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

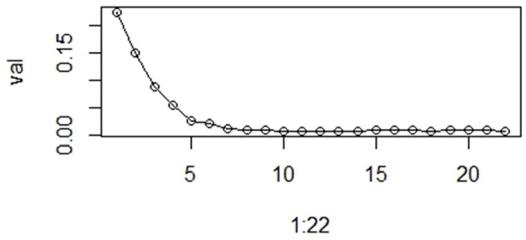

Figure S1 Scree plot to determine "k" value. The scree plot shows that the curve levels off at k=6.

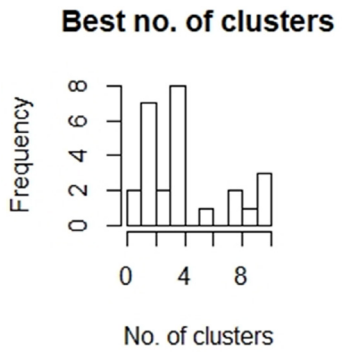

Figure S2 Histogram showing the best number of clusters. The histogram shows that the best number of clusters is 4.

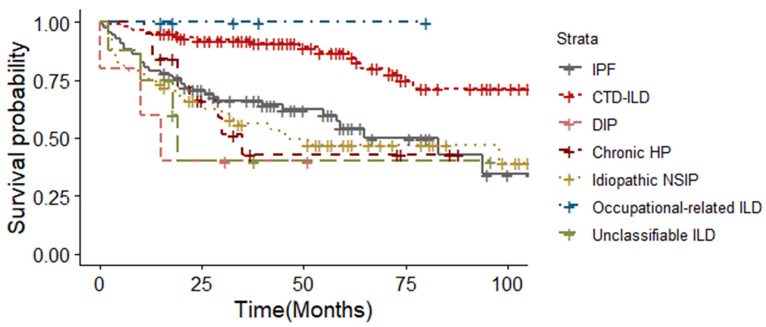

Figure S3 Kaplan-Meier curve comparing survival differences between ILD diagnosis for all-cause mortality. Diagnoses are coded according to the legend. CTD-ILD, connective tissue disease related interstitial lung disease; DIP, desquamative interstitial pneumonia; HP, hypersensitivity pneumonitis; ILD,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

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

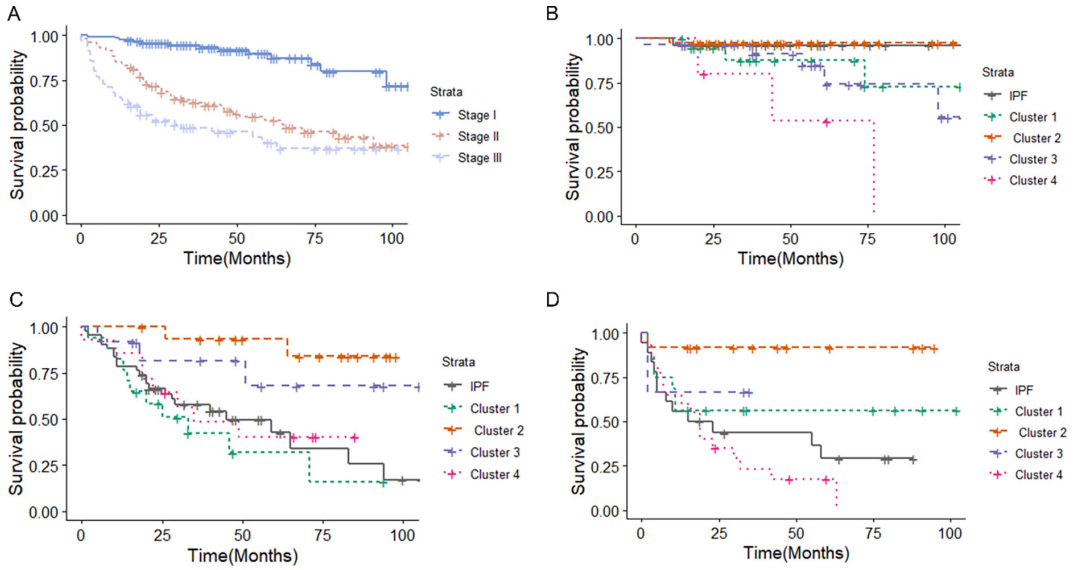

Figure S4 Kaplan-Meier curves comparing survival differences between GAP Stage for all-cause mortality (A); between clusters against IPF for all-cause mortality with GAP Stage I (B); between clusters against IPF for all-cause mortality with GAP Stage II (C); between clusters against IPF for all-cause 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.

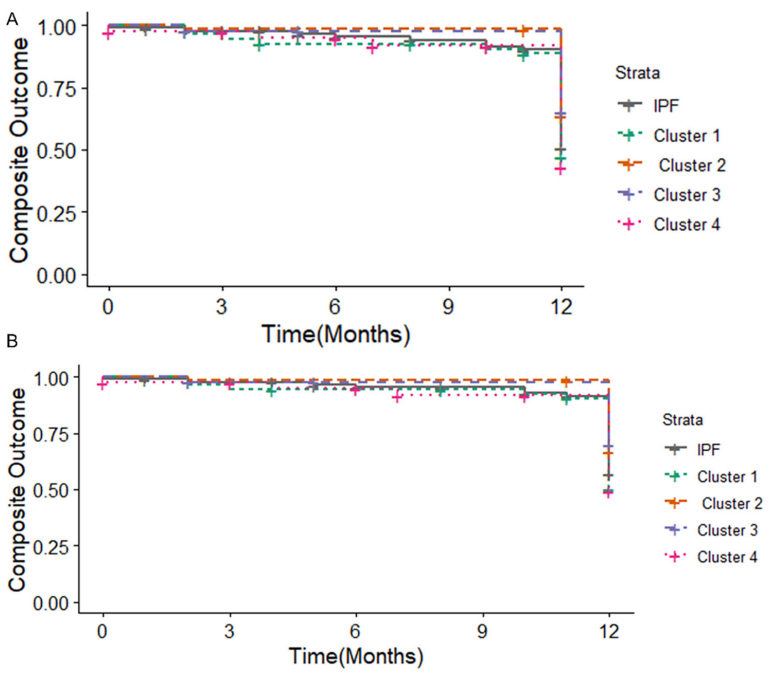

Figure S5 Kaplan-Meier curves comparing clusters against IPF for the composite outcome of decline in FVC of at least 5% of the predicted value from baseline or death at 12 months (A) and comparing clusters against IPF for the composite outcome of decline in FVC of at least 10% of the predicted value 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