

## RESEARCH ARTICLE

# Predicting clinical outcome with phenotypic clusters using quantitative CT fibrosis and emphysema features in patients with idiopathic pulmonary fibrosis

So Hyeon Bak<sup>1,3</sup>, Hye Yun Park<sup>4</sup>, Jin Hyun Nam<sup>5</sup>, Ho Yun Lee<sup>1,2\*</sup>, Jeong Hyun Lee<sup>1</sup>, Insuk Sohn<sup>6</sup>, Man Pyo Chung<sup>4\*</sup>

**1** Department of Radiology and Center for Imaging Science, Samsung Medical Center, Sungkyunkwan University School of Medicine, Seoul, Korea, **2** Department of Health Sciences and Technology, SAIHST, Sungkyunkwan University, Seoul, Korea, **3** Department of Radiology, School of Medicine, Kangwon National University, Chuncheon, Republic of Korea, **4** Division of Pulmonary and Critical Care Medicine, Department of Medicine, Sungkyunkwan University School of Medicine, Seoul, Republic of Korea, **5** Department of Public Health Sciences, Medical University of South Carolina, Charleston, SC, United States of America, **6** Statistics and Data Center, Research Institute for Future Medicine, Samsung Medical Center, Seoul, Republic of Korea

☯ These authors contributed equally to this work.

\* [hoyunlee96@gmail.com](mailto:hoyunlee96@gmail.com) (HYL); [mpchung@skku.edu](mailto:mpchung@skku.edu) (MPC)



## OPEN ACCESS

**Citation:** Bak SH, Park HY, Nam JH, Lee HY, Lee JH, Sohn I, et al. (2019) Predicting clinical outcome with phenotypic clusters using quantitative CT fibrosis and emphysema features in patients with idiopathic pulmonary fibrosis. PLoS ONE 14(4): e0215303. <https://doi.org/10.1371/journal.pone.0215303>

**Editor:** Konstantinos Kostikas, National and Kapodistrian University of Athens, SWITZERLAND

**Received:** October 31, 2018

**Accepted:** March 30, 2019

**Published:** April 18, 2019

**Copyright:** © 2019 Bak et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

**Data Availability Statement:** All relevant data are within the manuscript and its Supporting Information files.

**Funding:** This research (H.Y.L.) was supported by the Korea Health Technology R&D Project through the Korea Health Industry Development Institute (KHIDI), which was funded by the Ministry of Health & Welfare (HI17C0086) and the National Research Foundation of Korea (NRF) grant funded by the Korean government (MSIP; Ministry of

## Abstract

### Background

The clinical course of IPF varies. This study sought to identify phenotyping with quantitative computed tomography (CT) fibrosis and emphysema features using a cluster analysis and to assess prognostic impact among identified clusters in patient with idiopathic pulmonary fibrosis (IPF). Furthermore, we evaluated the impact of fibrosis and emphysema on lung function with development of a descriptive formula.

### Methods

This retrospective study included 205 patients with IPF. A texture-based automated system was used to quantify areas of normal, emphysema, ground-glass opacity, reticulation, consolidation, and honeycombing. Emphysema index was obtained by calculating the percentage of low attenuation area lower than -950HU. We used quantitative CT features and clinical features for clusters and assessed the association with prognosis. A formula was derived using fibrotic score and emphysema index on quantitative CT.

### Results

Three clusters were identified in IPF patients using a quantitative CT score and clinical values. Prognosis was better in cluster1, with a low extent of fibrosis and emphysema with high forced vital capacity (FVC) than cluster2 and cluster3 with higher fibrotic score and emphysema ( $p = 0.046$ , and  $p = 0.026$ ). In the developed formula [ $1.5670$ —fibrotic score(%)  $*0.04737$ —emphysema index $*0.00304$ ], a score greater  $\geq 0$  indicates coexisting of pulmonary fibrosis and emphysema at a significant extent despite of normal spirometric result.

Science, ICT & Future Planning) (No. NRF-2016R1A2B4013046 and NRF-2017M2A2A7A02018568).

**Competing interests:** The authors have declared that no competing interests exist.

## Conclusions

Cluster analysis identified distinct phenotypes, which predicted prognosis of clinical outcome. Formula using quantitative CT values is useful to assess extent of pulmonary fibrosis and emphysema with normal lung function in patients with IPF.

## Introduction

Idiopathic pulmonary fibrosis (IPF) is the most common form of chronic, progressive, interstitial pneumonia with restrictive ventilator dysfunction and reduced gas exchange [1, 2]. The prognosis of IPF is poor overall, with mean survival ranging from 2.5 to 5 years after definite diagnosis [3]. However, the clinical course of IPF varies substantially [4]. Predicting the clinical course or outcome for an individual patient is important but difficult. Clinical variables correlated with survival include age, sex, smoking status, dyspnea, pulmonary functions, digital clubbing, body mass index, and pulmonary hypertension [5]. Computed tomography (CT) plays a central role in the systemic assessment of patients with suspected IPF [6]. The extent of lung fibrosis on CT correlates with disease severity and mortality in interstitial lung disease [4, 7, 8]. Visual scoring of IPF by radiologists is limited by the availability of specialist radiologists and high interobserver variability and is somewhat subjective. There is moderate interobserver agreement among radiologists in identifying honeycombing, which is a diagnostic criterion of IPF [9]. Compared with visual assessment, quantitative analysis of IPF offers an objective, detailed, and reproducible measurement of the extent of IPF [1, 10].

