

# Development and Validation of a Comprehensive Multiparameter-based Scoring System to Assess Pulmonary Fibrosis Severity

**Chengsheng Yin**

Department of Respiratory, Shanghai Pulmonary Hospital, Tongji University, School of Medicine, Shanghai, China

**Yuan Zhang**

Department of respiratory, Shanghai Pulmonary Hospital, Tongji University, School of Medicine

**Yiliang Su**

Department of Respiratory, Shanghai Pulmonary Hospital, Tongji University, School of Medicine, Shanghai, China

**Feng Zhang**

Department of Respiratory, Shanghai Pulmonary Hospital, Tongji University, School of Medicine, Shanghai, China

**Jingyun Shi**

Department of Radiology, Shanghai Pulmonary Hospital, Tongji University, School of Medicine, Shanghai, China

**Yanan Chen**

Department of Radiology, Shanghai Pulmonary Hospital, Tongji University, School of Medicine, Shanghai, China

**Dong Weng**

Department of Respiratory, Shanghai Pulmonary Hospital, Tongji University, School of Medicine, Shanghai, China

**Tao Chen**

Department of Respiratory, Shanghai Pulmonary Hospital, Tongji University, School of Medicine, Shanghai, China

**Mengmeng Zhao**

Department of Respiratory, Shanghai Pulmonary Hospital, Tongji University, School of Medicine, Shanghai, China

**Ying Zhou**

Department of Respiratory, Shanghai Pulmonary Hospital, Tongji University, School of Medicine, Shanghai, China

**Xing Jiang**

Department of Respiratory, Shanghai Pulmonary Hospital, Tongji University, School of Medicine, Shanghai, China

**Aihong Zhang**

Department of Medical Statistics, Tongji University, School of Medicine, Shanghai 200092, P.R. China

**Huiping Li (✉ [liw2013@126.com](mailto:liw2013@126.com))**

Tongji University Affiliated Shanghai Pulmonary Hospital

---

**Research**

**Keywords:** Idiopathic pulmonary fibrosis, pulmonary fibrosis stage, disease severity grade, death risk, prediction model

**Posted Date:** February 26th, 2020

**DOI:** <https://doi.org/10.21203/rs.2.24588/v1>

**License:**  This work is licensed under a Creative Commons Attribution 4.0 International License. [Read Full License](#)

---

## Abstract

**Background** Survival time varies greatly in patients with idiopathic pulmonary fibrosis (IPF). An assessment method that can accurately assess the severity and prognosis of idiopathic pulmonary fibrosis is currently lacking. This study aimed to develop a new method, which can be easily used to assess pulmonary fibrosis severity.

**Method** 1. Development of a HRCT combined pulmonary function & physiological parameter (CTPF) assessment method: The method included two parts. 1) CT-based fibrosis staging: Four representative lung CT sections were selected and evenly divided into 100 small areas. The percentage of honeycomb lesion area in the four sections was determined fibrosis stage, 2) PF-based severity grade: FVC%pred, DLco%pred, SpO<sub>2</sub> age and gender were used to assess PF severity grade. 2. Validation of the new method: The method was used to assess 192 patients with IPF. Two radiologists used the CT-based fibrosis staging method to determine the fibrosis stage. Pulmonologist determined the PF severity grade. 3. Statistical analyses: Intra-group correlation coefficient to estimate the consistency between the CT scores from the two radiologists. Spearman correlation coefficient to evaluate the correlation between CT scores and lung function parameters. The competitive risk Fine–Gray model was used to analyze the relationship between CT-based stage/PF-based grade and prognosis. CT-based stage, PF-based grade, and GAP stage were used as predictors to predicted the death risk.

**Results** 1. The intra-group correlation coefficient of the CT scores of the two radiologists was 0.95,  $P < 0.05$ . 2. The CT scores negatively correlated with pulmonary function. 3. The CTPF comprehensive model, showed higher predictive accuracy.

**Conclusion** Combined CT-based staging and PF-based grading methods CTPF can be adopted easily in clinical practice, and can assess IPF severity and predict death risk more accurately.

## Background

Survival time varies greatly in patients with idiopathic pulmonary fibrosis (IPF). Some patients have slow disease progression and are stable for a long time, whereas others develop acute exacerbation and die quickly [1, 2]. How to accurately assess IPF severity and predict prognosis remains unanswered. A commonly accepted method that can provide accurate assessment for IPF severity and prognosis is currently still lacking [3].

The currently available IPF severity scoring methods include: 1) the clinical-radiographic-physiologic (CRP) scoring method published in 1986 by Leslie C. Watters et al [4, 5]. The CRP method uses the following 7 variables: degree of dyspnea, score of chest x-ray results, forced vital capacity (FVC), forced expiratory volume in one second (FEV<sub>1</sub>), intrathoracic gas volume (V<sub>tg</sub>), the ratio of diffusing capacity of the lung for Carbon Monoxide (DLco) to alveolar volume (VA) (DLco/VA), and resting alveolar-arterial oxygen partial pressure (AaPO<sub>2</sub>). This method has several disadvantages, such as requiring many parameters, using a complex calculation method, and using Chest x-ray, which often fails to reveal lung lesions such as fibrosis lesion. In 2001, Talmadge E. King et al. [6] improved the CRP scoring method by including additional parameters, such as gender, age, smoking status, and clubbing fingers, which further increases the complexity of this assessment method. 2) In 2002, Athol U. Wells et al [7] proposed a composite physiologic index (CPI) method to assess interstitial lung disease (ILD) severity by combining chest computed tomography (CT) results and pulmonary function parameters. They verified that the CPI method is superior in terms of estimating fibrosis severity and predicting survival to the method that only uses pulmonary function parameters. However, the calculation formula of the CPI method is complex, which limits its adoption in clinical practice. 3) Brett Ley, MD et al. [8] suggested a gender, age, and physiologic (GAP)-based method, which is based on the data of gender, age, FVC, and DLco. However, the GAP method does not include chest HRCT data, which is critical in IPF assessment. Thus the assessment accuracy of the GAP method is compromised. 4) Japanese researcher Ryo Okuda et al [9] proposed to use the two key arterial blood gas indicators, arterial partial pressure of oxygen (PaO<sub>2</sub>) and oxyhemoglobin saturation (SaO<sub>2</sub>%), to assess IPF severity. However, this method does not include HRCT data and pulmonary function parameters. In 2017, Hasti Robbie et al [10] analyzed the contributions of physiological parameters, histopathological parameters, imaging parameters, biomarkers to the assessment of IPF severity and concluded that using a single type of parameters to assess IPF severity has serious limitations.

In this study, based on the available scoring methods, we chose parameters that have been proven to have a good prognostic value and can be acquired easily in clinical practice to develop a new scoring method to assess pulmonary fibrosis severity (patent application number: 201910514972.5). In this method, we combined the data of HRCT, pulmonary function, and arterial oxygen saturation to estimate pulmonary fibrosis severity comprehensively, and we also validated the new scoring method ([www.clinicaltrials.gov](http://www.clinicaltrials.gov): ChiCTR-RRC-17010683). The new method may improve the accuracy of pulmonary fibrosis severity assessment and IPF prognosis prediction. This method is simple and can be adopted easily in clinical practice.

## Method

### Development of a New Scoring Method

### Pulmonary Fibrosis Staging by Chest High-resolution Computed Tomography (HRCT) (CT-based fibrosis staging, Fig. 1)

Based on the latest 2018 IPF guidelines [1], the common pathological manifestation of IPF is usual interstitial pneumonia (UIP). The imaging presentation of IPF is characterized by reticular infiltration shadow, linear shadow, honeycomb lung, and traction bronchiectasis. The severity and area of the lesions showing on chest CT images are important predictor of IPF mortality [11]. Previous imaging studies on interstitial pneumonia and IPF have proposed that honeycomb and stretch bronchiectasis can better predict the survival and prognosis of patients than other imaging characteristics such as reticular infiltration shadow

and linear shadow. Moreover, honeycomb and stretch bronchiectasis are the most representative imaging manifestations of pulmonary fibrosis [12–14]. Therefore, we choose the pathological range of honeycomb and traction bronchiectasis to evaluate the extent of pulmonary fibrosis.

