

# Cluster Analysis-based Clinical Phenotypes of Idiopathic Interstitial Pneumonias: Associations With Acute Exacerbation and Overall Survival

**Yoichiro Aoshima**

Hamamatsu University School of Medicine

**Masato Karayama** (✉ [karayama@hama-med.ac.jp](mailto:karayama@hama-med.ac.jp))

Second Division, Department of Internal Medicine, Hamamatsu University School of Medicine, 1-20-1 Handayama, Hamamatsu 431-3192, Japan <https://orcid.org/0000-0003-0455-4070>

**Yasuoki Horiike**

Hamamatsu University School of Medicine

**Kazutaka Mori**

Shizuoka City Shimizu Hospital: Shizuoka Shiritsu Shimizu Byoin

**Hideki Yasui**

Hamamatsu University School of Medicine

**Hironao Hozumi**

Hamamatsu University School of Medicine

**Yuzo Suzuki**

Hamamatsu University School of Medicine

**Kazuki Furuhashi**

Hamamatsu University School of Medicine

**Tomoyuki Fujisawa**

Hamamatsu University School of Medicine

**Noriyuki Enomoto**

Hamamatsu University School of Medicine

**Yutaro Nakamura**

Hamamatsu University School of Medicine

**Naoki Inui**

Hamamatsu University School of Medicine

**Takafumi Suda**

Hamamatsu University School of Medicine

---

**Research**

**Keywords:** idiopathic pulmonary fibrosis, interstitial lung disease, interstitial pneumonia, IPAF, IPF

**DOI:** <https://doi.org/10.21203/rs.3.rs-77412/v1>

**License:**  This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

---

# Abstract

*Background:* The precise classification of idiopathic interstitial pneumonia (IIP) is essential for selecting treatment as well as estimating clinical outcomes; however, this is sometimes difficult in clinical practice. Therefore, cluster analysis was used to identify the clinical phenotypes of IIPs, and its usefulness for predicting clinical outcomes was evaluated.

*Methods:* Cluster analysis was performed using clinical features including patients' demographics; histories; pulmonary function test data; and laboratory, physical and radiological findings.

*Results:* In 337 patients with IIPs, four clusters were identified: Cluster I, in which >80% of the patients had autoimmune features; Cluster II, which had the lowest rate of smoking, the lowest percent predicted forced vital capacity (%FVC) and the lowest body mass index (BMI); Cluster III, which had the highest rate of smoking, the highest rate of dust exposure, the second lowest %FVC and normal BMI; and Cluster IV, which exhibited maintenance of %FVC and normal BMI. Cluster IV had significantly longer overall survival than Clusters II and III. Clusters I and III had significantly longer overall survival than Cluster II. Clusters II and III had a significantly higher cumulative incidence of acute exacerbation than Cluster IV.

*Conclusions:* Cluster analysis using clinical features identified four clinical phenotypes of IIPs, which may be useful for predicting the risk of acute exacerbation and overall survival.

## Background

Idiopathic interstitial pneumonias (IIPs) consist of heterogeneous interstitial lung diseases of unknown aetiology. Based on their clinical, radiological and pathologic features, IIPs comprise several disease entities, such as idiopathic pulmonary fibrosis (IPF), nonspecific interstitial pneumonia (NSIP) and cryptogenic organizing pneumonia (COP).<sup>1</sup>

The precise classification of IIPs is essential for selecting treatment and predicting prognosis; however, this is not simple in clinical practice.<sup>1</sup> First, diagnosis at a single point is sometimes difficult, but this can be achieved by monitoring disease behaviour. For example, some patients with IIPs lacking honeycombing experience disease progression over several years, and they develop a usual interstitial pneumonia (UIP) pattern characterized by honeycombing on high-resolution computed tomography (HRCT) that is eventually diagnosed as IPF. In these patients, it is difficult to accurately diagnose the disease in the first examination. Additionally, patients with advanced IIPs other than IPF sometimes develop radiological and pathological honeycombing that mimics IPF.<sup>2</sup> Second, pathological patterns can vary among different sites of the lungs in patients with IIPs.<sup>3</sup> If surgical lung biopsy was not performed for representative lesions, it is difficult to ensure an accurate diagnosis.<sup>4,5</sup> Furthermore, discordance in diagnosis even among specialists is frequent.<sup>6,7</sup> Third, surgical lung biopsy, an essential procedure for the diagnosis of IIPs, cannot be performed in all patients.<sup>8</sup>

A new trend of classifying interstitial lung diseases (ILDs) using clinical characteristics, regardless of conventional disease entities, has emerged.<sup>9</sup> Recently, the concept of progressive fibrosing ILD (PF-ILD) was proposed, and this concept covers several IIPs featuring self-sustaining fibrosis, a progressive decline in lung function and early mortality.<sup>9</sup> PF-ILD comprises a comprehensive group of ILDs, including IIPs, connective tissue disease-associated ILD (CTD-ILD), sarcoidosis and chronic hypersensitivity pneumonia (CHP). Although the concept of PF-ILD has not been validated, it is expected to play a certain role from a therapeutic perspective.<sup>10</sup>

Clinical phenotyping using cluster analysis, which groups subjects according to the similarities and differences of their clinical features, has recently attracted attention for classifying heterogeneous diseases. For example, clinical phenotypes determined via cluster analysis using clinical data reveal distinct clinical outcomes in asthma and COPD.<sup>11,12</sup> In all ILDs, clinical phenotypes determined using cluster analysis illustrated considerable predictive accuracy for clinical outcomes.<sup>13</sup> However, more than one-third of studied patients had non-IIPs such as CTD-ILD and CHP, both of which exhibit distinct characteristics and different clinical outcomes from IIPs.<sup>13</sup>

We hypothesized that the clinical phenotypes of IIPs provide more useful information about clinical outcomes among patients. In this study, we performed cluster analysis using the clinical data of patients with IIPs and evaluated the clinical utility of the phenotypes.

## Methods

### Study design

This retrospective observational study followed the ethical standards of the Declaration of Helsinki. The study protocol was approved by the Institutional Review Board of Hamamatsu University School of Medicine (Hamamatsu, Japan, approval No. E15-197).