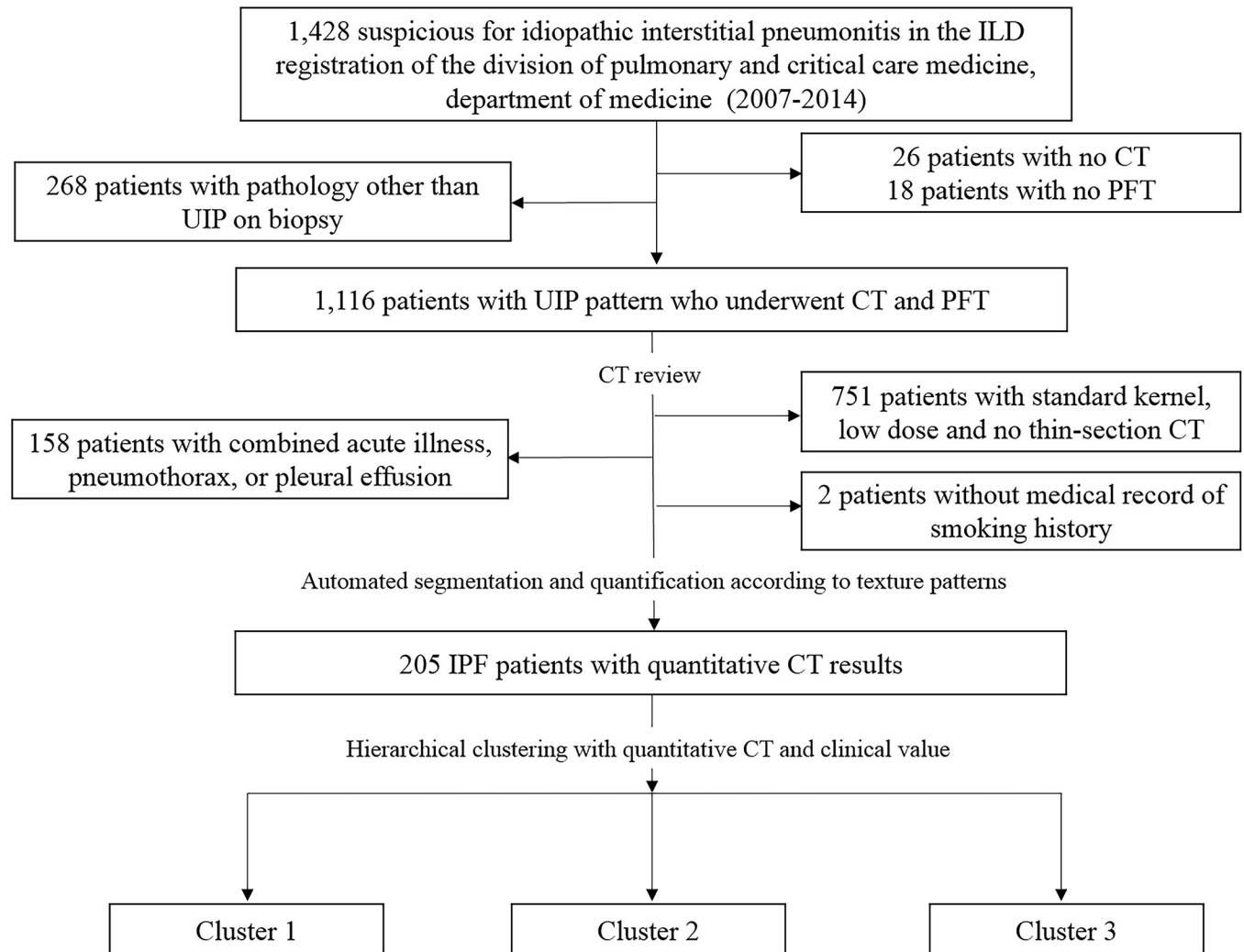
As for spirometry, forced vital capacity (FVC) has been identified as an indicator of disease progression in IPF and is widely used for the routine monitoring of IPF and as a primary endpoint in drug trials [11–13]. Meanwhile, approximately one-third of patients with IPF also have emphysema, and have a normal lung volume as the lung volume increases by 5–10% [11, 13, 14]. Furthermore, higher mortality and less extensive fibrosis have been reported in the patients with IPF combined with emphysema at diagnosis than those with IPF alone [13, 14]. Therefore, when emphysema coexists with pulmonary fibrosis, spirometric values may be misinterpreted as normal.

Cluster analysis has been used to identify homogeneous phenotypic clusters in patients with chronic obstructive pulmonary disease, interstitial lung disease (ILD), and asthma [15–17]. The purpose of our study was two-fold. Firstly, we sought to identify radiologic-based phenotyping with a quantitative CT fibrotic score and emphysema features using a cluster analysis in IPF patients, ultimately reflecting prognostic differences among the identified clusters. Secondly, we evaluated the impact of fibrosis and emphysema on lung function in IPF patients with development of a descriptive formula.

## Materials and methods

### Study population

The Institutional Review Board of our institute approved this retrospective study (approval #2013-09-119), and the requirement for informed consent was waived. From 2007 to 2014, 1117 patients were diagnosed with IPF and underwent CT and pulmonary function test (PFT). The usual interstitial pneumonia (UIP) diagnosis was based on the American Thoracic Society/European Respiratory Society (ATS/ERS) statement with the presence of a UIP pattern on CT in patients not subjected to surgical lung biopsy or specific combinations of CT and surgical lung biopsy patterns in those who received a surgical lung biopsy [2, 18]. Nine hundred nine patients were excluded from our study, including 158 patients with combined acute



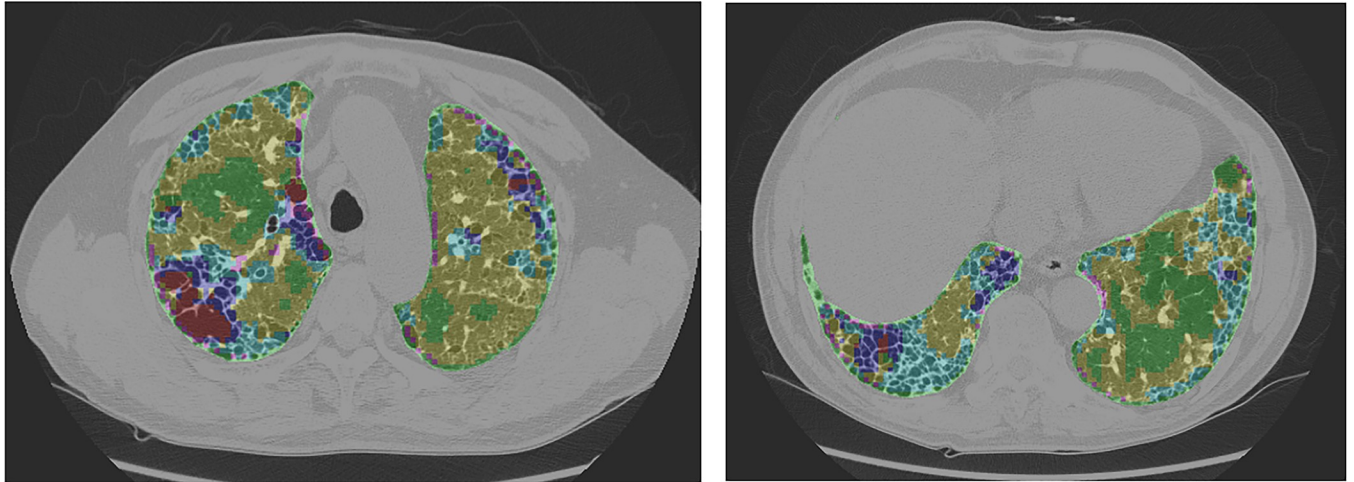
**Fig 1. Flow chart of study design.** UIP = usual interstitial pneumonitis; IPF = idiopathic pulmonary fibrosis; CT = computed tomography.

<https://doi.org/10.1371/journal.pone.0215303.g001>

illness, pneumothorax, or pleural effusion and 751 patients who underwent CT with a standard kernel, low dose, or thick section protocol. We further excluded 2 patients without medical record of smoking history. Final sample included 205 IPF patients with volumetric thin-section CT and PFT (Fig 1).

### Texture-based automated quantification

In-house software was developed for texture-based automated quantification using Visual C++ (Visual Studio 2013, Microsoft) and the Insight Segmentation and Registration Toolkit (ITK, version 4.7) (National Library of Medicine). Whole-lung segmentation was automatically extracted. Based on the texture-based automated system, six regional patterns of normal, reticulation, ground-glass opacity (GGO), consolidation, honeycombing, and emphysema were automatically quantified (Fig 2). Fibrotic score (FS) was defined as the sum of the extent of fibrosis-associated GGO, reticulation, and honeycombing. The methods of texture-based automated quantification have been previously described in detail [10]. Emphysema index was obtained by calculating the percentage of the low attenuation area lower than -950HU.