We combined the theories proposed in the previous studies [15, 16] and calculus principles to design a "four-section honeycomb lung percentage" method. We selected the following four representative lung CT sections to semi-quantitatively estimate the extent of honeycomb lesion in the entire lung: the aortic arch section, the tracheal bifurcation section, the section of basal (dorsal) segment of the tracheal bifurcation at the inferior lobes, and the section below the right lung apex. Each section included both the left and the right lungs. The largest transverse diameter line of each lung section was evenly divided into three parts, and then the lung section was divided into inner, middle, and outer sections by drawing lines starting from the dividing points along the shape of patient's thorax. The outer lung section was then evenly divided into 6 small areas; the middle lung Sect. 4–5 areas (5 areas for a large middle section); the inner lung Sect. 2 areas. Thus, in total, the 4 CT sections comprised 8 lung sections (4 left + 4 right lung) and were evenly divided into approximately 100 small areas (12 or 13 per lung section × 8). Each small area was scored as 1 when there was positive honeycomb lesion in the area, and the total score of the entire lung was used as the total honeycomb lung score. The traction bronchiectasis score was calculated in the same way as the honeycomb score. The honeycomb lung percentage was calculated as: (total honeycomb score + total traction bronchiectasis score) ÷ total number of the small areas × 100%. For example, if 8 lung sections were evenly divided into 100 small areas and 30 of them were scored as positive honeycomb lung or traction bronchiectasis, then the honeycomb lung percentage was 30%. According to Lynch et al [17], lung fibrosis can be staged based on the following lung CT characteristics: stage I: there is reticular and linear shadow but no honeycomb lesion; stage II: honeycomb lesion area is < 25% of the entire lung; stage III: honeycomb lesion area is 25%-49%; stage IV: honeycomb lesion area is 50%-75%; stage V: honeycomb lesion area is > 75%.

The thickness of HRCT section was 1-1.5 mm; section spacing was 2 cm. Patients were in supine or prone position. The minimum exposure was 200 mA per second. The definition of honeycomb lung followed the criteria recommended by the Fleischner Society Guidelines [11].

## **Assess Pulmonary Fibrosis Severity by Using Multi-parameter-based Comprehensive Scoring Method**

Patients' baseline physiological condition and lung function parameters are important predictors for survival [7–9, 18]. We compared the advantages and disadvantages of the existing pulmonary fibrosis severity scoring methods, including CRP, GAP, CPI, and JRS (Table 1) and chose the 5 parameters that are of important predictive values and are relatively easy to be collected in clinical practice: FVC%pred, DLco%pred, oxygen saturation of peripheral blood (SpO<sub>2</sub>%), age, and gender. We used the 5 parameters to evaluate the disease severity. We followed the previous studies [4–9] to define a multiparameter-based (parameters of pulmonary function and physiological condition, PF-based grading) comprehensive scoring criteria to estimate disease severity. We then combined this PF-based grading method with the CT-based pulmonary fibrosis staging method to develop a new scoring method (CTPF) to assess pulmonary fibrosis severity (Table 2).

Table 1  
Comparison of Different Pulmonary Staging Methods

Scoring Method	Parameters												
	gender	age	FVC%	DLco%	TLC%	FEV <sub>1</sub> %	Lung capacity (Vtg)	HRCT	X-ray	PaO <sub>2</sub>	SpO <sub>2</sub> %	Smoking	C
GAP	√	√	√	√									
CPI			√	√		√							
CRP					√		√		√	√		√	
JRS										√	√		
Accessibility	Easy	Easy	Easy	Easy	Easy	Easy	Require a comprehensive device to measure lung function	Easy	Easy but images overlap	Require arterial blood	Easy	Difficult to quantification	
Importance	Y	Y	Y	Y	Affected by multiple factors	Correlate to airway disease	?	Y	Y	Y	Y	?	?
Parameters in our method	√	√	√	√				√			√		
Notes: Y: The parameter is important. ?: The importance of the parameter is currently unknown. √: The parameter was included in the model of this study.													
SpO <sub>2</sub> %; oxygen saturation of peripheral blood. SpO <sub>2</sub> is the resting arterial oxygen saturation measured at fingertips.													
FVC: forced vital capacity. FVC% pred: the percentage of the actual FVC over the predicted FVC. FEV <sub>1</sub> : forced expiratory volume in one second. FEV <sub>1</sub> % pred: the percentage of the actual FEV <sub>1</sub> over the predicted FEV <sub>1</sub> . DLco: diffusing capacity of the lung for carbon monoxide. DLco% pred: the percentage of the actual DLco over the predicted DLco. FEV <sub>1</sub> /FVC%: the percentage of the actual FEV <sub>1</sub> over the predicted FEV <sub>1</sub> .													
GAP (gender, age, and physiologic variables) stage followed the recommendation by Brett Ley, and a higher stage represented a greater death risk.													
CPI: composite physiologic index. In 2002, Athol U. Wells and colleagues proposed to use CPI, which combined chest CT and pulmonary functional parameters (ILDs). A higher CPI represents a more severe ILD.													
CRP: clinical-radiographic-physiologic. Leslie C. Watters et al published the CRP system in 1986.													
JRS: Ryo Okuda et al proposed the IPF staging method in 2004.													
HRCT: high-resolution computed tomography.													

Table 2  
Criteria for CT-based Pulmonary Fibrosis Staging and PF-based Severity Grading

PF scoring criteria	SpO <sub>2</sub> %		FVC% pred		DLco% pred				Age (year)			Gender		Total severity score	Criteria 1		
	≥ 95%	90%-94%	≤ 89%	> 75%	50%-75%	< 50%	> 55%	36%-55%	< 36%	Cannot complete	≤ 60	61-65	> 65			M	F
	0	1	2	0	1	2	0	1	2	3	0	1	2	1	0		0-3
CT-based staging criteria	I	there is reticular and linear shadow but no honeycomb lesion;															
	II	honeycomb lesion area is < 25% of the entire lung;															
	III	honeycomb lesion area is 25%-49%															
	IV	honeycomb lesion area is 50-75%;															
	V	honeycomb lesion area is > 75%.															
CTPF stage presentation example	Fibrosis stage/severity										Definition						
	II a										Fibrosis stage II and IPF severity grade a (mild)						
Notes: SpO <sub>2</sub> %; Oxygen saturation of peripheral blood. SpO <sub>2</sub> % is the resting arterial oxygen saturation measured at fingertips. FVC: forced vital capacity. FVC% the actual FVC over the predicted FVC. DLco: diffusing capacity of the lung for carbon monoxide. DLco% pred: the percentage of the actual DLco over the pre																	

## Validation of the New Scoring (CTPF) Method

### Patients' Clinical Data

We retrospectively analyzed the medical records and survival status of 212 patients who were diagnosed with IPF in the Department of Respiratory Medicine of Shanghai Pulmonary Hospital from 2011 to 2017. The final IPF diagnosis was confirmed by the multidisciplinary group of the hospital following the diagnostic criteria of the 2018 IPF international guidelines [1]. Patients' gender, age, lung function, SpO<sub>2</sub>% (or SaO<sub>2</sub>%), chest HRCT, occupation, and smoking history were collected. All the 212 patients were followed up in clinic visits (and telephone follow-up). The follow-up data included patient survival, time of death (the year and month of death), cause of death, occurrence of other complications, whether undergoing lung transplantation, and the time of lung transplantation. The last follow-up date was November 30, 2018. The flow chart of patient screening and enrollment and the follow-up results are presented in Fig. 2.

### Scoring the Clinical Data

Two radiologists used the CT-based pulmonary fibrosis staging method described above to evaluate patients' chest HRCT images. The average scores from the two radiologists were used as patients' final lung fibrosis scores, and then the scores were used to stage pulmonary fibrosis according to the criteria described in Table 2. Patients' age, gender, FVC%pred, DLco%pred, and SpO<sub>2</sub>% were scored according to the criteria in Table 2, and the total scores were used to estimate PF-based disease severity according to the criteria in Table 2. The definition of disease severity is: score 0-3 for grade (a) mild; score 4-6 for grade (b) moderate; score 7-10 for grade (c) severe. The CT-based stage and the PF-based severity were combined to determine patients' CTPF stage (Examples are presented in Additional file).

### Statistical Analyses

Measurement data are expressed as mean ± standard deviation (SD). Count data are presented as percentage (%) or proportion (%). Intra-group correlation coefficient was calculated to estimate the CT score consistency between the two radiologists [19, 20]. Spearman correlation coefficient was calculated to analyze the correlation between CT-based fibrosis scores and pulmonary function parameters (FVC%pred, DLco%pred, SpO<sub>2</sub>%) and CPI index. The competition risk (Fine-Gray) model was used to analyze the relationship between prognosis (cumulative mortality) and the CT-based fibrosis stage and the PF-based severity grade [21]. Patients' survival period was defined from the time when patients' data were acquired to the time of death endpoint or the last follow-up visit. The time unit was month. The death endpoint of this study was defined as the death caused by lung diseases (IPF exacerbation or IPF combined with lung cancer). Lung transplantation is considered to be the most effective treatment for patients with IPF, so the occurrence of lung transplantation was considered as a competitive risk event in this study [22]. Other types of data were treated as censored data.