### Patients

The medical records of consecutive patients with IIPs who were diagnosed at Hamamatsu Medical University Hospital between September 2004 and August 2018 were retrospectively analysed. The IIP diagnosis followed American Thoracic Society (ATS)/European Respiratory Society (ERS) guidelines.<sup>1,14-17</sup> Patients who exhibited acute exacerbation (AE) at the first visit were excluded.<sup>18</sup>

### Data collection

The following data were collected at the time of IIP diagnosis : age, sex, body mass index (BMI), pack-year smoking history, history of dust exposure, laboratory data (C-reactive protein [CRP], lactate dehydrogenase [LDH], albumin, Krebs von den Lungen-6 [KL-6], surfactant protein D [SP-D]), pulmonary function (percent predicted forced vital capacity [%FVC], percent predicted forced expiratory volume in 1 s [%FEV<sub>1</sub>]) and the clinical and pathological diagnoses (if performed). The existence of emphysema and

honeycombing was evaluated via chest high-resolution computed tomography (HRCT) at the diagnosis. The definitions of emphysema and honeycombing were described elsewhere.<sup>19</sup> Chest HRCT data were reviewed by two independent observers who were masked to patient data. Autoimmune features were recorded according to the interstitial pneumonia with autoimmune features (IPAF) diagnostic criteria,<sup>20</sup> namely positivity for two of the following three domains: clinical domain (extra-thoracic symptoms associated with autoimmune diseases), serologic domain (serum autoantibodies) and morphologic domain (HRCT and histopathologic findings and multi-compartment involvement other than IIPs).<sup>20</sup>

## Statistical analysis

Hierarchical clustering was performed using age, sex, BMI, histories of smoking and dust exposure, autoimmune features, laboratory data (CRP, LDH, albumin, KL-6, SP-D), pulmonary function (%FVC, %FEV<sub>1</sub>) and HRCT findings (emphysema, honeycombing) to identify clinical IIP subtypes. The number of clusters in which the scree plot of the distances between the clusters in a dendrogram of hierarchical clustering rose sharply, indicating different characteristics between clusters, was determined. Overall survival (OS) and the time to the first acute exacerbation (AE) were measured from the IIP diagnosis. The Kaplan–Meier method and log-rank test were used to analyse OS, and Grey’s test was used to analyse the time to the first AE. Wilcoxon’s signed-rank test was used for continuous variables, and Fisher’s exact test was applied for categorical variables. All comparisons among clusters were adjusted using Bonferroni’s correction. Data were expressed as the median (range) unless otherwise indicated. All statistical tests were two-sided, and  $p < 0.05$  indicated significance. All values were analysed using EZR (Saitama Medical Center, Jichi Medical University, Saitama, Japan), which is a graphical user interface for R (The R Foundation for Statistical Computing, Vienna, Austria), except the cluster analysis, which was performed using JMP v13.0.0 (SAS Institute Japan, Tokyo, Japan).

## Results

### Baseline characteristics

In total, 337 patients with IIPs were screened, and 54 were excluded because of missing data ( $n = 41$ ) and AE at the time of diagnosis ( $n = 13$ ), leaving 283 patients (Fig. 1). The clinical characteristics of these patients are presented in Table 1. Sixty-eight (24.0%) patients underwent surgical lung biopsy. Additionally, 94 (33.2%), 19 (6.7%), 16 (5.6%), 12 (4.2%) and 4 (1.4%) patients were diagnosed with IPF (clinical,  $n = 66$ ; pathological,  $n = 28$ ), COP, NSIP, pleuroparenchymal fibroelastosis (PPFE) and desquamative interstitial pneumonia (DIP)/respiratory bronchiolitis-associated interstitial lung disease (RB-ILD), respectively (Table 2). The remaining 138 patients (48.7%) had unclassifiable IIPs. Thirty-nine (13.7%) patients had autoimmune features, including 6 (17.4%), 39 (100%) and 36 (92.3%) patients in the clinical, serologic and morphologic domains, respectively. The median observation time was 39.4 months (range, 1.0–165.0 months).

Table 1  
Characteristics of the study patients

	All, n = 283	Cluster I, n = 35	Cluster II, n = 62	Cluster III, n = 89	Cluster IV, n = 97
Age	70.6 (20–90)	70.7 (52–83) <sup>#</sup>	75.9 (55–90) <sup>§, ¶</sup>	68.9 (20–89)	69.2 (41–88)
Sex, male	214 (75.6)	22 (62.8) <sup>§</sup>	49 (79.0) <sup>§, ¶</sup>	83 (93.2) <sup>¶</sup>	60 (61.8)
Body mass index	22.8 (13.1–36.8)	22.9 (18.1–29.7) <sup>#</sup>	20.8 (13.1–25.8) <sup>§, ¶</sup>	24.9 (19.4–36.8) <sup>¶</sup>	22.1 (14.9–29.8)
Smoking history	182 (64.3)	19 (54.2) <sup>§</sup>	31 (50.0) <sup>§</sup>	79 (88.7) <sup>¶</sup>	53 (54.6)
Pack-year smoking	42 (2.5–165)	30 (18.8–66) <sup>§</sup>	40 (5–100) <sup>§</sup>	49 (5–165) <sup>¶</sup>	40 (2.5–90)
Dust exposure	69 (24.3)	10 (28.5) <sup>#</sup>	6 (9.6) <sup>§</sup>	40 (44.9) <sup>¶</sup>	13 (13.4)
Organic materials	2 (0.7)	0 (0)	0 (0)	1 (1.1)	1 (1.0)
Inorganic material	67 (23.6)	10 (28.5)	6 (9.6)	39 (43.8)	12 (12.3)
Autoimmune features	39 (13.7)	30 (85.7)	6 (9.6)	1 (1.1)	2 (2.0)
Spirometry					
% predicted FVC	78.5 (27.6–124.0)	81.8 (31.6–113.0) <sup>#</sup>	64.5 (27.6–116.0) <sup>§, ¶</sup>	73.0 (39.8–112.0) <sup>¶</sup>	89.1 (40.9–124.0)
% predicted FEV <sub>1</sub>	81.0 (31.9–131.6)	82.3 (66.1–114.0) <sup>#, §</sup>	68.9 (31.9–100.0) <sup>§, ¶</sup>	72.4 (41.4–117.0) <sup>¶</sup>	87.6 (61.0–131.6)
FEV <sub>1</sub> /FVC	82.4 (53.4–100.0)	81.7 (65.8–100.0)	87.5 (63.4–100.0) <sup>§</sup>	81.0 (53.4–98.7)	82.3 (61.3–98.8)
Laboratory data					
CRP (mg/dL)	0.20 (0.01–25.5)	0.27 (0.01–4.8) <sup>#, ¶</sup>	1.13 (0.02–25.5) <sup>§, ¶</sup>	0.23 (0.02–9.0) <sup>¶</sup>	0.10 (0.01–4.7)
Data are presented as the median (range) or number (%).					
CRP, C-reactive protein; FEV <sub>1</sub> , forced expiratory volume in 1 s; FVC, forced vital capacity; IIPs, idiopathic interstitial pneumonias; IPAF, interstitial pneumonia with autoimmune features; KL-6, Krebs von den Lungen-6; LDH, lactate dehydrogenase; SP-D, pulmonary surfactant protein-D					
<sup>#</sup> <i>p</i> < 0.05 compared with Cluster II					
<sup>§</sup> <i>p</i> < 0.05 compared with Cluster III					
<sup>¶</sup> <i>p</i> < 0.05 compared with Cluster IV					