**Fig 2. Texture-based automated quantification was performed on volumetric thin-section CT in patients with IPF and shows the extent of regional patterns with different colors.** Green = normal, yellow = ground-glass opacity, cyan = reticulation, red = emphysema, pink = consolidation, blue honeycombing. IPF = idiopathic pulmonary fibrosis; CT = computed tomography.

<https://doi.org/10.1371/journal.pone.0215303.g002>

## Statistical analysis

Parametric data are expressed as mean  $\pm$  standard deviation (SD), whereas non-parametric data are expressed as numbers and percentages. Quantitative CT and clinical characteristics including patterns for normal, reticulation, consolidation, GGO, and honeycombing, age, sex, smoking, and PFT were divided into three clusters by Consensus Clustering with the agglomerative hierarchical clustering algorithm using Spearman correlation distance. Differences in clinical and quantitative CT values among the three clusters were tested using Fisher's exact test, Chi-square test, ANOVA test, and Kruskal-Wallis test. Survival curves were estimated using Kaplan-Meier methods, and survival analysis was performed using Cox regression with Bonferroni's correction. Two-sided P-values less than 0.05 were considered significant.

For the formula to evaluate the impact of fibrosis and emphysema on pulmonary function in patients with IPF, we considered several methods, including linear discriminant analysis, logistic regression, cut-off criterion, and support vector machines (SVMs). Ultimately, SVMs were confirmed based on cross-validation (CV) accuracy. SVMs separate the data into different classes using a hyper-plane  $\mathbf{w} \cdot \mathbf{x} + b = 0$  corresponding to the decision function  $f(\mathbf{x}) = \text{sign}(\mathbf{w} \cdot \mathbf{x} + b)$ , which minimizes  $\frac{1}{2} \|\mathbf{w}\|^2 + C \sum_i \xi_i$  subject to  $y_i(\mathbf{w} \cdot \mathbf{x}_i + b) \geq 1 - \xi_i, \xi_i \geq 0$ , where  $\xi_i$  is a slack variable, and  $C$  is a tuning parameter. Here, a category of normal or restrictive was used as response variable  $y$ , and FS (%) and emphysema index were used as  $\mathbf{x}$ . To make a decision function using SVMs, kernel function and tuning parameter  $C$  were needed a priori. Based on 100 repetitions of 5-fold CV, we selected a linear kernel and  $C = 8$ , which maximized CV accuracy. For the linear kernel function, polynomial, sigmoid, and Gaussian RBF were considered, with the range of  $C$  from 0 to 10. To estimate the formula, the R package "kernlab" was used. Data analysis was performed using R 3.4.3 (Vienna, Austria; <http://www.R-project.org/>) and SAS version 9.4 (SAS Institute Inc., Cary, NC).

## Results

### Demographic characteristics

Of the 205 included patients, 78.0% were male, the mean age was 65.8 years (range; 45 to 96 years), 31.2% performed surgical lung biopsy, the mean FVC was 79.0%, and mean diffusing

capacity of the lung carbon monoxide (DLco) was 62.9%. When assessing fibrosis and emphysema on CT with texture-based automated quantification, the mean score was 3,794,216.5 mm<sup>3</sup> for whole lung volume, 44,974.8 mm<sup>3</sup> for honeycombing, 278,098.0 for reticulation, 776,471.6 for GGO, and 44,003.8 for consolidation. The mean emphysema index was 5.6. Of the 205 patients, 17 (8.3%) had an obstructive and combined pattern, 93 (45.4%) had a normal pattern, and 95 (46.3%) had a restrictive pattern in PFT. Characteristics of the 205 patients are summarized in Table 1.

### Clustering analysis

We performed Consensus hierarchical clustering with quantitative CT and clinical features for 205 patients. Consensus clustering analysis was used to divide the patients into three clusters (Fig 3). Subjects grouped into cluster 1 tended to be female and had higher FVC on PFT. The prevalence of reticulation, honeycombing, and emphysema was low, while the prevalence of GGOs was high. Subjects in cluster 2 were predominantly male and had higher FVC on PFT, along with a high prevalence of honeycombing, reticulation, and emphysema, but were the lowest for GGOs on CT. Subjects in cluster 3 had the lowest FVC on PFT; the greatest prevalence of honeycombing, consolidation, and GGOs; and the lowest normal volume on CT. DLco was 64.8% in cluster 1, 52.8% in cluster 2, and 42.3% in cluster 3. Cluster 1 had significantly lower DLco than cluster 2 ( $p = 0.023$ ) and cluster 3 ( $p = 0.005$ ). However, there was no significant difference in DLco between cluster 2 and 3 ( $p = 0.403$ ). Sixty-three of the 176 patients in cluster 1, 0 in 90 patients in cluster 2, and 1 in 9 in cluster 3 were diagnosed as IPF by surgical lung biopsy. A summary of the comparisons among the three clusters is shown in Table 2 and Fig 4.