We used the following strategies to develop and evaluate disease prognosis prediction models: (1) Considered lung transplantation occurrence as a competitive risk event and used CT-based stage, PF-based grade, and CTPF comprehensive stage as predictors. To estimate the accuracy of prediction models, we included the GAP staging method proposed by Brett Ley, MD et al [8] in our analysis. We used all the data and the Fine-Gray regression analysis to establish 4 death-risk prediction models: CT-based fibrosis stage model, PF-based severity grade model, CTPF combined stage model, and GAP stage model. (2) The Bootstrap cross-validation method was used to validate the predictive effectiveness of the 4 models, and the validation was repeated 1000 times to obtain the following average indexes of model prediction accuracy: area under the ROC curve (AUC), Brier score, and a calibration curve. The AUC value reflects the discrimination of the models. It is generally accepted that the model has a satisfactory discrimination to death risk from a disease when AUC is > 75%. The calibration curve reflects the consistency between the predicted risk and the actual risk. The Brier scores reflect both the discrimination and calibration of a model. The smaller the Brier score is, the better the discrimination and calibration of a model is [21]. (3) Prepared a nomogram to display the CTPF model-predicted one-, two-, and three-year cumulative risk of death in patients with different CT stage and PF grade [23].

The statistical software used in this study was IBM SPSS24.0, Stata/MP14.0 and R3.4.3 software.

## Results

### Patients' Clinical Characteristics

Patient screening flow chart is displayed in Fig. 2. A total of 212 patients with IPF were screened, and 192 of them met the inclusion criteria [1] and were included to validate the CTPF comprehensive staging method. Of the 192 included patients, 86 survived; 74 died; 32 were lost to follow-up; 15 patients underwent lung transplantation. Patients' general clinical characteristics are displayed in Table 3. The mean age was  $64.1 \pm 7.7$  (years) and the average survival time was  $28.1 \pm 19.5$  (months). The majority of the patients were men (183/192, 95.3%) and had a history of smoking (138/192, 71.9%). Most of the patients had a CT-based fibrosis stage of II-IV.

Table 3  
Patients' General Clinical Characteristics

Patient Data	Values
Median age years	64.1 ± 7.7
Male/female	183/9
Smokers/Never-smokers	138/54
Survival time(months)	28.1 ± 19.5
SpO <sub>2</sub> %	95.4 ± 3.2
FVC% pred	72.6 ± 20.3
FEV <sub>1</sub> % pred	75.4 ± 20.6
DLco% pred	52.3 ± 28.8
FEV <sub>1</sub> /FVC%	83.5 ± 7.8
CT Score values by Reviewer 1	24.4 ± 14.1
CT Score values by Reviewer 2	24.7 ± 14.4
CT-based stage I/II/III/IV/V	0/107/72/13/0
PF-based grade a/b/c	86/77/29
GAP stage I/II/III	97/65/30
CPI	52.3 ± 18.4
Notes: Measurement data are presented as mean ± standard deviation (SD). Count data are presented as percentage or proportion.	
SpO <sub>2</sub> %; oxygen saturation of peripheral blood. SpO <sub>2</sub> is the resting arterial oxygen saturation measured at fingertips. FVC: forced vital capacity. FVC% pred: the percentage of the actual FVC over the predicted FVC. FEV <sub>1</sub> : forced expiratory volume in one second. FEV <sub>1</sub> % pred: the percentage of the actual FEV <sub>1</sub> over the predicted FEV <sub>1</sub> . DLco: diffusing capacity of the lung for carbon monoxide. DLco% pred: the percentage of the actual DLco over the predicted DLco. FEV <sub>1</sub> /FVC%: the percentage of FEV <sub>1</sub> over FVC. CT Score values by reviewer 1 and CT Score values by reviewer 2 were the scores from the two radiologists using the "4-section honeycomb lung percentage" method to score patients' HRCT imaging results. CT-based stage: The stage was determined by using the average score of the two radiologists and following the criteria described in Table 2. PF-based grade: The grade was determined by using the pulmonary function and physiological parameters (age, gender, FVC%pred, DLco%pred, and SpO <sub>2</sub> %) and following the description in Table 2. The grade was defined as: mild (a), moderate (b), and severe (c). GAP (gender, age, and physiologic variables) stage followed the recommendation by Brett Ley, and a higher stage represented a greater death risk. CPI: composite physiologic index. In 2002, Athol U. Wells and colleagues proposed to use CPI, which combined chest CT and pulmonary functional parameters, to assess the severity of interstitial lung diseases (ILDs). A higher CPI represents a more severe ILD.	

### The Relationship Between CT-based Stage/PF-based Severity and Pulmonary Function and Death Risk

The average CT scores of the 192 patients from the two radiologists using the "4-section honeycomb percentage" method were  $24.4 \pm 14.1$  and  $24.7 \pm 14.4$ , respectively; the highest scores were 67 and 65, respectively, and the lowest values were 1 and 3, respectively (Table 3). The intra-group correlation coefficient of the scores from the two radiologists was 0.95 ( $P < 0.05$ ). For each patient, the mean CT score from the two radiologists was used as the final CT score. The final CT scores were used in the Spearman correlation analysis to assess the correlation between the CT scores and pulmonary function parameters (Fig. 3). The CT scores negatively correlated with FVC%pred ( $r_s = -0.47$ ,  $P < 0.01$ , Fig. 3A), DLco%pred ( $r_s = -0.66$ ,  $P < 0.01$ , Fig. 3B), and SpO<sub>2</sub>% ( $r_s = -0.40$ ,  $P < 0.01$ , Fig. 3C) and positively correlated with CPI index ( $r_s = 0.63$ ,  $P < 0.01$ , Fig. 3D), which represented ILD severity. These data support that the "4-section honeycomb lung percentage" scoring method can effectively represent the severity of pulmonary fibrosis.

To analyze the correlation between CT-based stage and death risk, we performed Fine–Gray univariate regression (Fig. 4A) and multivariate regression to eliminate the potential confounding effects from the PF-based grade (Fig. 4B). Both analyses revealed that CT stage positively correlated with death risk. Similarly, both Fine–Gray univariate regression (Fig. 4C) and multivariate regression to eliminate the potential confounding effects from the CT-based stage (Fig. 4D) found that PF-based grade also positively correlated with death risk.

## CTPF stage

HRCT images of two representative cases are displayed in Additional file. Figure.S1 Example. A shows that the patient was CT-based stage III and PF-staged grade c and thus CTPF stage III c. The patient developed IPF exacerbation and died 23 months after the patient’s clinical data were acquired for the assessment of this study. Figure.S1 Example. B shows CT-based stage II and PF-based grade a and thus CTPF stage II a, and this patient survived well in the 39-month follow-up visit.

Table 4 displays the results from 4 Fine-Gray competitive risk regression prediction models. The predictive factors of the four models were CT-based stage, PF-based grade, CTPF comprehensive stage, and GAP stage, respectively. The CT model, PF model, and GAP model demonstrated that CT-based stage, PF-based grade, and GAP stage were risk factors for death from IPF. The CTPF model showed that CT-based stage and PF-based grade were independent predictors of death from IPF regardless of the type (univariate or multivariate) of the analysis.

Table 4  
Fine–Gray Death Risk Regression Analysis Results From 4 Prediction Models

	Hazard Ratio (HR)	P-value	95% CI
<b>Model CT</b>			
CT II	referent		
CT III	2.22	0.001	1.36 to 3.63
CT IV	5.32	0.001	1.97 to 14.39
<b>Model PF</b>			
PF(a)	referent		
PF(b)	1.99	< 0.001	1.18 to 3.34
PF(c)	4.39	< 0.001	2.22 to 8.70
<b>Model CTPF</b>			
CT II	referent		
CT III	1.76	0.039	1.03 to 3.00
CT IV	3.10	0.059	0.96 to 10.04
PF(a)	referent		
PF(b)	1.68	0.066	0.97 to 2.92
PF(c)	2.79	0.011	1.27 to 6.13
<b>Model GAP</b>			
GAP I	referent		
GAP II	2.30	0.002	1.37 to 3.87
GAP III	3.31	< 0.001	1.71 to 6.43
Notes: CI: confidence interval. Model CT: CT-based stage was used in the univariate Fine–Gray death risk regression analysis. Model PF: PF-based grade was used in the univariate Fine–Gray death risk regression analysis. Model CTPF: CTPF comprehensive stage was used in the multivariate Fine–Gray death risk regression analysis. Model GAP: GAP stage proposed by Brett Ley was used in univariate Fine–Gray death risk regression analysis. CT II: Honeycomb lesion area was < 25% of the entire lung. CT III: Honeycomb lesion area was 25%-49% of the entire lung. CT IV: Honeycomb lesion area was 50%-75%. PF-based grade was determined by assessing the scores of age, gender, FVC%pred, DLco%pred, and SpO <sub>2</sub> % according to the criteria in Table 2 and adding the scores. PF (a): score 0–3. PF(b): score 4–6. PF(c): score 7–10. GAP I: score 0–3. GAP II: score 4–5. GAP III: score 6–8.			

