2.632, 9.000	<0.001
Stage III	7.655	4.063, 14.380	<0.001
Cluster within GAP Stage			
Stage I (IPF as reference)			0.002
Cluster 1	3.542	0.367, 34.205	0.274
Cluster 2	0.551	0.034, 8.843	0.674
Cluster 3	4.016	0.464, 34.747	0.207
Cluster 4	14.489	1.501, 139.840	0.021
Stage II (IPF as reference)			0.003
Cluster 1	1.353	0.659, 2.781	0.410
Cluster 2	0.129	0.030, 0.550	0.006
Cluster 3	0.339	0.101, 1.135	0.080
Cluster 4	0.876	0.389, 1.973	0.750
Stage III (IPF as reference)			0.010
Cluster 1	0.647	0.255, 1.645	0.361
Cluster 2	0.096	0.0125, 0.739	0.024
Cluster 3	0.537	0.069, 4.164	0.552
Cluster 4	1.372	0.650, 2.894	0.407

GAP, gender-age-physiology; IPF, idiopathic pulmonary fibrosis; CTD-ILD, connective tissue disease related interstitial lung disease; HP, hypersensitivity pneumonitis; ILD, interstitial lung disease; NSIP, non-specific interstitial pneumonia; DIP, desquamative interstitial pneumonia.

https://dx.doi.org/10.21037/jtd-22-40

Table S6 Comparison in the annual rate of change of FVC according to antifibrotic use[a]

Treatment	Antifibrotic use	No antifibrotic use	P value
Cluster 3 annual change in FVC from baseline (mL)[b]	−3.66 (2.82)	−4.33 (1.10)	0.647
Cluster 4 annual change in FVC from baseline (mL)[c]	−50.9 (7.92)	−59.2 (13.3)	0.093

[a], Cluster 1 did not have any subjects on antifibrotic therapy, Cluster 2 has 1 subject on antifibrotic therapy. [b], 6 subjects from Cluster 3 were on antifibrotic therapy. [c], 3 subjects from Cluster 4 were on antifibrotic therapy. FVC, forced vital capacity.

https://dx.doi.org/10.21037/jtd-22-40