**Table 1. Characteristics of 205 patients with IPF.**

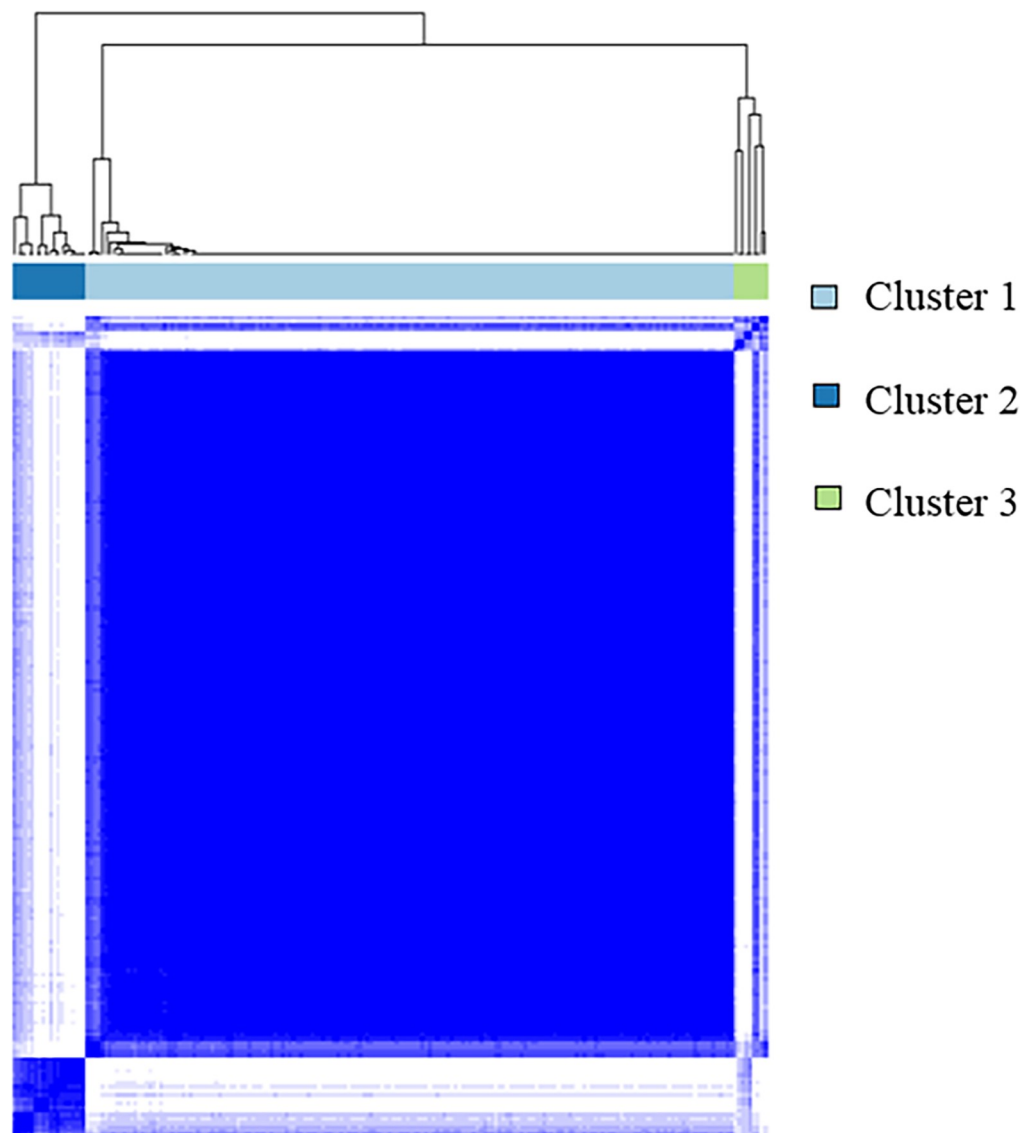
Characteristics	Value
Age	65.8 ± 7.6
Sex (male)*	160 (78.0)
Pulmonary Function Tests	
FVC(%)	79.0 ± 17.6
FEV <sub>1</sub> (%)	89.3 ± 19.3
FEV <sub>1</sub> /FVC ratio	79.9 ± 7.1
DL <sub>CO</sub> (%)	62.9 ± 18.9
TLC (%)	77.6 ± 14.8
CPI	38.5 ± 14.5
Pattern (mm <sup>3</sup> )	
Whole lung volume	3,797,216.5 ± 986,814.5
Normal	2,588,276.5 ± 1,306,646.9 (68.2%)
Honeycombing	44,974.8 ± 86,107.1 (1.2%)
GGO	776,471.6 ± 497,800.3 (20.5%)
Consolidation	44,003.8 ± 23,634.1 (1.2%)
Reticulation	278,098.0 ± 304,913.2 (7.3%)
Emphysema index	5.6 ± 4.6

CPI = composite physiologic index; DL<sub>CO</sub> = diffusing capacity of the lung for carbon monoxide; FEV<sub>1</sub> = forced expiratory volume in 1 second; FVC = forced vital capacity; IPF = idiopathic pulmonary fibrosis; GGO = ground-glass opacity

Data are shown as mean ± standard deviation, and data in parentheses are percentages.

\*Data are numbers of male and data in parentheses are percentages.

<https://doi.org/10.1371/journal.pone.0215303.t001>



**Fig 3. Consensus Cumulative Distribution Function (CDF) Pot using quantitative CT and clinical values.**

<https://doi.org/10.1371/journal.pone.0215303.g003>

### Survival

Survival was evaluated according to IPF classification. There were significant differences in survival among the three clusters by Consensus clustering ( $p = 0.019$ ). Cluster 1, with a low fibrotic score and high FVC, showed good prognosis when compared with cluster 2 ( $p = 0.046$ ) and cluster 3 ( $p = 0.026$ ) with higher fibrotic scores. Cluster 3 had a worse prognosis than cluster 2, but the difference was not significant ( $p = 0.520$ ) (Fig 5).

### Formula based on quantitative CT for assessing impact of fibrosis and emphysema

Of the 205 patients, except for 17 with an obstructive and combined pattern on PFT, 188 patients were divided into two subgroups according to the results of the PFT: restrictive (95

**Table 2. Characteristics of the three clusters according to Consensus clustering using quantitative CT and clinical features.**

Characteristics	Cluster 1 (n = 176)	Cluster 2 (n = 20)	Cluster 3 (n = 9)	P-value	P-value†	P-value‡	P-value§
Sex (male)*	132 (75%)	20 (100%)	8 (88.9%)	0.015	0.026	1	0.931
Age	65.5 ± 7.6	68.6 ± 6.7	64.9 ± 8.2	0.273			
Smoking status				0.247			
Current smoker	15 (8.5)	4 (20)	1 (11.1)				
Ex-smoker	93 (52.8)	12 (60)	4 (44.4)				
Non-smoker	68 (38.7)	4 (20)	4 (44.4)				
Pulmonary Function Tests							
FVC	79.7 ± 17.5	79.2 ± 14.4	63.8 ± 20.9	0.029	1	0.027	0.086
FEV <sub>1</sub>	90.5 ± 19.6	86.6 ± 12.8	72.9 ± 18.9	0.023	1	0.029	0.088
FEV <sub>1</sub> /FVC	80.1 ± 6.5	76.7 ± 9.3	82.7 ± 11.2	0.157			
Pattern (mm <sup>3</sup> )							
Normal	2,613,015.9 ± 1,325,027.6	2,972,241.2 ± 983,977.0	1,251,227 ± 661,978.1	0.003	0.724	0.008	< .0001
Honeycombing	15,657.6 ± 25,565.1	208,361.6 ± 111,838.9	255,208.6 ± 98,461.7	< .0001	< .0001	< .0001	0.688
GGO	807,501.4 ± 495,473.7	326,882.0 ± 235,448.5	1,168,755.0 ± 343,862.6	< .0001	< .0001	0.042	0.688
Consolidation	44,132.5 ± 23,557.5	35,070.8 ± 18,399.7	61,336.6 ± 27,604.0	0.024	0.266	0.129	0.031
Reticulation	245,720.0 ± 259,780.7	483,797.7 ± 538,742.5	454,157.5 ± 203,685.5	0.000	0.029	0.005	0.744
Emphysema index	4.6 ± 3.4	13.4 ± 5.9	9.0 ± 4.1	< .0001	< .0001	0.004	0.121

FEV<sub>1</sub> = forced expiratory volume in 1 second; FVC = forced vital capacity; GGO = ground-glass opacity