The AUC versus time plot from the Bootstrap cross-validation model is displayed in Fig. 5. Compared with the other three prediction models (CT model, PF model, and GAP model), the AUC value calculated from the CTPF model was the best; both the one-year and the two-year AUC values of the CTPF model were > 75%. Figure 6A is the nomogram showing CTPF-based death risk prediction, which was prepared from the CT-based stage and PF-based grade multivariate Fine-Gray regression coefficients. Figures 6B, 6C, and 6D show the calibration curves of the four prediction models after Bootstrap cross-validation. The CTPF model had the best stability. The one-, two-, and three-year cumulative death risks of patients at different CTPF stage are displayed in Table 5. When patients had the same CT-based stage, their cumulative death risk increased as their PF-based grade increased. When patients had the same PF-based grade, their cumulative death risk increased as their CT-based stage increased. Thus, combination of CT-based stage and PF-based grade could improve the accuracy of death risk prediction.

Table 5  
CTPF Model-predicted one-, two-, and three-year accumulative death risk of patients at different CTPF stage

CTPF stage	1-y Cumulative mortality %	2-y Cumulative mortality %	3-y Cumulative mortality %
II a	4.81	13.07	17.50
II b	7.95	20.98	27.63
II c	12.84	32.34	41.51
III a	8.29	21.82	28.67
III b	13.54	33.88	43.33
III c	21.44	49.65	61.02
IV a	14.18	35.25	44.94
IV b	22.66	51.84	63.32
IV c	34.70	70.23	81.05

Notes: CTPF stage: CTPF-based comprehensive stage.

II a: CT stage II and PF grade a; II b: CT stage II and PF grade b; II c: CT stage II and PF grade c; III a: CT stage III and PF grade a; III b: CT stage III and PF grade b; III c: CT stage III and PF grade c; IV a: CT stage IV and PF grade a; IV b: CT stage IV and PF grade b; IV c: CT stage IV and PF grade c.

CT II: Honeycomb lesion area was < 25% of the entire lung. CT III: Honeycomb lesion area was 25%-49% of the entire lung. CT IV: Honeycomb lesion area was 50%-75%. PF-based grade was determined by assessing the scores of age, gender, FVC%pred, DLco%pred, and SpO<sub>2</sub>% according to the criteria in Table 2 and adding the scores. PF (a): score 0–3. PF(b): score 4–6. PF(c): score 7–10.

## Comparison Between CTPF Stage and GAP Stage

The CTPF model, which combined CT-based stage and PF-based grade, predicted that the AUC values of one-, two-, and three-year accumulative death risk were 78.6 (95% CI: 58.6–93.0), 77.8 (95% CI: 58.8–83.4), and 73.4 (95% CI: 57.5.1–85.1), respectively. The AUC values from the GAP model were 73.9 (95% CI: 58.4–87.7), 72.3 (95% CI: 58.8–83.4), and 70.8 (95% CI: 57.8.1–82.9), respectively. These data and the calibration curves after the cross-validation (Fig. 6) suggest that the CTPF stage appears to be more accurate for predicting death risk than the other 3 models.

## Discussion

Comparison of several available IPF staging methods (Table 1) shows that the staging results from some methods, such as the GAP and JRS methods, fail to accurately reflect IPF severity and predict prognosis because the methods include too few parameters. The calculation methods in the CRP and CPI scoring systems are too complex to be adopted in clinical practice [10]. Therefore, a new scoring method that can accurately assess IPF severity, predict prognosis, and can be used easily is greatly needed.

Chest HRCT is one of the common clinical examinations to diagnose IPF and assess IPF severity and prognosis. Honeycomb lung is the most representative lesion of pulmonary fibrosis, and the area of honeycomb lesion directly correlates to IPF prognosis [11–14, 16].

Currently, CT scoring for IPF patients includes manual semi-quantitative evaluation and total quantitative evaluation by artificial intelligence. Although the manual method is simple to use, the evaluation results are susceptible to the wide variation from different evaluators [24–26]. We took applicability in clinical practice into consideration and based on calculus principles to develop a “four-section honeycomb lung percentage” method, which can determine the proportion of honeycomb lung accurately and reduce inter-evaluator variation. In the current study, two radiologists reviewed patient HRCT results and determined the honeycomb lung percentage independently. The consistency coefficient of the two radiologists’ scoring results was 0.95 ( $P < 0.05$ ), and the fibrosis stage determined according to the honeycomb percentage was also consistent in the two radiologists. In addition, the CT-based stage negatively correlated with patients’ lung function parameters (FVC%pred, DLco%pred, and SpO<sub>2</sub>%) and positively correlated with CPI index (Fig. 3). The CPI index reflects IPF severity. Patients with higher CT-based stage had a greater accumulative death risk. These results indicate that our CT-based fibrosis staging method may effectively reflect IPF severity and prognosis.

Previous studies have shown that age, gender, oxygen use at rest, lower FVC %pred and lower DLco % pred were associated closely with risk of death in patients with IPF [4–8, 18, 27]. Thus, we selected the 5 important and clinical easily available lung function and physiological parameters, FVC%pred, DLco%pred, SpO<sub>2</sub>%, age, and gender to assess IPF severity grade (PF-based severity grade). Both our univariate and multivariate regression analysis revealed that PF-based severity grade was an independent risk factor for death from IPF.

Compared with the CT-based fibrosis staging method, the PF-based severity grading method, and the GAP staging method, the CTPF comprehensive staging method, which combined the CT-based fibrosis staging and the PF-based severity grading methods, showed the best AUC value, Brier score, and stability in terms of predicting death risk. For example, the case presented in Figure S1 was CTPF stage III c, and his predicted 2-year death risk was 49.65% according to Table 5. The patient died of acute IPF exacerbation 23 months after his clinical data were collected for the assessment in this study. The case in Figure S2 was CTPF stage II a, which corresponded to a predicted 3-year death risk of only 17.50%. This patient survived well 39 months after his data were collected for the assessment. These results support that our CTPF comprehensive staging method can accurately predict patient death risk.

Lung transplantation has been considered to be an effective treatment for improving the survival of patients with IPF. Thus, we used lung transplantation as a competitive risk of death to calculate death risk when we validated the new CTPF comprehensive staging method. However, lung transplantation also has a death risk [21]. In 2015, Yusen, RD et al [28] reported that the global lung transplantation one-year and three-year death risk was 20% and 35%, respectively. When the death risk (Table 5) calculated from the CTPF staging method was higher than lung transplantation death risk, lung transplantation should be recommended to patients. Therefore, our CTPF comprehensive evaluation method may be useful when physicians determine an optimal time of lung transplantation for patients with IPF. The majority of the included cases were CT stage II-IV in the current study. A large sample size including patients with a more diverse CT stage should be used in future investigations.

## Conclusion

This study developed a new method of “four-section honeycomb lung percentage” on HRCT combined other comprehensive multiparameter (CTPF) for evaluating pulmonary fibrosis severity. This new method can effectively assess IPF severity and predict death risk. Compared with existing assessment methods, the CTPF method used comprehensive parameters, was simple and easy to be used in clinical practice, and showed high accuracy.

## Abbreviations

IPF

Idiopathic pulmonary fibrosis

CRP

Clinical-radiographic-physiologic

FVC

Forced vital capacity

DLco

Diffusing capacity of the lung for carbon Monoxide

CT

Computed tomography

CPI

Composite physiologic index

GAP

Gender, age, and physiologic variables

SaO<sub>2</sub>

Oxyhemoglobin saturation

SpO<sub>2</sub>

Oxygen saturation of peripheral blood

HRCT

High-resolution computed tomography

UIP

Usual interstitial pneumonia

JRS

Japanese Respiratory Society

PF

Pulmonary function & physiological features

CTPF

HRCT combined pulmonary function & physiological features

AUC

area under curve

## Declarations

### Ethics Statement

The study was approved by the Institutional Ethics Committee of Shanghai Pulmonary Hospital (No. K17-006)

### Consent for publication

Not applicable.

### Availability of data and materials

The datasets used and analyzed during the current study are available from the corresponding author on reasonable request.

### Competing interests

The authors declare that they have no competing interests.

## Funding

This study was funded by grants from the National Science Foundation of China (Grant No: 81730002, 81670055, 81670056, 91442103, 81500052, and 81570057), Ministry of Science and Technology of the People's Republic of China (2016YFC1100200 and 2016YFC1100204), Shanghai Family Planning Commission Health Industry Clinical Research Project(20184Y0084), and National Science Foundation of Shanghai (18ZR1431400)

## Acknowledgement

Not applicable

## Author Contributions

HP Li, Chengsheng Yin, Aihong Zhang, Yuan Zhang, Yiliang Su, Fen Zhang, participated in the conception, hypothesis and design of the study. Chengsheng Yin performed the experiments. Chengsheng Yin, Aihong Zhang carried out the statistical analyses. All authors contributed to interpretation of the data. Jingyun Shi and Yanan Chen participated in CT imaging evaluation. Chengsheng Yin, Aihong Zhang and HP Li wrote the manuscript and all authors made critical revisions. All authors read and approved the final manuscript.