	All, n = 283	Cluster I, n = 35	Cluster II, n = 62	Cluster III, n = 89	Cluster IV, n = 97
LDH (U/L)	225 (94–569)	230 (94–319) <sup>¶</sup>	213 (132–450) <sup>§</sup>	241 (170–569) <sup>¶</sup>	205 (141–334)
Albumin (g/dL)	4.1 (1.9–4.9)	4.0 (2.4–4.9) <sup>#, §, ¶</sup>	3.6 (1.9–4.5) <sup>§, ¶</sup>	4.1 (2.8–4.8) <sup>¶</sup>	4.2 (3.4–4.9)
KL-6 (U/mL)	829 (112–5710)	1010 (164–3651)	664 (112–2070) <sup>§</sup>	1148 (177–5710) <sup>¶</sup>	644 (196–3378)
SP-D (ng/mL)	179 (17.2–1160)	166 (25.4–533) <sup>§</sup>	173 (33.4–1130) <sup>§</sup>	218 (49.8–1160) <sup>¶</sup>	154 (17.2–525)
CT findings					
Emphysema	90 (31.8)	11 (31.4)	14 (22.5) <sup>§</sup>	35 (39.3)	30 (30.9)
Honeycombing	82 (28.9)	5 (14.2) <sup>#, §</sup>	23 (37.0) <sup>¶</sup>	34 (38.2) <sup>¶</sup>	20 (20.6)
Treatments					
Steroids	96 (33.9)	15 (42.8)	21 (33.8)	36 (40.4) <sup>¶</sup>	24 (24.7)
Immunosuppressants	35 (12.3)	4 (11.4)	5 (8.0)	17 (19.1)	9 (9.2)
Antifibrotic agents	56 (19.7)	3 (8.5) <sup>§</sup>	13 (20.9)	28 (31.4) <sup>¶</sup>	12 (12.3)
Data are presented as the median (range) or number (%).					
CRP, C-reactive protein; FEV <sub>1</sub> , forced expiratory volume in 1 s; FVC, forced vital capacity; IIPs, idiopathic interstitial pneumonias; IPAF, interstitial pneumonia with autoimmune features; KL-6, Krebs von den Lungen-6; LDH, lactate dehydrogenase; SP-D, pulmonary surfactant protein-D					
<sup>#</sup> <i>p</i> < 0.05 compared with Cluster II					
<sup>§</sup> <i>p</i> < 0.05 compared with Cluster III					
<sup>¶</sup> <i>p</i> < 0.05 compared with Cluster IV					

Table 2  
Characteristics of the study patients

	All, n = 283	Cluster I, n = 35	Cluster II, n = 62	Cluster III, n = 89	Cluster IV, n = 97
IPF	94 (33.2)	6 (17.1)	22 (35.4)	46 (51.6)	20 (20.6)
NSIP	16 (5.6)	5 (14.2)	0 (0)	5 (5.6)	6 (6.1)
COP	19 (6.7)	3 (8.5)	8 (12.9)	3 (3.3)	5 (5.1)
DIP / RB-ILD	4 (1.4)	2 (5.7)	0 (0)	0 (0)	2 (2.0)
PPFE	12 (4.2)	0 (0)	8 (12.9)	0 (0)	4 (4.1)
Unclassifiable IIPs	138 (48.7)	19 (54.2)	24 (38.7)	35 (39.3)	60 (61.8)
Data are expressed as number (%).					
COP, cryptogenic organizing pneumonia; DIP, desquamative interstitial pneumonia; IIP, idiopathic interstitial pneumonia; IPF, idiopathic pulmonary fibrosis; NSIP, non-specific interstitial pneumonia; PPFE, pleuroparenchymal fibroelastosis; RB-ILD, respiratory bronchiolitis-associated interstitial lung disease					

## Clinical characteristics of the clusters

Four clusters were identified in the cluster analysis using clinical data (Supplementary Fig. 1).

Cluster I (n = 35) was an intermediate-aged cohort (70.7 years) with low proportions of males (62.8%) and smokers (54.2%; Table 1, Fig. 2A). Cluster I featured preserved %FVC (81.8%), the lowest honeycombing rate (14.2%) and the highest rate of autoimmune features (85.7%). Regarding laboratory data, Cluster I had the second highest KL-6 level (1010 U/mL), a slightly increased SP-D level (166 ng/mL) and normal CRP, LDH and albumin levels. In Cluster I, six (17.1%), five (14.2%), three (8.5%) and two (5.7%) patients were diagnosed with IPF (pathological, n = 3; clinical, n = 3), NSIP, COP and DIP, respectively. The remaining 19 (54.2%) patients had unclassifiable IIP (Table 2).