Data are shown as mean ± standard deviation

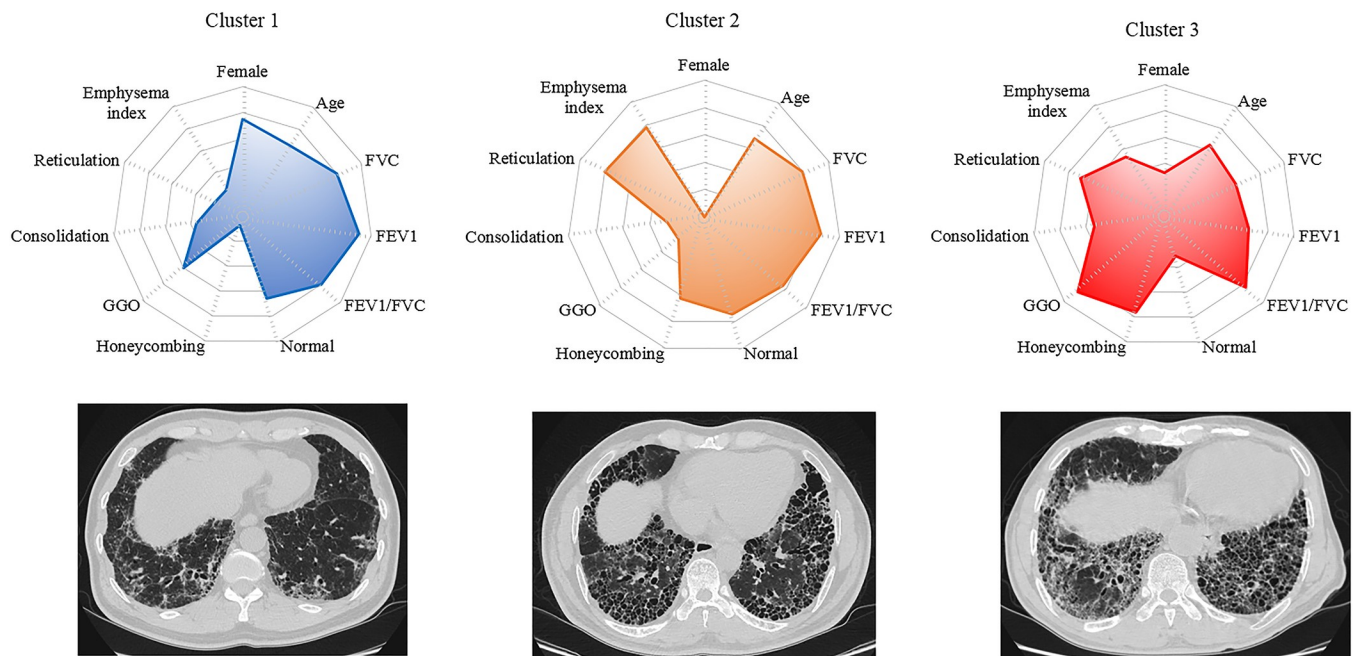
\*Data are numbers of male, and data in parentheses are percentages.

†P-value indicates difference between cluster 1 and cluster 2.

‡ P-value indicates difference between cluster 1 and cluster 3.

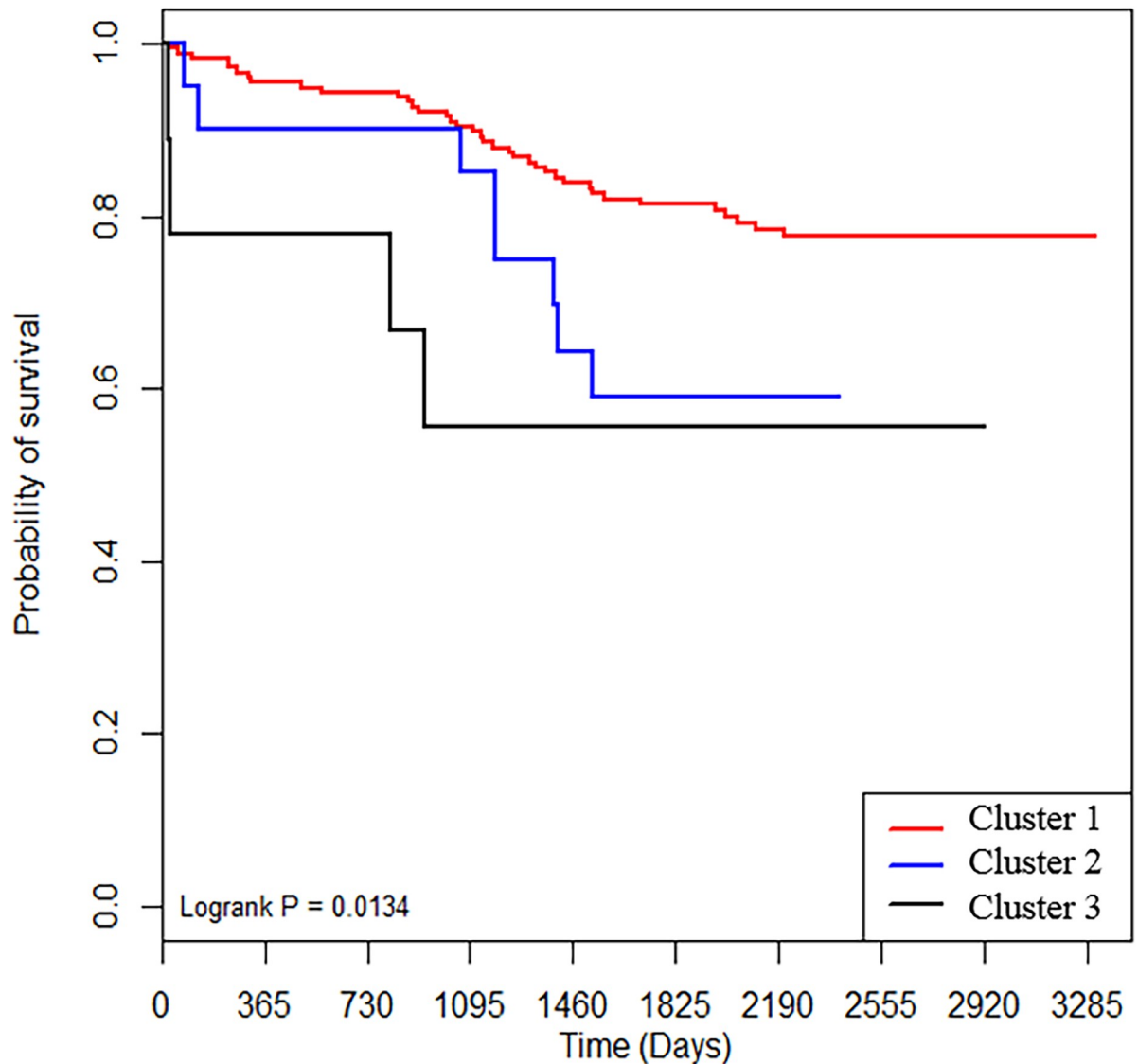
§ P-value indicates difference between cluster 2 and cluster 3.

<https://doi.org/10.1371/journal.pone.0215303.t002>



**Fig 4. Radar plot according to phenotypic clusters (top) and CT images corresponding to each clusters (bottom).**

<https://doi.org/10.1371/journal.pone.0215303.g004>



**Fig 5. Survival in patients with IPF.** Patients were substratified according to Consensus clustering. IPF = idiopathic pulmonary fibrosis.