## Authors' information

Chengsheng Yin<sup>1†</sup> Yuan Zhang<sup>1†</sup> Yiliang Su<sup>1†</sup> Feng Zhang<sup>1†</sup> Jingyun Shi<sup>2</sup> Yanan Chen<sup>2</sup> Dong Weng<sup>1</sup> Tao Chen<sup>1</sup> Mengmeng Zhao<sup>1</sup> Ying Zhou<sup>1</sup> Xing Jiang<sup>1</sup> Aihong Zhang<sup>3§</sup> Huiping Li<sup>1\*</sup>

Affiliations:1. Department of Respiratory Medicine, Shanghai Pulmonary Hospital, Tongji University, School of Medicine, Shanghai 200433, China.2. Department of Radiology, Shanghai Pulmonary Hospital, Tongji University, School of Medicine, Shanghai 200433, China.3. Department of Medical Statistics, Tongji University, School of Medicine, Shanghai 200092, China.

†Contributed equally to this work

\*Corresponding author: Prof. Dr. Hui-Ping Li, MD, 507 Zheng Min Road, Shanghai 200433, China. Email: liw2013@126.com; Fax: 86-21-65111298; Phone: 86-21-65115006-2103

Co-corresponding author: Prof. Ai-Hong Zhang, PhD, Department of Medical Statistics, Tongji University, School of Medicine, Shanghai 200092, P.R. China. Email: zhangah@tongji.edu.cn

## References

1. Raghu G, Remy-Jardin M, Myers JL, et al. Diagnosis of Idiopathic Pulmonary Fibrosis. An Official ATS/ERS/JRS/ALAT Clinical Practice Guideline. *Am J Respir Crit Care Med* 2018; 198(5): e44-e68.
2. Ley B, Collard HR, King TE, Jr. Clinical course and prediction of survival in idiopathic pulmonary fibrosis. *Am J Respir Crit Care Med* 2011; 183(4): 431-40.
3. Gonnella JS, Hornbrook MC, Louis DZ. Staging of Disease - a Case-Mix Measurement. *Jama-J Am Med Assoc* 1984; 251(5): 637-44.
4. Watters LC, King TE, Schwarz MI, Waldron JA, Stanford RE, Cherniack RM. A clinical, radiographic, and physiologic scoring system for the longitudinal assessment of patients with idiopathic pulmonary fibrosis. *Am Rev Respir Dis* 1986; 133(1): 97-103.
5. Watters LC, Schwarz MI, Cherniack RM, et al. Idiopathic Pulmonary Fibrosis - Pretreatment Bronchoalveolar Lavage Cellular-Constituents and Their Relationships with Lung Histopathology and Clinical-Response to Therapy. *Am Rev Respir Dis* 1987; 135(3): 696-704.
6. King TE, Tooze JA, Schwarz MI, Brown KR, Cherniack RM. Predicting survival in idiopathic pulmonary fibrosis: Scoring system and survival model. *Am J Resp Crit Care* 2001; 164(7): 1171-81.
7. Wells AU, Desai SR, Rubens MB, et al. Idiopathic pulmonary fibrosis - A composite physiologic index derived from disease extent observed by computed tomography. *Am J Resp Crit Care* 2003; 167(7): 962-9.
8. Ley B, Ryerson CJ, Vittinghoff E, et al. A Multidimensional Index and Staging System for Idiopathic Pulmonary Fibrosis. *Ann Intern Med* 2012; 156(10): 684-U58.
9. Okuda R, Hagiwara E, Baba T, Kitamura H, Kato T, Ogura T. Safety and efficacy of pirfenidone in idiopathic pulmonary fibrosis in clinical practice. *Respiratory medicine* 2013; 107(9): 1431-7.
10. Robbie H, Daccord C, Chua F, Devaraj A. Evaluating disease severity in idiopathic pulmonary fibrosis. *Eur Respir Rev* 2017; 26(145).
11. Best AC, Meng JF, Lynch AM, et al. Idiopathic pulmonary fibrosis: Physiologic tests, quantitative CT indexes, and CT visual scores as predictors of mortality. *Radiology* 2008; 246(3): 935-40.
12. Rosas IO, Yao JH, Avila NA, Chow CK, Gahl WA, Gochuico BR. Automated Quantification of High-Resolution CT Scan Findings in Individuals at Risk for Pulmonary Fibrosis. *Chest* 2011; 140(6): 1590-7.
13. Watadani T, Sakai F, Johkoh T, et al. Interobserver Variability in the CT Assessment of Honeycombing in the Lungs. *Radiology* 2013; 266(3): 936-44.
14. Jacob J, Hirani N, van Moorsel CHM, et al. Predicting outcomes in rheumatoid arthritis related interstitial lung disease. *Eur Respir J* 2019; 53(1).

15. Su Y, Gu H, Weng D, et al. Association of serum levels of laminin, type IV collagen, procollagen III N-terminal peptide, and hyaluronic acid with the progression of interstitial lung disease. *Medicine (Baltimore)* 2017; 96(18): e6617.
16. Sverzellati N, Devaraj A, Desai SR, Quigley M, Wells AU, Hansell DM. Method for Minimizing Observer Variation for the Quantitation of High-Resolution Computed Tomographic Signs of Lung Disease. *J Comput Assist Tomo* 2011; 35(5): 596-601.
17. Lynch DA, Godwin JD, Safrin S, et al. High-resolution computed tomography in idiopathic pulmonary fibrosis: diagnosis and prognosis. *Am J Respir Crit Care Med* 2005; 172(4): 488-93.
18. Snyder L, Neely ML, Hellkamp AS, et al. Predictors of death or lung transplant after a diagnosis of idiopathic pulmonary fibrosis: insights from the IPF-PRO Registry. *Respiratory research* 2019; 20(1): 105.
19. Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics* 1977; 33(1): 159-74.
20. Kundel HL, Polansky M. Measurement of observer agreement. *Radiology* 2003; 228(2): 303-8.
21. Zhang ZH, Cortese G, Combesure C, et al. Overview of model validation for survival regression model with competing risks using melanoma study data. *Ann Transl Med* 2018; 6(16).
22. Thabut G, Christie JD, Ravaud P, et al. Survival After Bilateral Versus Single-Lung Transplantation for Idiopathic Pulmonary Fibrosis. *Ann Intern Med* 2009; 151(11): 767-W253.
23. Zhang ZH, Geskus RB, Kattan MW, Zhang HY, Liu TY. Nomogram for survival analysis in the presence of competing risks. *Ann Transl Med* 2017; 5(20).
24. Demedts M, Costabel U. ATS/ERS international multidisciplinary consensus classification of the idiopathic interstitial pneumonias. *Eur Respir J* 2002; 19(5): 794-6.
25. Flaherty KR, Mumford JA, Murray S, et al. Prognostic implications of physiologic and radiographic changes in idiopathic interstitial pneumonia. *Am J Respir Crit Care Med* 2003; 168(5): 543-8.
26. 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.
27. Zappala CJ, Latsi PI, Nicholson AG, et al. Marginal decline in forced vital capacity is associated with a poor outcome in idiopathic pulmonary fibrosis. *European Respiratory Journal* 2010; 35(4): 830-5.
28. Yusen RD, Edwards LB, Kucheryavaya AY, et al. The Registry of the International Society for Heart and Lung Transplantation: Thirty-second Official Adult Lung and Heart-Lung Transplantation Report–2015; Focus Theme: Early Graft Failure. *The Journal of heart and lung transplantation: the official publication of the International Society for Heart Transplantation* 2015; 34(10): 1264-77.