Cluster II (n = 62) had the oldest population (75.9 years), the smallest percentage of smokers (50.0%), the lowest BMI (20.8) and the lowest %FVC (64.5%; Table 1, Fig. 2B), as well as the second largest male population (79.0%). Emphysema was least frequent in this group (22.5%), but the second highest rate of honeycombing was noted (37.0%). In laboratory findings, Cluster II had the highest CRP level (1.13 mg/dL), the lowest serum albumin level (3.6 g/dL) and slightly increased KL-6 (664 U/mL) and SP-D levels (173 ng/mL). In total, 22 (35.4%), 8 (12.9%) and 8 (12.9%) patients were diagnosed with IPF (clinical, n = 20; pathological, n = 2), PPFE and COP, respectively. The remaining 24 (38.7%) patients had unclassifiable IIP (Table 2).

Cluster III (n = 89) had the youngest population (68.9 years), the highest proportion of males (93.2%), the highest BMI (24.9) and the highest frequencies of smoking (88.7%) and dust exposure (44.9%; Table 1,

Fig. 2C). This cluster had the second lowest %FVC (73.0%). Cluster III had the highest LDH (241 U/L), KL-6 (1148 U/mL) and SP-D levels (218 ng/mL) and normal CRP and albumin levels. The rates of emphysema and honeycombing were highest in this group (39.3 and 38.2%, respectively). In total, 46 (51.6%), 5 (5.6%) and 3 (3.3%) patients were diagnosed with IPF (clinical,  $n = 29$ ; pathological,  $n = 17$ ), NSIP and COP, respectively. The remaining 35 (39.3%) patients had unclassifiable IIP (Table 2).

Cluster IV ( $n = 97$ ) had the second youngest population (69.2 years), the highest proportion of females (38.2%), normal BMI (22.1) and lower frequencies of smoking (54.6%) and dust exposure (13.4%; Table 1, Fig. 2D). Cluster IV had the highest %FVC (89.1%) and normal CRP, LDH and albumin levels. Cluster IV had the lowest KL-6 (644 U/mL) and SP-D levels (154 ng/mL). The rates of emphysema and honeycombing were 30.9 and 20.6%, respectively. Overall, 20 (20.6%), 6 (6.1%), 5 (5.1%), 4 (4.1%) and 2 patients (2.0%) were diagnosed with IPF (clinical,  $n = 13$ ; pathological,  $n = 7$ ), NSIP, and COP, PPFE and DIP, respectively. The remaining 60 (61.8%) patients had unclassifiable IIPs (Table 2).

## Differences in clinical outcomes among the clusters

OS was significantly longer in Cluster IV than in Clusters II ( $p < 0.01$ ) and III ( $p = 0.01$ ; Fig. 3A). Clusters I and III had significantly longer OS than Cluster II (both  $p < 0.01$ ). The 5-year OS rates were 87.7, 79.0, 70.1 and 32.3% in Clusters IV, I, III and II, respectively, and the 10-year OS rates were 66.6%, 57.6%, 41.4% and 32.3%, respectively.

Clusters II and III had significantly higher cumulative incidences of AE than Cluster IV (both  $p = 0.03$ ; Fig. 4A). The 5-year cumulative incidence rates of AE were 20.6%, 19.9%, 9.7% and 8.0% in Clusters II, III, I, and IV, respectively, and the 10-year cumulative incidence rates were 20.6%, 23.0%, 21.7% and 13.8%, respectively.

Next, OS and AE were evaluated according to the IIP diagnosis. Patients with NSIP had significantly longer OS than those with IPF ( $p = 0.01$ ), whereas there was no significant difference among the other groups (Fig. 3B). Patients with IPF had a significantly higher cumulative incidence of AE than those with COP ( $p = 0.01$ ) and unclassifiable IIPs ( $p < 0.01$ ), whereas there was no significant difference among the other groups (Fig. 4B).

## Discussion

Four IIP clusters with distinct clinical features and outcomes were identified. Cluster I, in which > 80% of the patients had autoimmune features, featured a low risk for AE and good OS. Clusters II and III had the greatest decreases in %FVC and the highest honeycombing rates. More than 80% of patients with IPF belonged to these two clusters. Cluster II had the lowest %FVC, BMI and albumin levels. Cluster III had the second lowest %FVC and normal BMI and albumin levels, as well as the highest LDH, KL-6 and SP-D levels. OS was worst in Cluster II, followed by Cluster III. Both clusters had high risks of AEs. Cluster IV, representing the mildest IIPs with preserved %FVC and almost normal laboratory data, had the best OS and a low risk of AE. In our patients, more than half did not undergo surgical lung biopsies. Additionally,

the radiologic findings used in the cluster analysis were simple; namely the existence of honeycombing and emphysema on chest CT. Our clustering of patients with IIPs using clinical data is highly feasible in practice, and this strategy may provide useful information about patients' clinical outcomes, even without detailed pathological or radiologic findings.

Interestingly, our cluster analysis revealed that the clinical features associated with collagen vascular diseases were important components of the clinical phenotypes of IIPs. A considerable number of patients with IIPs have some features of collagen vascular disease without fulfilling the defined criteria of any collagen vascular diseases. Several classifications have been proposed for such patients, but controversies remain unresolved.<sup>21–24</sup> The IPAF concept was proposed by ATS/ERS to establish a unified platform for such patients with IIPs.<sup>20</sup> Several studies reported that patients with IPAF have better OS and lower risks of AE than their counterparts.<sup>25,26</sup> Similarly, Cluster I, in which most patients had autoimmune features, was linked to better clinical outcomes than Clusters II and III. Additionally, patients with IPAF are younger and more commonly female than those without IPAF, coinciding with the characteristics of Cluster I.<sup>27–29</sup> The clinical significance and utility of the IPAF concept are poorly validated. Additionally, several studies reported that the IPAF criteria must be revisited. In this context, our simple clustering successfully extracts patients with IIPs and autoimmune features who have distinct outcomes (Cluster I).