<https://doi.org/10.1371/journal.pone.0215303.g005>

patients) and normal pattern (93 patients). A formula reflecting spirometric results was derived using FS and emphysema index on quantitative CT in patients with IPF. The FS was different between the restrictive and normal subgroups based on PFT, and the restrictive subgroup had higher FS than those in the normal subgroup (restrictive  $43.4 \pm 20.5$ , normal  $24.3 \pm 17.8$ ;  $p = 0.000$ ). However, emphysema index was not significantly different between the two groups (restrictive  $5.7 \pm 3.9$ , normal  $4.9 \pm 4.7$ ;  $p = 0.210$ ). The formula to assess the extent of FS and emphysema in patients with IPF is as follows:  $1.5670 - \text{FS}(\%) * 0.04737 - \text{emphysema index} * 0.00304$ . A score greater than 0 indicates coexistence of pulmonary fibrosis and emphysema as a significant extent despite of normal spirometric result. The accuracy of the formula which is the probability that the PFT normal or restrictive patients are exactly classified by formular was 0.67, and the true positive rate (TPR) which is the probability that the PFT normal patients are classified as normal was 0.71.



## Discussion

In the present study, quantitative CT features dealing both with fibrotic score and emphysema index were used for clustering of radiologic phenotyping in patients with IPF, yielding three clusters. The radiologic phenotypic subgroups identified using cluster analysis according quantitative CT features differed substantially in survival. Furthermore, we identified that the developed formula could provide important information about extent of fibrosis and emphysema with normal lung function in patients with IPF.

Although IPF is considered an incurable disease with poor prognosis, it has a variety of natural courses, with rapid decline, slower progression, or relative stability [19]. It is difficult to predict prognosis because of the diversity of clinical courses in IPF, and studies have sought to find simple and accurate prognostic factors [3]. Disease extent on CT has been used as a prognostic determinant of IPF. Visually stratifying the extent of honeycombing on CT may help predict prognosis in IPF with emphysema [20]. However, visual scoring of IPF by radiologists shows high interobserver variability and is somewhat subjective. Quantitative CT analysis of lung fibrosis is less time-consuming, more objective, and highly reproducible [1]. Some studies have shown that automated quantification is comparable or superior to visual scoring for evaluating pulmonary function in patients with IPF [21, 22]. A reticular opacity less than 22.05% in extent by automated quantitative assessment of regional CT pattern is related to stability in IPF [10]. Areas of increased lung attenuation via quantitative analysis on CT are associated with increased mortality, and quantitative analysis may be a clinically useful risk stratification tool in patients with interstitial lung disease [23].

The GAP (gender, age, and physiology) model predicts mortality in patients with IPF [24, 25]. Ley et al. [4] identified a CT-GAP model that replaces the DLco with a quantitative CT fibrosis score, with a comparable performance. Based on these studies, a computed-based CT image analysis was performed in patients with IPF, and some cases have already utilized the analysis in clinical practice at specialist centers [1]. In a recent study, stratifying ILD patients with a cluster analysis was predictive of disease progression and survival [15]. Therefore, we carried out a cluster analysis with quantitative features of fibrosis and emphysema on CT, classified the discrete clusters, and then assessed association with prognosis. The first cluster was composed of subjects with high FVC, high GGO, and low honeycombing and emphysema. Subjects with severe honeycombing, reticulation, and emphysema were predominantly grouped into the latter two clusters.

We found that prognosis was the best in cluster 1, with a low extent of fibrosis with emphysema and high FVC. Although they were not statistically different, cluster 2 tended to have a better prognosis than cluster 3. In cluster 2, the degree of honeycombing, reticulation, and emphysema was not significantly different from cluster 3, whereas the extent of GGO was higher. In many studies, honeycombing and reticulation are defined as a CT fibrosis score, and prognosis is worse with an increasing extent of honeycombing and reticulation [4, 10, 20]. Our study found that the extent of GGO in patients with severe fibrosis may affect prognosis. In IPF patients, GGOs should be divided into two categories including acute inflammation and chronic fibrosis. GGOs are admixed with reticulation and honeycombing, and these GGOs should be regarded as fibrotic process [6]. However, extensive pure GGOs in the non-fibrotic areas of the lung suggests acute exacerbation or infection. Our study shows active inflammation in cluster 1 and GGOs associated with reticulation in cluster 3. Furthermore, Moon et al. [26] showed that GGO ratio correlates with survival in patients with diffuse ILD. Therefore, the increased extent of GGO associated with acute inflammation or fine fibrosis has an adverse effect on patient prognosis.

PFT is commonly used as a biomarker associated with prognosis. FVC with high precision and reproducibility is linked to disease progression and mortality in IPF, and FVC decreases

by an average of 150–250 mL per year in untreated IPF patients [27]. However, we found that FVC was similar between cluster 2 with severe fibrosis and cluster 1 with low fibrosis. The two clusters were significantly different in the extent of emphysema as well as fibrosis. Cluster 2 had the highest emphysema index. Approximately one-third of patients with IPF also have emphysema [14]. The presence of emphysema has been associated with mortality in patients with IPF; however, the results are inconsistent among studies [14, 28, 29]. Patients with IPF and emphysema show preservation of spirometric values and lung volumes despite extensive lung disease on CT and marked impairment of gas exchange, because emphysema leads to normalized lung volume as the compliance increased by emphysema compensates physically for the decreased compliance caused by fibrosis [11, 30]. A previous study showed that FVC is negatively correlated with extent of fibrosis, but positively correlated with emphysema [10]. According to Cottin et al. [31], FVC measurement is not appropriate for monitoring disease progression in IPF patients with emphysema extent greater than or equal to 15% on CT. Therefore, FVC does not accurately reflect disease in patients with IPF and should not be considered as an endpoint.

To overcome limitation of FVC in patients with IPF and emphysema, we developed a formula to encapsulate both fibrosis and emphysema based on the results of automated quantitative assessment of regional CT patterns:  $1.5670 - \text{FS}(\%) * 0.04737 - \text{emphysema index} * 0.00304$ . The accuracy and TPR of the formula were 0.67 and 0.71, respectively, and this means that the spirometric results of about 71% patients with IPF and PFT normal pattern which have the score greater than 0 can be interpreted as normal due to coexistence of pulmonary fibrosis and emphysema as significant extent. However, in case of the patient with IPF which has the score greater 0 with severe fibrosis and high emphysema index, there are two cases; 1) the patient with IPF and PFT normal pattern, 2) the patient with IPF and PFT restrictive pattern. Both cases mean that the spirometric result is misinterpreted as normal. Thus based on the application of our formula, PFT results can be more cautiously interpreted while considering fibrosis and emphysema on CT in patients with IPF.

This study has several limitations. First, chest CT scans were not obtained using the same CT scanner and protocol because of the retrospective study design. Texture-based quantification can be influenced by dose, CT values, and reconstruction algorithms. We excluded 751 patients with standard kernel, low dose, or no thin-section protocols to minimize these effects. Second, in Consensus clustering, the numbers of patients in clusters 2 and 3 were small, and cluster 1 with mild fibrosis contained the majority. These small numbers were due to the exclusion of many patients with IPF in order to minimize the effect of differences in CT quantification, resulting in a study population of only 205 patients. Third, complications such as pulmonary hypertension may affect the survival of patients with IPF. Therefore, the quantitative CT features and cluster analysis we conducted may not represent mortality due to other complications. Fourth, this investigation was conducted at a single center, and so our results are not conclusive and require validation by prospective studies in a larger cohort.