## Figures

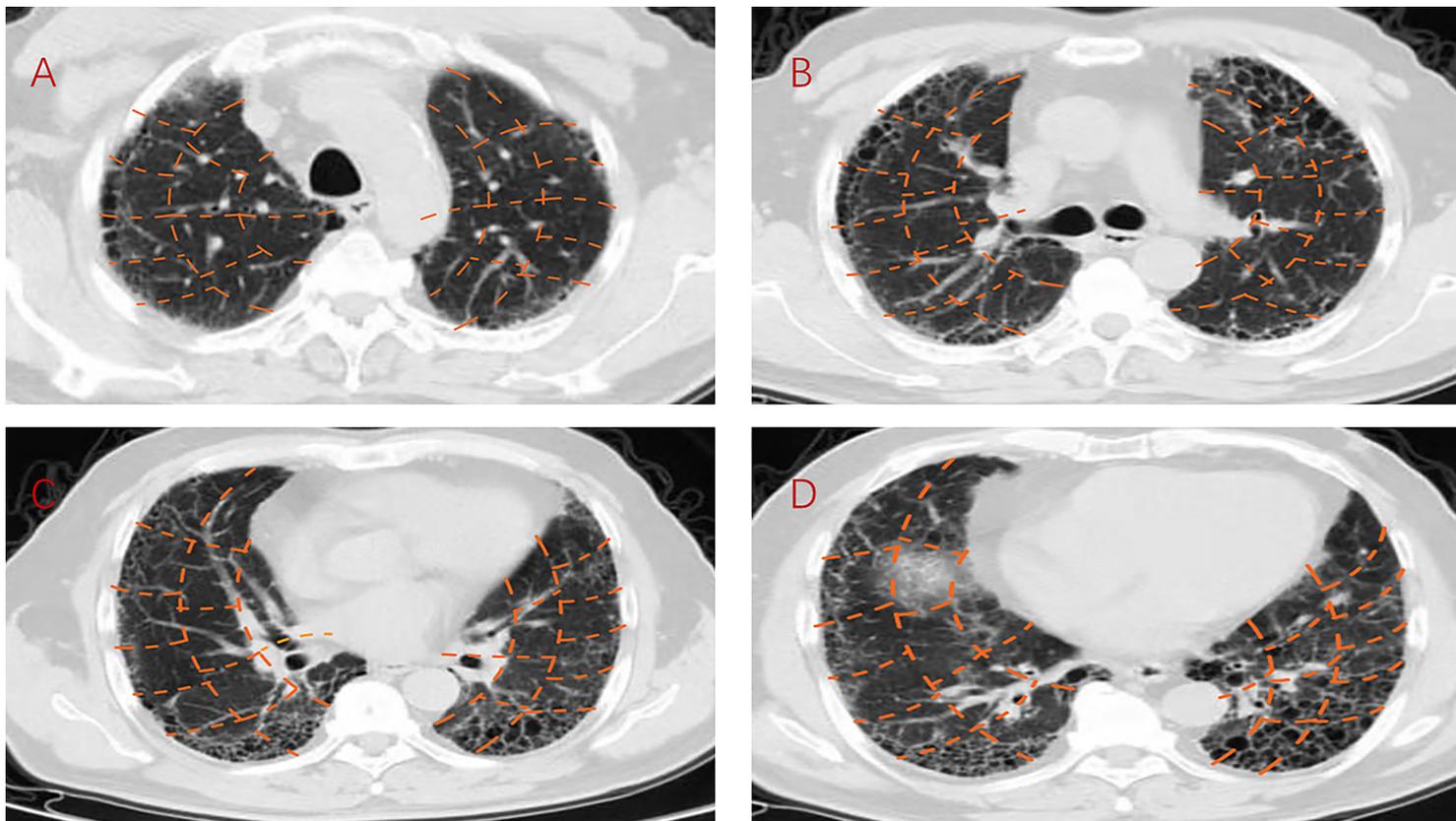
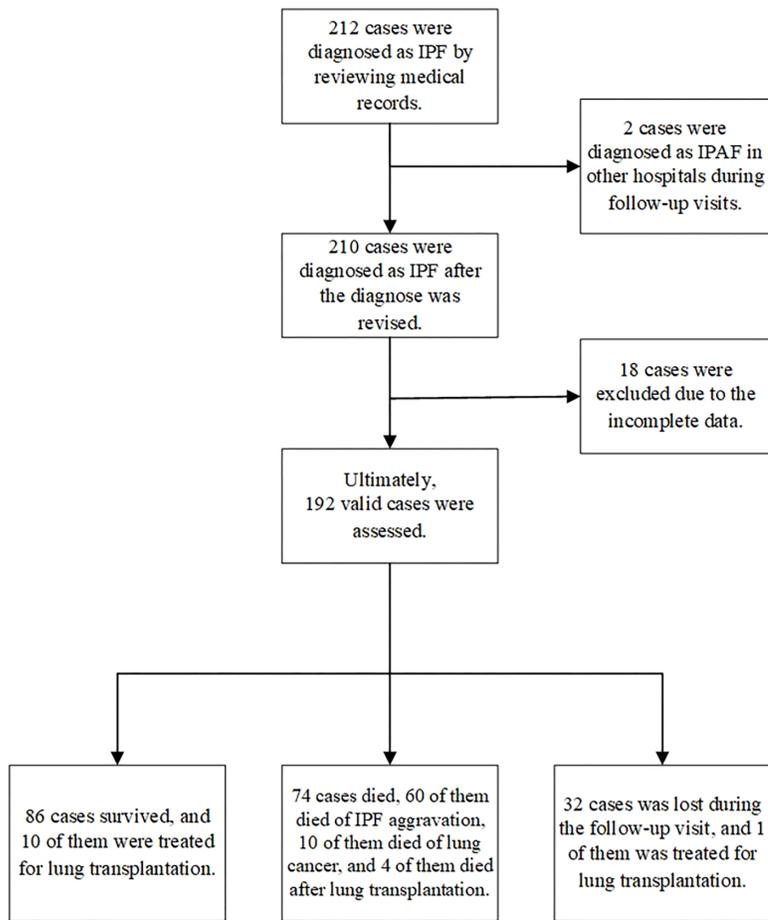


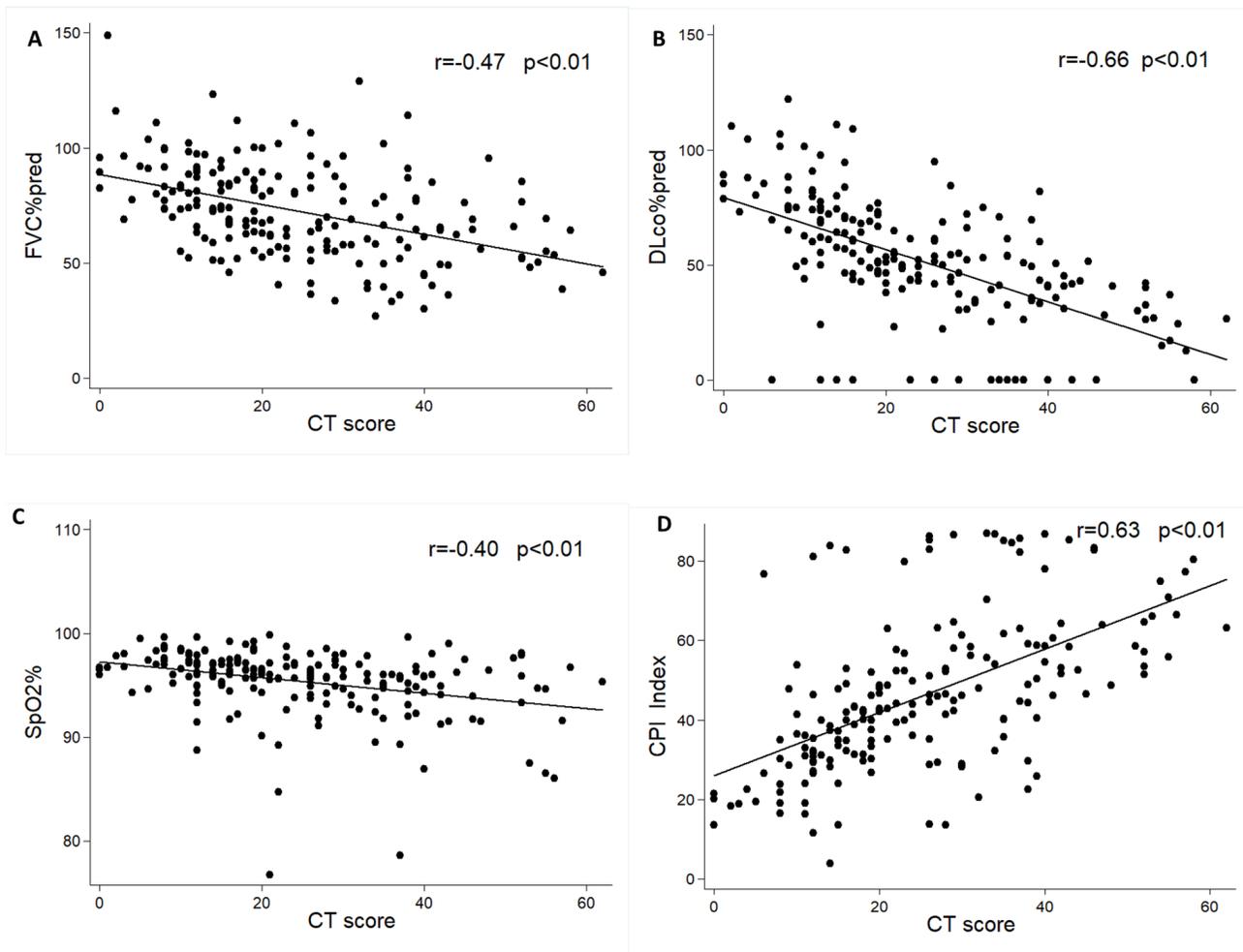
Figure 1

CT scoring method: 4-section honeycomb lung percentage Four representative lung CT sections were selected for semi-quantitative estimation: A. the aortic arch section, B. the tracheal bifurcation section, C. the section of basal (dorsal) segment of the tracheal bifurcation at the inferior lobes, D. the section below the right lung apex. Each section included both the left and the right lungs. The largest transverse diameter line of each lung section was evenly divided into three parts, and then the lung section was divided into inner, middle, and outer sections by drawing lines starting from the dividing points along the shape of patient's thorax. The outer lung section was then evenly divided into 6 small areas; the middle lung section 4-5 areas (5 areas for a large middle section); the inner lung section 2 areas. Thus, in total, the 4 CT sections comprised 8 lung sections (4 left + 4 right lung) and were evenly divided into approximately 100 small areas (12 or 13 per lung section × 8). Each small area was scored as 1 when there was positive honeycomb lesion in the area, and the total score of the entire lung was used as the total honeycomb lung score. The total traction bronchiectasis score was calculated in the same way as the total honeycomb lung score. The honeycomb lung percentage was calculated as: (total honeycomb score + total traction bronchiectasis score) ÷ total number of the small areas × 100%. For example, if the 8 lung sections were evenly divided into 100 small areas and 30 of them were scored as positive honeycomb lung or traction bronchiectasis, then the honeycomb lung percentage was 30%