Clusters II and III were characterized by advanced fibrosis accompanied by decreased %FVC and honeycombing, which may lead to poor prognoses and high risks of AE. Most study patients with IPF belonged to these clusters. Patients with IPF have higher AE rates and worse OS than those with other IIPs.<sup>30</sup> The high prevalence of IPF might have strong effects on clinical outcomes in Clusters II and III. However, other factors determined the characteristics of these clusters. More than 50% of patients in these clusters had non-IPF IIPs. Specifically, OS was poorest in Cluster II despite its lower proportion of patients with IPF than Cluster III. In this study, we identified age, low serum albumin levels, decreased %FVC and radiologic honeycombing as independent risk factors for OS in all patients with IIP (Supplementary Table 1). Additionally, the latter two were also risk factors for AE (Supplementary Table 2). Among the two clusters, Cluster II was associated with significantly older patient age, lower serum albumin levels, lower %FVC and higher rates of radiologic honeycombing than Cluster III. These differences may explain the clinical outcomes of Clusters II and III.

The current study had several limitations. First, the numbers of clusters affected the results. If we had employed three clusters, Clusters III and IV would have been merged (Supplementary Fig. 1). However, Clusters III and IV had distinct clinical characteristics and clinical outcomes, necessitating their separation. Alternatively, if we had employed five clusters, Cluster II would have been divided into two clusters. One of the new clusters had a trend toward a lower proportion of males, lower BMI and lower %FVC than the other cluster (Supplementary Tables 4–5, Supplementary Figs. 2–3). Although these two new clusters might provide additional information, they comprised small numbers of patients (not enough to reach statistical significance); thus, they were grouped together. Second, the numbers and/or optimal combination of variables used for the cluster analysis were not validated. We employed clinical

variables that could be representative of demographic, historical, physical, laboratory and radiographic information. However, there were limited data available because the study was retrospective. The selection of variables can vary according to the purpose of the classification, such as predicting OS, AE risk or therapeutic response. Further studies are warranted to optimize the clinical classifications of IIPs.

## Conclusion

Cluster analysis using clinical data identified four IIP phenotypes with distinct differences in the AE risks and OS.

## Declarations

**Ethics approval and consent to participate:** The study protocol was approved by the Institutional Review Board of Hamamatsu University School of Medicine (Hamamatsu, Japan, approval No. E15-197). Informed consents were waived because of a retrospective observational study.

**Consent for publication:** Not applicable

**Availability of data and materials:** The datasets used and/or analysed during the current study are available from the corresponding author on reasonable request.

**Competing interests:** All authors have no conflict of interest in relation to this manuscript.

**Funding:** This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

### Author contributions:

YA: interpretation of data, writing-original draft, MK: conceptualization, methodology, interpretation of data, writing-review and editing, and supervision. KM: acquisition and analysis, and interpretation of data, YA, HY: acquisition and analysis. HH, YS, KF: conceptualization and supervision. TF, NE, YN, NI: conceptualization, supervision, and writing-review and editing. TS: conceptualization, methodology, writing-review and editing, and supervision.

**Acknowledgement:** We thank Joe Barber Jr., PhD, from Edanz Group (<https://en-author-services.edanzgroup.com/ac>) for editing a draft of this manuscript.

## References

1. Travis WD, Costabel U, Hansell DM, King TE Jr, Lynch DA, Nicholson AG, Ryerson CJ, Ryu JH, Selman M, Wells AU, 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.

2. Watadani T, Sakai F, Johkoh T, Noma S, Akira M, Fujimoto K, Bankier AA, Lee KS, Müller NL, Song JW, et al. Interobserver variability in the CT assessment of honeycombing in the lungs. *Radiology*. 2013;266:936–44.
3. Barratt SL, Creamer A, Hayton C, Chaudhuri N. Idiopathic pulmonary fibrosis (IPF): An Overview. *J Clin Med*. 2018;7:201.
4. Yagihashi K, Huckleberry J, Colby TV, Tazelaar HD, Zach J, Sundaram B, Pipavath S, Schwarz MI, Lynch DA. Radiologic-pathologic discordance in biopsy-proven usual interstitial pneumonia. *Eur Respir J*. 2016;47:1189–97.
5. Hutchinson JP, Fogarty AW, McKeever TM, Hubbard RB. In-hospital mortality after surgical lung biopsy for interstitial lung disease in the United States. 2000 to 2011. *Am J Respir Crit Care Med*. 2016;193:1161–7.
6. Sumikawa H, Johkoh T, Colby TV, Ichikado K, Suga M, Taniguchi H, Kondoh Y, Ogura T, Arakawa H, Fujimoto K, et al. Computed tomography findings in pathological usual interstitial pneumonia: relationship to survival. *Am J Respir Crit Care Med*. 2008;177:433–9.
7. Sverzellati N, Wells AU, Tomassetti S, Desai SR, Copley SJ, Aziz ZA, Zompatori M, Chilosi M, Nicholson AG, Poletti V, et al. Biopsy-proved idiopathic pulmonary fibrosis: spectrum of nondiagnostic thin-section CT diagnoses. *Radiology*. 2010;254:957–64.
8. Ryerson CJ, Urbania TH, Richeldi L, Mooney JJ, Lee JS, Jones KD, Elicker BM, Koth LL, King TE Jr, Wolters PJ, et al. Prevalence and prognosis of unclassifiable interstitial lung disease. *Eur Respir J*. 2013;42:750–7.
9. Cottin V, Hirani NA, Hotchkiss DL, Nambiar AM, Ogura T, Otaola M, Skowasch D, Park JS, Poonyagariyagorn HK, Wuyts W, et al. Presentation, diagnosis and clinical course of the spectrum of progressive-fibrosing interstitial lung diseases. *Eur Respir Rev*. 2018;27:180076.
10. Flaherty KR, Wells AU, Cottin V, Devaraj A, Walsh SLF, Inoue Y, Richeldi L, Kolb M, Tetzlaff K, Stowasser S, et al. Nintedanib in progressive fibrosing interstitial lung diseases. *N Engl J Med*. 2019;381:1718–27.
11. Rennard SI, Locantore N, Delafont B, Tal-Singer R, Silverman EK, Vestbo J, Miller BE, Bakke P, Celli B, Calverley PM, et al. Identification of five chronic obstructive pulmonary disease subgroups with different prognoses in the ECLIPSE cohort using cluster analysis. *Ann Am Thorac Soc*. 2015;12:303–12.
12. Moore WC, Meyers DA, Wenzel SE, Teague WG, Li H, Li X, D'Agostino R Jr, Castro M, Curran-Everett D, Fitzpatrick AM, et al. Identification of asthma phenotypes using cluster analysis in the Severe Asthma Research Program. *Am J Respir Crit Care Med*. 2010;181:315–23.
13. Adegunsoye A, Oldham JM, Chung JH, Montner SM, Lee C, Witt LJ, Stahlbaum D, Bermea RS, Chen LW, Hsu S, et al. Phenotypic clusters predict outcomes in a longitudinal interstitial lung disease cohort. *Chest*. 2018;153:349–60.
14. Fischer A, Antoniou KM, Brown KK, Cadranel J, Corte TJ, du Bois RM, Lee JS, Leslie KO, Lynch DA, Matteson EL, et al. An official European Respiratory Society/American Thoracic Society research