## Conclusion

In conclusion, quantitative CT features related to both fibrosis and emphysema were used to classify cases into three discrete subgroups with unique radiologic phenotypes and different survival. Prognosis worsens as the extent of quantitative CT fibrosis increases, and the extent of GGO in patients with severe fibrosis may affect prognosis. In addition, cluster analysis showed that emphysema could affect FVC. The relative extent of quantitative emphysema and fibrosis on CT could help to predict lung function based on the developed formula.

## Supporting information

**S1 Dataset. Dataset of patients with IPF.**  
(XLSX)

## Acknowledgments

The authors would like to thank Seoyeon Jeon and Sook Lee, who assisted with quantitative CT analysis.

## Author Contributions

**Conceptualization:** So Hyeon Bak, Hye Yun Park, Jin Hyun Nam, Ho Yun Lee, Man Pyo Chung.

**Data curation:** So Hyeon Bak, Hye Yun Park, Jin Hyun Nam, Ho Yun Lee, Jeong Hyun Lee, Man Pyo Chung.

**Formal analysis:** So Hyeon Bak, Hye Yun Park, Jin Hyun Nam, Ho Yun Lee, Insuk Sohn, Man Pyo Chung.

**Funding acquisition:** Ho Yun Lee.

**Methodology:** So Hyeon Bak, Hye Yun Park, Jin Hyun Nam, Ho Yun Lee, Jeong Hyun Lee, Insuk Sohn, Man Pyo Chung.

**Supervision:** So Hyeon Bak, Hye Yun Park, Jin Hyun Nam, Ho Yun Lee, Jeong Hyun Lee, Insuk Sohn, Man Pyo Chung.

**Validation:** Hye Yun Park.

**Visualization:** So Hyeon Bak, Hye Yun Park, Jin Hyun Nam, Ho Yun Lee, Man Pyo Chung.

**Writing – original draft:** So Hyeon Bak, Hye Yun Park, Jin Hyun Nam, Ho Yun Lee, Insuk Sohn, Man Pyo Chung.

**Writing – review & editing:** So Hyeon Bak, Hye Yun Park, Jin Hyun Nam, Ho Yun Lee, Jeong Hyun Lee, Insuk Sohn, Man Pyo Chung.

## References

1. Ohkubo H, Nakagawa H, Niimi A. Computer-based quantitative computed tomography image analysis in idiopathic pulmonary fibrosis: A mini review. *Respiratory investigation*. 2018; 56(1):5–13. Epub 2018/01/13. <https://doi.org/10.1016/j.resinv.2017.10.003> PMID: 29325682.
2. Raghu G, Collard HR, Egan JJ, Martinez FJ, Behr J, Brown KK, et al. An official ATS/ERS/JRS/ALAT statement: idiopathic pulmonary fibrosis: evidence-based guidelines for diagnosis and management. *American journal of respiratory and critical care medicine*. 2011; 183(6):788–824. Epub 2011/04/08. <https://doi.org/10.1164/rccm.2009-040GL> PMID: 21471066; PubMed Central PMCID: PMC5450933.
3. Fujimoto H, Kobayashi T, Azuma A. Idiopathic Pulmonary Fibrosis: Treatment and Prognosis. *Clinical medicine insights Circulatory, respiratory and pulmonary medicine*. 2015; 9(Suppl 1):179–85. Epub 2016/12/17. <https://doi.org/10.4137/CCRP.M.S23321> PMID: 27980445; PubMed Central PMCID: PMC5147432.
4. Ley B, Elicker BM, Hartman TE, Ryerson CJ, Vittinghoff E, Ryu JH, et al. Idiopathic pulmonary fibrosis: CT and risk of death. *Radiology*. 2014; 273(2):570–9. Epub 2014/06/14. <https://doi.org/10.1148/radiol.14130216> PMID: 24927326; PubMed Central PMCID: PMC54334234.
5. Ley B, Collard HR, King TE Jr. Clinical course and prediction of survival in idiopathic pulmonary fibrosis. *American journal of respiratory and critical care medicine*. 2011; 183(4):431–40. Epub 2010/10/12. <https://doi.org/10.1164/rccm.201006-0894CI> PMID: 20935110.