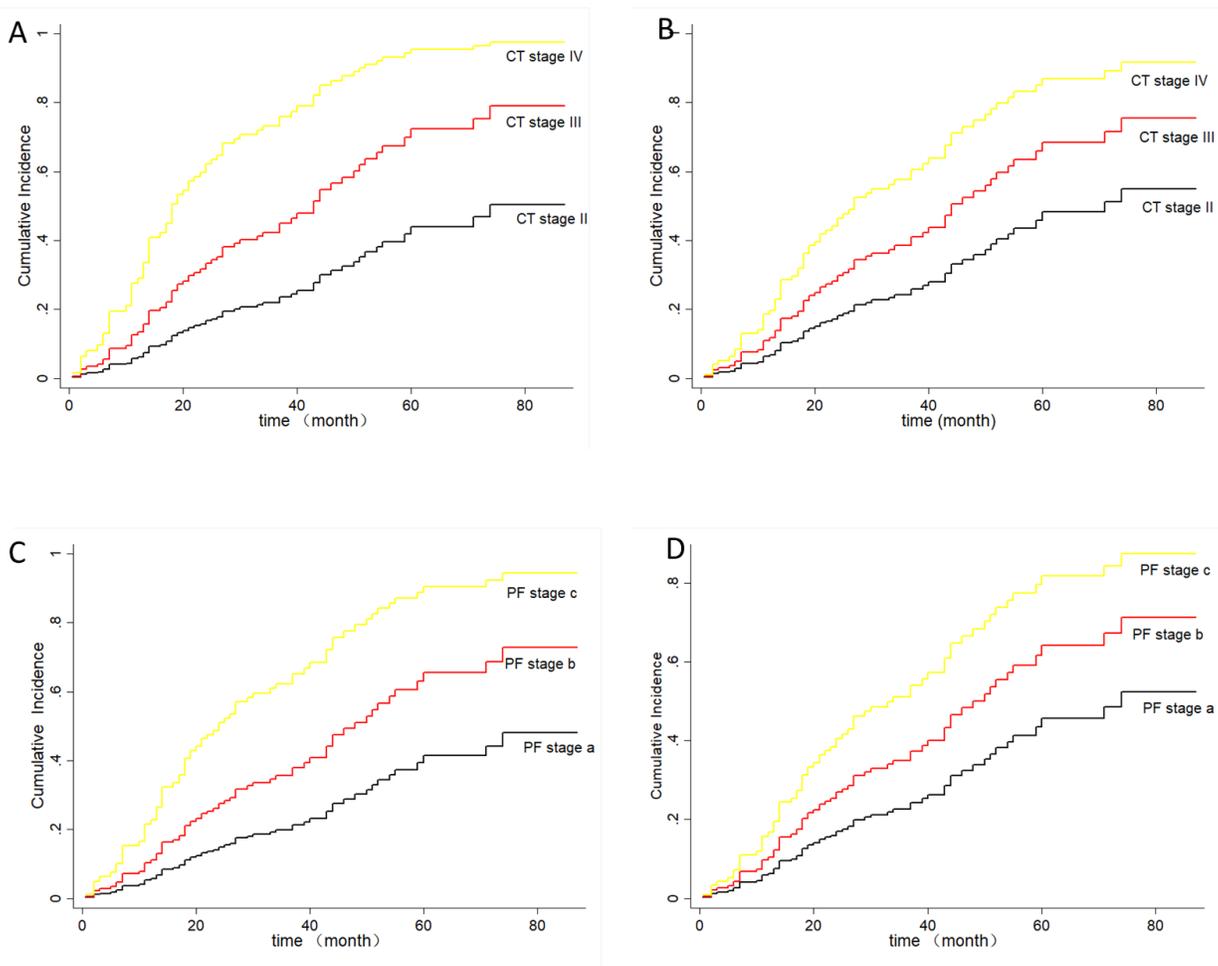
**Figure 2**

Patient Screening Flowchart A total of 212 patients with IPF were screened. IPF diagnosis followed the 2018 IPF diagnosis and treatment guidelines. Of the 212 cases, 20 were excluded (2 cases of interstitial pneumonia with autoimmune features diagnosed during follow-up visits + 18 cases of insufficient CT and pulmonary function data); 192 of them were included to the analysis. Of the 192 included patients, 86 survived (including 10 cases of lung transplantation); 74 died (10 cases of lung cancer + 60 cases of IPF exacerbation + 4 cases of lung transplantation); 32 were lost to follow-up (one case of lung transplantation); 15 patients underwent lung transplantation.