- statement: interstitial pneumonia with autoimmune features. *Eur Respir J*. 2015;46:976–87.
15. 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.
  16. Raghu G, Collard HR, Egan JJ, Martinez FJ, Behr J, Brown KK, Colby TV, Cordier JF, Flaherty KR, Lasky JA, 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.
  17. Raghu G, Remy-Jardin M, Myers JL, Richeldi L, Ryerson CJ, Lederer DJ, Behr J, Cottin V, Danoff SK, Morell F, 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.
  18. Collard HR, Ryerson CJ, Corte TJ, Jenkins G, Kondoh Y, Lederer DJ, Lee JS, Maher TM, Wells AU, Antoniou KM, et al. Acute exacerbation of idiopathic pulmonary fibrosis an international working group report. *Am J Respir Crit Care Med*. 2016;194:265–75.
  19. Hansell DM, Bankier AA, MacMahon H, McLoud TC, Müller NL, Remy J. Fleischner Society: glossary of terms for thoracic imaging. *Radiology*. 2008;246:697–722.
  20. Fischer A, Antoniou KM, Brown KK, Cadranel J, Corte TJ, du Bois RM, Lee JS, Leslie KO, Lynch DA, Matteson EL, et al. An official European Respiratory Society / American Thoracic Society research statement: interstitial pneumonia with autoimmune features. *Eur Respir J*. 2015;46:976–87.
  21. Corte TJ, Copley SJ, Desai SR, Zappala CJ, Hansell DM, Nicholson AG, Colby TV, Renzoni E, Maher TM, Wells AU. Significance of connective tissue disease features in idiopathic interstitial pneumonia. *Eur Respir J*. 2012;39:661–8.
  22. Kinder BW, Collard HR, Koth L, Daikh DI, Wolters PJ, Elicker B, Jones KD, King TE Jr. Idiopathic nonspecific interstitial pneumonia: lung manifestation of undifferentiated connective tissue disease? *Am J Respir Crit Care Med*. 2007;176:691–7.
  23. Vij R, Noth I, Strek ME. Autoimmune-featured interstitial lung disease: a distinct entity. *Chest*. 2011;140:1292–9.
  24. Omote N, Taniguchi H, Kondoh Y, Watanabe N, Sakamoto K, Kimura T, Kataoka K, Johkoh T, Fujimoto K, Fukuoka J, et al. Lung-dominant connective tissue disease: clinical, radiologic, and histologic features. *Chest*. 2015;148:1438–46.
  25. Yoshimura K, Kono M, Enomoto Y, Nishimoto K, Oyama Y, Yasui H, Hozumi H, Karayama M, Suzuki Y, Furuhashi K, et al. Distinctive characteristics and prognostic significance of interstitial pneumonia with autoimmune features in patients with chronic fibrosing interstitial pneumonia. *Respir Med*. 2018;137:167–75.
  26. Lim JU, Gil BM, Kang HS, Oh J, Kim YH, Kwon SS. Interstitial pneumonia with autoimmune features show better survival and less exacerbations compared to idiopathic pulmonary fibrosis. *BMC Pulm Med*. 2019;19:120.
  27. Ahmad K, Barba T, Gamondes D, Ginoux M, Khouatra C, Spagnolo P, Strek M, Thivolet-Béjui F, Traclet J, Cottin V. Interstitial pneumonia with autoimmune features: Clinical, radiologic, and histological

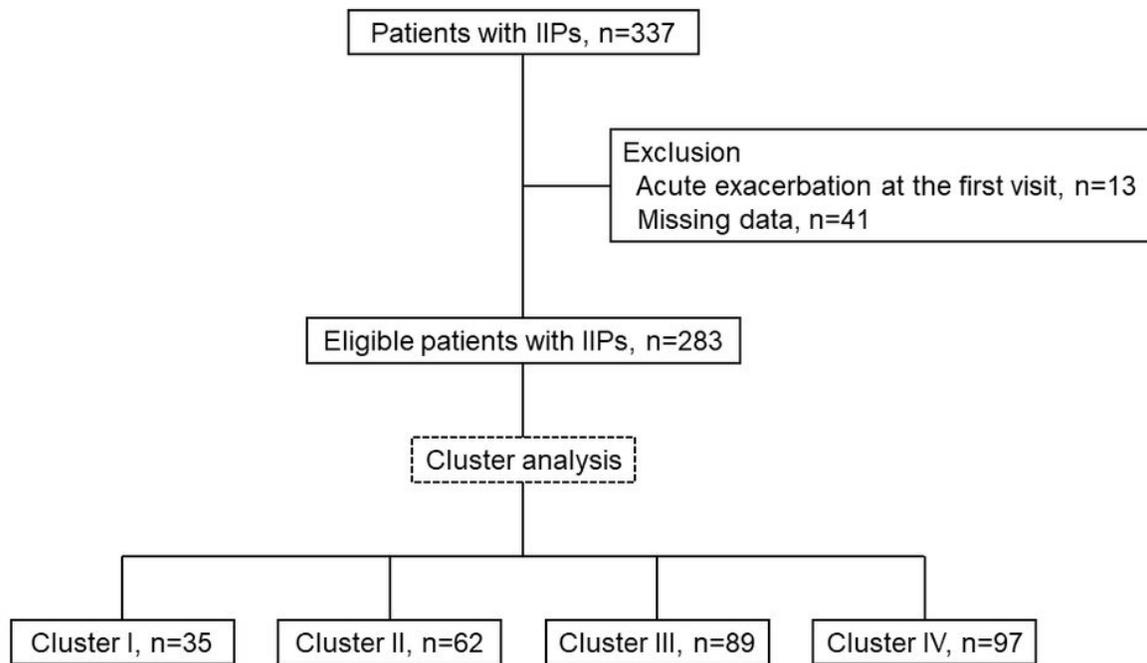
characteristics and outcome in a series of 57 patients. *Respir Med.* 2017;123:56–2.