6. Lynch DA, Sverzellati N, Travis WD, Brown KK, Colby TV, Galvin JR, et al. Diagnostic criteria for idiopathic pulmonary fibrosis: a Fleischner Society White Paper. *The Lancet Respiratory medicine*. 2018; 6(2):138–53. Epub 2017/11/21. [https://doi.org/10.1016/S2213-2600\(17\)30433-2](https://doi.org/10.1016/S2213-2600(17)30433-2) PMID: 29154106.
7. Best AC, Meng J, Lynch AM, Bozic CM, Miller D, Grunwald GK, et al. Idiopathic pulmonary fibrosis: physiologic tests, quantitative CT indexes, and CT visual scores as predictors of mortality. *Radiology*. 2008; 246(3):935–40. Epub 2008/02/01. <https://doi.org/10.1148/radiol.2463062200> PMID: 18235106.
8. Hansell DM, Goldin JG, King TE Jr., Lynch DA, Richeldi L, Wells AU. CT staging and monitoring of fibrotic interstitial lung diseases in clinical practice and treatment trials: a position paper from the Fleischner Society. *The Lancet Respiratory medicine*. 2015; 3(6):483–96. Epub 2015/05/16. [https://doi.org/10.1016/S2213-2600\(15\)00096-X](https://doi.org/10.1016/S2213-2600(15)00096-X) PMID: 25975761.
9. Walsh SL, Calandriello L, Sverzellati N, Wells AU, Hansell DM. Interobserver agreement for the ATS/ERS/JRS/ALAT criteria for a UIP pattern on CT. *Thorax*. 2016; 71(1):45–51. Epub 2015/11/21. <https://doi.org/10.1136/thoraxjnl-2015-207252> PMID: 26585524.
10. Park HJ, Lee SM, Song JW, Lee SM, Oh SY, Kim N, et al. Texture-Based Automated Quantitative Assessment of Regional Patterns on Initial CT in Patients With Idiopathic Pulmonary Fibrosis: Relationship to Decline in Forced Vital Capacity. *AJR American journal of roentgenology*. 2016; 207(5):976–83. Epub 2016/10/22. <https://doi.org/10.2214/AJR.16.16054> PMID: 27533069.
11. Behr J. Disease Progression in Idiopathic Pulmonary Fibrosis. FVC Is Not Enough. *American journal of respiratory and critical care medicine*. 2017; 196(9):1094–5. Epub 2017/07/13. <https://doi.org/10.1164/rccm.201706-1246ED> PMID: 28700255.
12. Collard HR, King TE Jr., Bartelson BB, Vourlekis JS, Schwarz MI, Brown KK. Changes in clinical and physiologic variables predict survival in idiopathic pulmonary fibrosis. *American journal of respiratory and critical care medicine*. 2003; 168(5):538–42. Epub 2003/05/30. <https://doi.org/10.1164/rccm.200211-1311OC> PMID: 12773325.
13. Cottin V. Combined pulmonary fibrosis and emphysema: bad and ugly all the same? *The European respiratory journal*. 2017; 50(1). Epub 2017/07/07. <https://doi.org/10.1183/13993003.00846-2017> PMID: 28679618.
14. Mejia M, Carrillo G, Rojas-Serrano J, Estrada A, Suarez T, Alonso D, et al. Idiopathic pulmonary fibrosis and emphysema: decreased survival associated with severe pulmonary arterial hypertension. *Chest*. 2009; 136(1):10–5. Epub 2009/02/20. <https://doi.org/10.1378/chest.08-2306> PMID: 19225068.
15. Adegunsoye A, Oldham JM, Chung JH, Montner SM, Lee C, Witt LJ, et al. Phenotypic Clusters Predict Outcomes in a Longitudinal Interstitial Lung Disease Cohort. *Chest*. 2018; 153(2):349–60. Epub 2017/10/02. <https://doi.org/10.1016/j.chest.2017.09.026> PMID: 28964798; PubMed Central PMCID: PMC5815877.
16. Choi S, Hoffman EA, Wenzel SE, Castro M, Fain S, Jarjour N, et al. Quantitative computed tomographic imaging-based clustering differentiates asthmatic subgroups with distinctive clinical phenotypes. *The Journal of allergy and clinical immunology*. 2017; 140(3):690–700.e8. Epub 2017/02/02. <https://doi.org/10.1016/j.jaci.2016.11.053> PMID: 28143694; PubMed Central PMCID: PMC5534190.
17. Rennard SI, Locantore N, Delafont B, Tal-Singer R, Silverman EK, Vestbo J, et al. Identification of five chronic obstructive pulmonary disease subgroups with different prognoses in the ECLIPSE cohort using cluster analysis. *Annals of the American Thoracic Society*. 2015; 12(3):303–12. Epub 2015/02/03. <https://doi.org/10.1513/AnnalsATS.201403-125OC> PMID: 25642832.
18. Travis WD, Costabel U, Hansell DM, King TE Jr., Lynch DA, Nicholson AG, et al. An official American Thoracic Society/European Respiratory Society statement: Update of the international multidisciplinary classification of the idiopathic interstitial pneumonias. *American journal of respiratory and critical care medicine*. 2013; 188(6):733–48. Epub 2013/09/17. <https://doi.org/10.1164/rccm.201308-1483ST> PMID: 24032382; PubMed Central PMCID: PMC5803655.
19. Kim HJ, Perlman D, Tomic R. Natural history of idiopathic pulmonary fibrosis. *Respiratory medicine*. 2015; 109(6):661–70. Epub 2015/03/03. <https://doi.org/10.1016/j.rmed.2015.02.002> PMID: 25727856.
20. Kim YS, Jin GY, Chae KJ, Han YM, Chon SB, Lee YS, et al. Visually stratified CT honeycombing as a survival predictor in combined pulmonary fibrosis and emphysema. *The British journal of radiology*. 2015; 88(1055):20150545. Epub 2015/09/22. <https://doi.org/10.1259/bjr.20150545> PMID: 26388110; PubMed Central PMCID: PMC58474367.
21. Jacob J, Bartholmai BJ, Rajagopalan S, Kokosi M, Maher TM, Nair A, et al. Functional and prognostic effects when emphysema complicates idiopathic pulmonary fibrosis. *The European respiratory journal*. 2017; 50(1). Epub 2017/07/07. <https://doi.org/10.1183/13993003.00379-2017> PMID: 28679612.
22. Yoon RG, Seo JB, Kim N, Lee HJ, Lee SM, Lee YK, et al. Quantitative assessment of change in regional disease patterns on serial HRCT of fibrotic interstitial pneumonia with texture-based automated quantification system. *European radiology*. 2013; 23(3):692–701. Epub 2012/08/25. <https://doi.org/10.1007/s00330-012-2634-8> PMID: 22918563.

23. Podolanczuk AJ, Oelsner EC, Barr RG, Bernstein EJ, Hoffman EA, Easthausen IJ, et al. High-Attenuation Areas on Chest Computed Tomography and Clinical Respiratory Outcomes in Community-Dwelling Adults. *American journal of respiratory and critical care medicine*. 2017; 196(11):1434–42. Epub 2017/06/15. <https://doi.org/10.1164/rccm.201703-0555OC> PMID: 28613921; PubMed Central PMCID: PMC5736977.
24. Kolb M, Collard HR. Staging of idiopathic pulmonary fibrosis: past, present and future. *European respiratory review: an official journal of the European Respiratory Society*. 2014; 23(132):220–4. Epub 2014/06/02. <https://doi.org/10.1183/09059180.00002114> PMID: 24881076.
25. Ley B, Ryerson CJ, Vittinghoff E, Ryu JH, Tomassetti S, Lee JS, et al. A multidimensional index and staging system for idiopathic pulmonary fibrosis. *Annals of internal medicine*. 2012; 156(10):684–91. Epub 2012/05/16. <https://doi.org/10.7326/0003-4819-156-10-201205150-00004> PMID: 22586007.
26. Moon JW, Bae JP, Lee HY, Kim N, Chung MP, Park HY, et al. Perfusion- and pattern-based quantitative CT indexes using contrast-enhanced dual-energy computed tomography in diffuse interstitial lung disease: relationships with physiologic impairment and prediction of prognosis. *European radiology*. 2016; 26(5):1368–77. Epub 2015/08/09. <https://doi.org/10.1007/s00330-015-3946-2> PMID: 26253261.
27. Behr J, Bonella F, Bonnet R, Glaser S, Grohe C, Gunther A, et al. [Position Paper: Significance of the Forced Vital Capacity in Idiopathic Pulmonary Fibrosis]. *Pneumologie (Stuttgart, Germany)*. 2015; 69(8):455–8. Epub 2015/08/01. <https://doi.org/10.1055/s-0034-1392602> PMID: 26227628.
28. Cottin V, Nunes H, Brillet PY, Delaval P, Devouassoux G, Tillie-Leblond I, et al. Combined pulmonary fibrosis and emphysema: a distinct underrecognised entity. *The European respiratory journal*. 2005; 26(4):586–93. Epub 2005/10/06. <https://doi.org/10.1183/09031936.05.00021005> PMID: 16204587.
29. Kurashima K, Takayanagi N, Tsuchiya N, Kanauchi T, Ueda M, Hoshi T, et al. The effect of emphysema on lung function and survival in patients with idiopathic pulmonary fibrosis. *Respirology (Carlton, Vic)*. 2010; 15(5):843–8. Epub 2010/06/16. <https://doi.org/10.1111/j.1440-1843.2010.01778.x> PMID: 20546187.
30. Jankowich MD, Rounds SIS. Combined pulmonary fibrosis and emphysema syndrome: a review. *Chest*. 2012; 141(1):222–31. Epub 2012/01/05. <https://doi.org/10.1378/chest.11-1062> PMID: 22215830; PubMed Central PMCID: PMC3251269.
31. Cottin V, Hansell DM, Sverzellati N, Weycker D, Antoniou KM, Atwood M, et al. Effect of Emphysema Extent on Serial Lung Function in Patients with Idiopathic Pulmonary Fibrosis. *American journal of respiratory and critical care medicine*. 2017; 196(9):1162–71. Epub 2017/06/29. <https://doi.org/10.1164/rccm.201612-2492OC> PMID: 28657784.