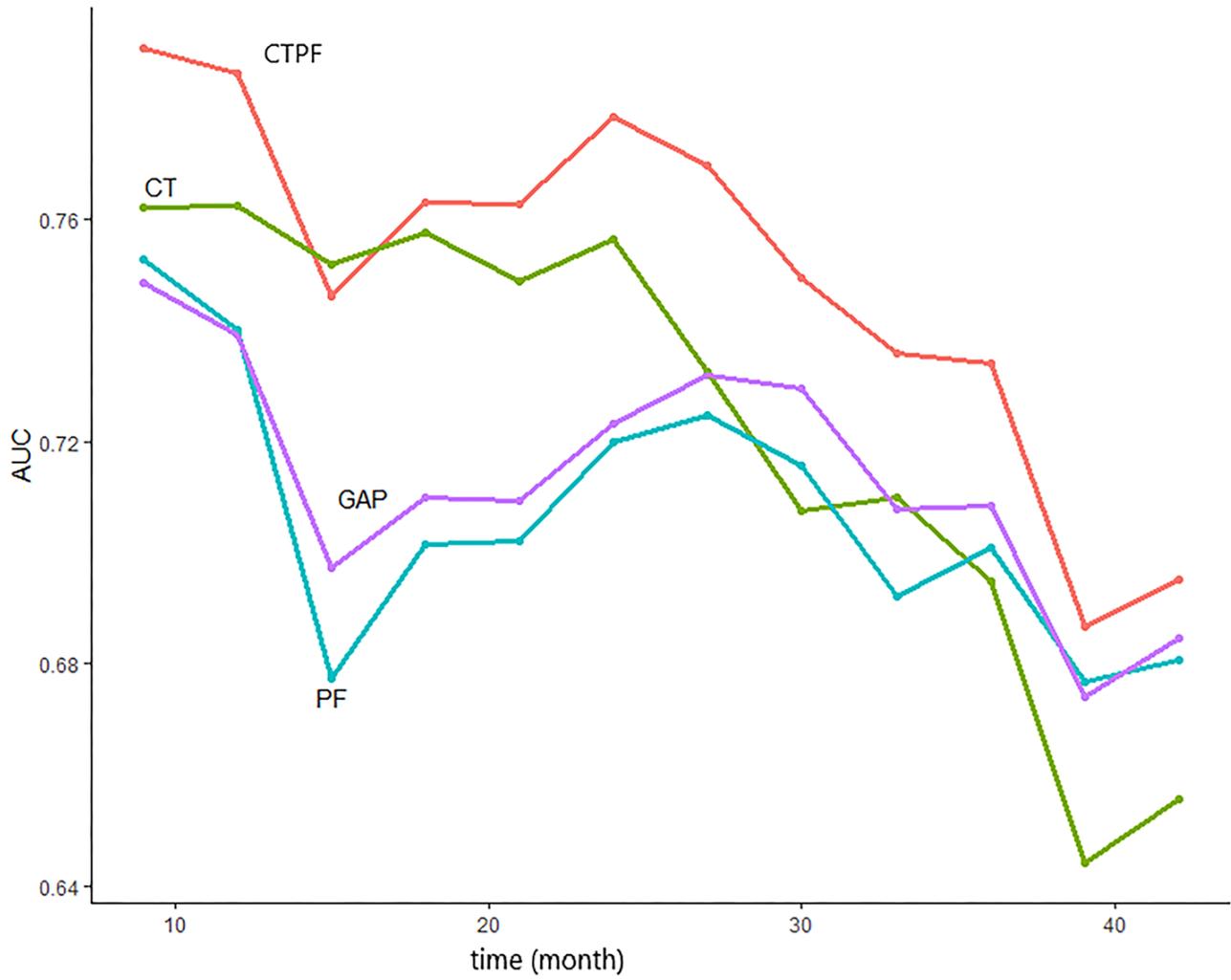


**Figure 3**

The correlation between CT scores and pulmonary function Spearman correlation analysis was performed. CT scores negatively correlated with FVC%pred (A,  $r_s = -0.47$ ,  $P < 0.01$ ), DLco%pred (B,  $r_s = -0.66$ ,  $P < 0.01$ ), and SpO2% (C,  $r_s = -0.40$ ,  $P < 0.01$ ) and positively correlated with CPI index (D,  $r_s = 0.63$ ,  $P < 0.01$ ). SpO2%: oxygen saturation of peripheral blood. FVC: forced vital capacity. FVC% pred: actual FVC/predicted FVC  $\times 100\%$ . DLco: diffusing capacity of the lung for carbon Monoxide. DLco% pred: actual DLco/ predicted DLco  $\times 100\%$ . CPI: composite physiologic index. In 2002, Athol U. Wells and colleagues proposed to use CPI, which combined chest CT and pulmonary functional parameters, to assess the severity of interstitial lung diseases (ILDs). A higher CPI represents a more severe ILD. CT score: mean CT score from the two radiologists.

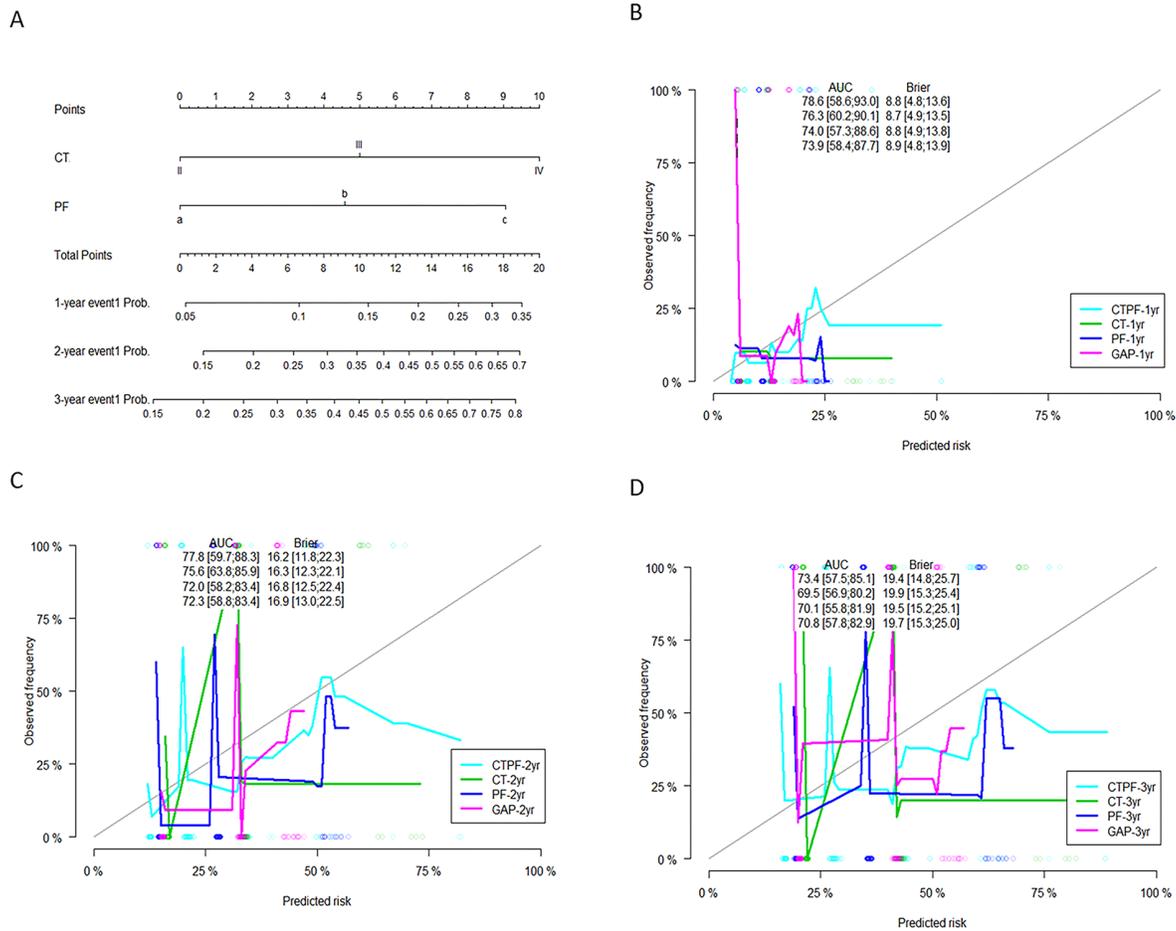


**Figure 4**  
 Correlation Between CT-based Stage and Death Risk & Correlation Between PF-based Grade and Death Risk A. Fine-Gray univariate regression to analyze the correlation between CT stage and death risk. B. Fine-Gray multivariate regression (eliminating the potential confounding effects of PF-based grade) to analyze the correlation between CT stage and death risk. Both analyses show a positive correlation between CT-based stage and death risk. C. Fine-Gray univariate regression to analyze the correlation between PF-based grade and death risk. D. Fine-Gray multivariate regression (eliminating the potential confounding effects of CT-based stage) to analyze the correlation between PF-based grade and death risk. Both analyses show a positive correlation between PF-based grade and death risk. CT stage: The stage was determined by using the average score of the two radiologists and following the criteria described in Table 2. The definition of CT-based stage was: honeycomb lung < 25% was Stage II; honeycomb lung 25%-49% Stage III; honeycomb lung 50%-75% Stage IV; honeycomb lung >75% Stage V. PF-based grade: The grade was determined by using the pulmonary function and physiological parameters (age, gender, FVC%pred, DLco%pred, and SpO2%) and following the description in Table 2. The grade was defined as: PF score 0-3 was mild (a); PF score 4-6 moderate (b); PF score 7-10 severe (c).



**Figure 5**

AUC Versus Time Plots from the 4 Predictive Models The 4 curves are the AUC (area under curve)-versus-time plots. CT-based stage, PF-based grade, CTPF comprehensive stage, and GAP stage were used in the 4 models, respectively, to predict death risk. Bootstrap cross-validation was used to assess the predictive effectiveness of the models.



**Figure 6**

A. Death risk nomogram from the CTPF prediction model. B, C, and D are the calibration curves after cross-validation, which were prepared by using CT-based stage, PF-based grade, CTPF-based stage, and GAP stage as the four prediction models to predict one- (B), two- (C), and three-year (D) cumulative death risks. The CTPF model-predicted AUCs of one-, two-, and three-year death risk were 78.6 (95%CI: 58.6-93.0), 77.8 (95%CI: 58.8-83.4), and 73.4 (95%CI: 57.5-85.1), respectively. The one-, two-, and three-year Brier values were 8.8 (95%CI: 4.8-13.6), 16.2 (95%CI: 11.8-22.3), and 19.4 (95%CI: 14.8-25.7), respectively. The CT model-predicted AUCs of one-, two-, and three-year death risk were 76.3 (95%CI: 60.2-90.1), 75.6 (95%CI: 63.8-85.9), 69.5 (95%CI: 56.9-80.2), respectively. The one-, two-, and three-year Brier values were 8.7 (95%CI: 4.9-13.5), 16.3 (95%CI: 12.3-22.1), and 19.9 (95%CI: 15.3-25.4). The PF model-predicted AUCs of one-, two-, and three-year death risk were 74.0 (95%CI: 57.4-88.6), 72.0 (95%CI: 58.2-83.4), and 70.1 (95%CI: 55.8-81.9), respectively. The one-, two-, and three-year Brier values were 8.8 (95%CI: 4.9-13.8), 16.8 (95%CI: 12.5-22.4), and 19.5 (95%CI: 15.2-25.1), respectively. The GAP model-predicted AUCs of one-, two-, and three-year death risk were 73.9 (95%CI: 58.4-87.7), 72.3 (95%CI: 58.8-83.4), and 70.8 (95%CI: 57.8-82.9), respectively. The one-, two-, and three-year Brier values were 8.9 (95%CI: 4.8-13.9), 16.9 (95%CI: 13.0-22.5), 19.7 (95%CI: 15.3-25.0), respectively. The CTPF model shows the best AUC value, Brier value, and stability, indicating the best predictive effectiveness.

## Supplementary Files

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

- [Additionalfile.docx](#)