28. Chartrand S, Swigris JJ, Stanchev L, Lee JS, Brown KK, Fischer A. Clinical features and natural history of interstitial pneumonia with autoimmune features: A single center experience. *Respir Med.* 2016;119:150–4.

29. Oldham JM, Adegunsoye A, Valenzi E, Lee C, Witt L, Chen L, Husain AN, Montner S, Chung JH, Cottin V, et al. Characterisation of patients with interstitial pneumonia with autoimmune features. *Eur Respir J.* 2016;47:1767–75.

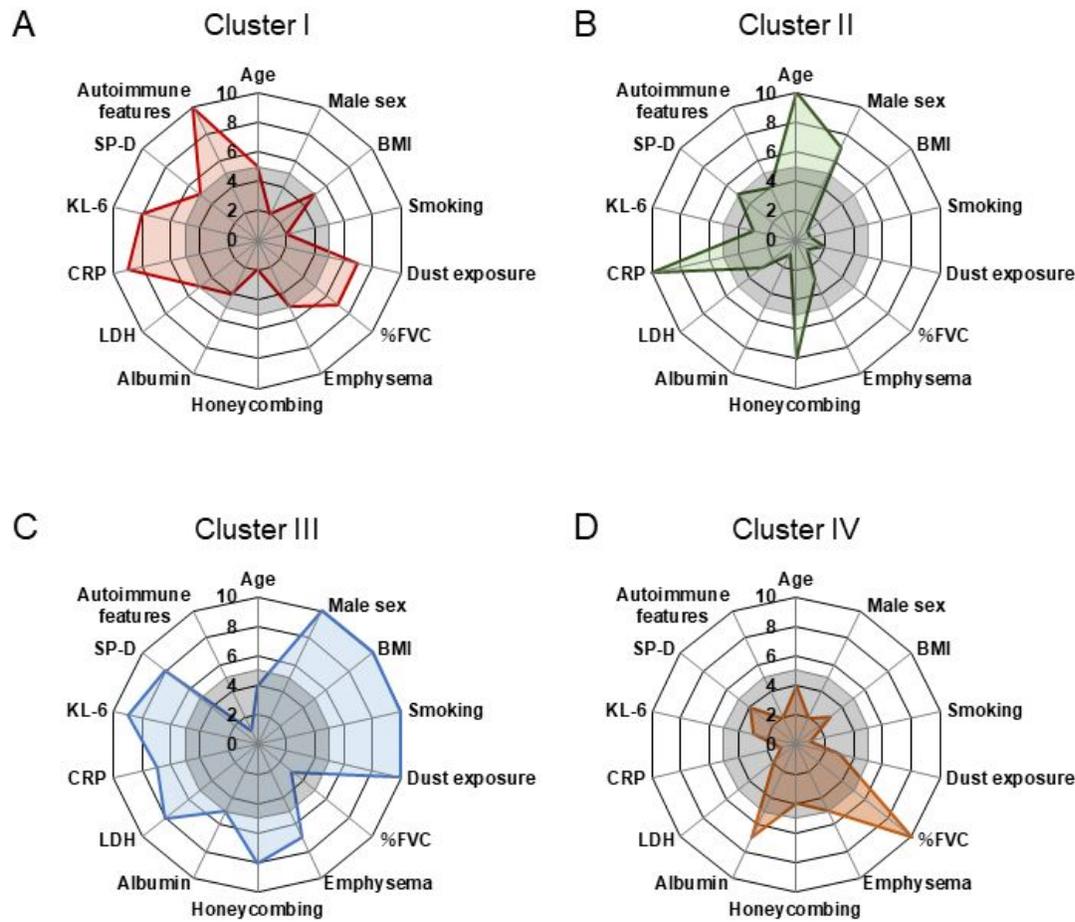
30. Leuschner G, Behr J. Acute exacerbation in interstitial lung disease. *Front Med.* 2017;4:176.

## Figures



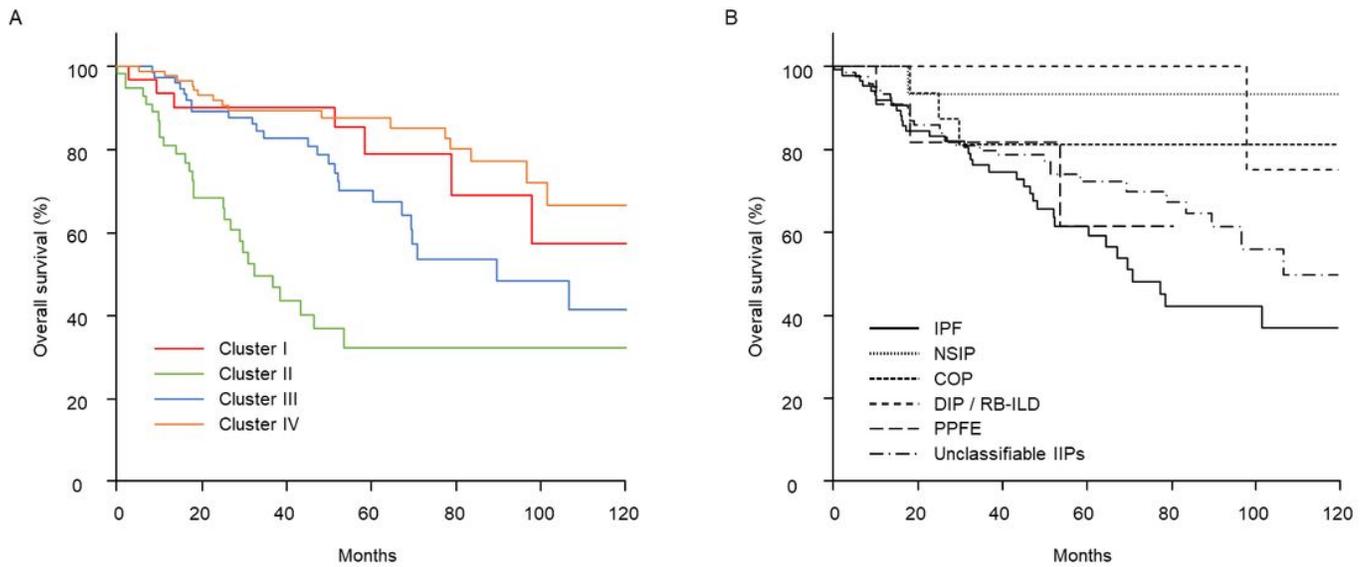
**Figure 1**

Flow diagram of the study patients IIP, idiopathic interstitial pneumonia



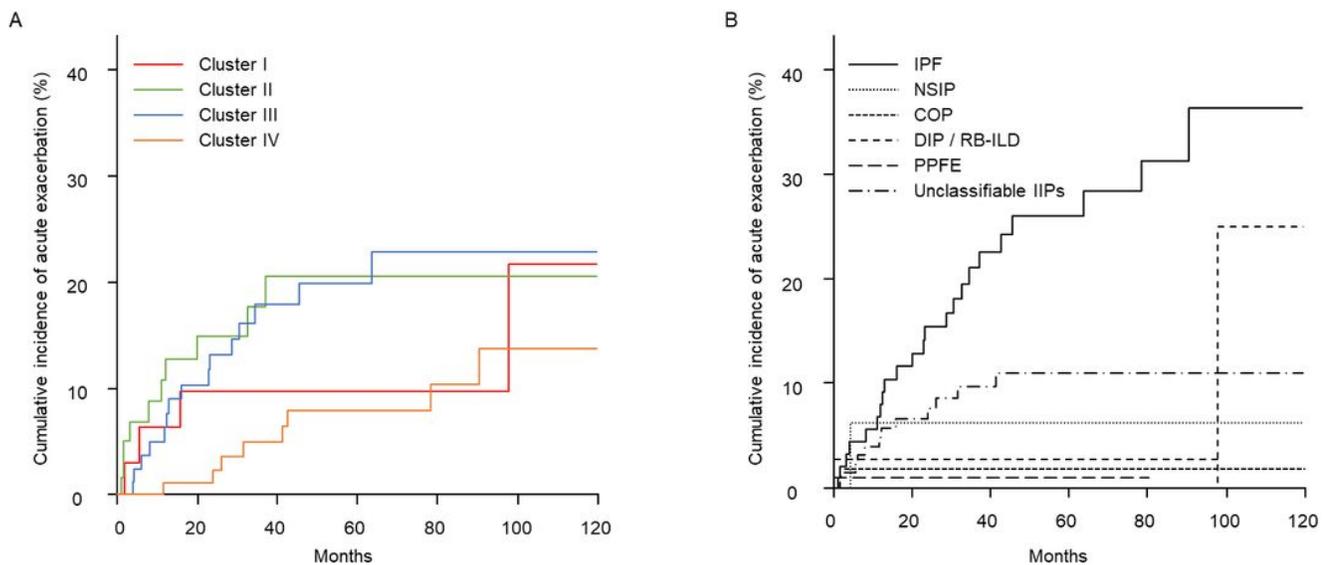
**Figure 2**

Radar plot of clinical features of the four clusters A) Cluster I, B) Cluster II, C) Cluster III, and D) Cluster IV. Plot scale: 1, <0.4 SDs below the mean; 2, 0.3–0.4 SDs below the mean; 3, 0.2–0.3 SDs below the mean; 4, 0.1–0.2 SDs below the mean; 5, mean to 0.1 SDs below the mean; 6, mean to 0.1 SDs above the mean; 7, 0.1–0.2 SDs above the mean; 8, 0.2–0.3 SDs above the mean; 9, 0.3–0.4 SDs above the mean; 10, >0.4 SDs above the mean. BMI, body mass index; CRP, C-reactive protein; IPAF, interstitial pneumonia with autoimmune features; KL-6, Krebs von den Lungen-6; %FVC, percent predicted forced vital capacity; SP-D, pulmonary surfactant protein-D



**Figure 3**

Overall survival (OS) A) OS was significantly longer in Cluster IV (orange line) than in Clusters I (red line,  $p = 0.53$ ), II (green line,  $p < 0.01$ ), and III (blue line,  $p = 0.01$ ). OS was significantly longer in Cluster I than in Clusters II ( $p < 0.01$ ) and III ( $p = 0.21$ ). OS was significantly longer in Cluster III than in Cluster II ( $p < 0.01$ ). B) Patients with idiopathic pulmonary fibrosis (IPF, continuous line) had significantly longer OS than those with non-specific interstitial pneumonia (NSIP, dotted line,  $p = 0.01$ ), whereas there was no significant difference among the other idiopathic interstitial pneumonia groups. COP, cryptogenic organizing pneumonia (short dashed line); DIP/ RB-ILD, desquamative interstitial pneumonia/respiratory bronchitis-associated interstitial lung disease (middle dashed line); PPFE, pleuroparenchymal fibroelastosis (long dashed line)



**Figure 4**

Cumulative incidence of acute exacerbation A) Clusters II (green line) and III (blue line) had significantly higher cumulative incidence rates of acute exacerbation than Cluster IV (orange line, both  $p = 0.03$ ). B) Patients with idiopathic pulmonary fibrosis (IPF, continuous line) had a significantly higher cumulative incidence rate of acute exacerbations than those with cryptogenic organizing pneumonia (COP, small dashed line,  $p = 0.01$ ) and unclassifiable IIPs (long dashed short dashed line,  $p < 0.01$ ), whereas there was no significant difference among the other idiopathic interstitial pneumonia groups. DIP/ RB-ILD, desquamative interstitial pneumonia/ respiratory bronchitis-associated interstitial lung disease (middle dashed line); NSIP, non-specific interstitial pneumonia (dotted line); PPFE, pleuroparenchymal fibroelastosis (long dashed line)

## Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [SupplementaryTables.docx](#)
- [SupplementaryFigures.docx](#)