European Radiology
https://doi.org/10.1007/s00330-019-06402-6

CHEST

Prediction of idiopathic pulmonary fibrosis progression using early quantitative changes on CT imaging for a short term of clinical 18–24-month follow-ups

Grace Hyun J. Kim[1,2] · Stephan S. Weigt[3] · John A. Belperio[3] · Matthew S. Brown[1] · Yu Shi[2] · Joshua H. Lai[1] · Jonathan G. Goldin[1]

Received: 29 January 2019 / Revised: 21 July 2019 / Accepted: 29 July 2019

Abstract

Objective High-resolution computed tomography (HRCT) plays an indispensable role in the diagnosis of idiopathic pulmonary fibrosis (IPF). Due to unpredictability in progression and the short median survival of 2–5 years, it is critical to delineate the patients with rapid progression. The aim is to evaluate the predictability of IPF progression using the early quantitative changes.

Methods Automated texture-based quantitative lung fibrosis (QLF) was calculated from the anonymized HRCT. Two datasets were collected retrospectively: (1) a pilot study of 35 subjects with three sequential scans (baseline and 6 and 12 months) to obtain a threshold, where visual assessments were stable at 6 months but worsened at 12 months; (2) 157 independent subjects to test the threshold. Landmark Cox regressions were used to compare the progression-free survival (PFS) defined by pulmonary function using the threshold from the early changes in QLF. C-indexes were reported as estimations of the concordance of prediction.

Results A threshold of 4% QLF change at 6 months corresponded to the mean change that worsened on HRCT visually at 12 months from the pilot study. Using the threshold, significant differences were found in the independent dataset (hazard ratio (HZ) = 5.92, $p = 0.001$ by Cox model, C-index = 0.71 at the most severe lobe; and HZ = 3.22, $p = 0.012$, C-index = 0.68 in the whole lung). Median PFS was 11.9 months for subjects with ≥ 4% changes, whereas median PFS was greater than 18 months for subjects with < 4% changes at the most severe lobe.

Conclusion Early structural changes on HRCT using a quantitative score can predict progression in lung function.

Key Points

- *Changes on HRCT using quantitative texture-based scores can play a pivotal role for providing information and an aid tool for timely management decision for patients with IPF.*
- *Quantitative changes on HRCT of 4% or more, which matched 6-month prior changes with visual assessment of worsening, can play a pivotal role for providing prediction of clinical progression by 3–5 folds higher in the next incidence, compared with those of subjects with less than 4% changes.*
- *Early structural changes of 4% or more in a paired HRCT scans derived by quantitative scores can predict the progression in lung function in 1–2 years in subjects with IPF, which is critical information for timely management decision for subjects with IPF where the median survival is 2 to 5 years.*

Keywords Idiopathic pulmonary fibrosis · Quantitative evaluation · Progressive · Prediction · Interstitial lung disease

✉ Grace Hyun J. Kim
gracekim@mednet.ucla.edu

[1] David Geffen School of Medicine, Radiological Science, UCLA, 924 Westwood Blvd. Ste 650, Box 957319, Los Angeles, CA 90095-7319, USA

[2] Fielding School of Public Health, Biostatistics, Computer Vision and Imaging Biomarkers, UCLA, 924 Westwood Blvd. Ste 650, Box 957319, Los Angeles, CA 90095-7319, USA

[3] UCLA Med-Pul Critical Care, Los Angeles, USA

Abbreviation

ALAT	Latin American Thoracic Society
ATS	American Thoracic Society
DLCO	Diffusing capacity of the lungs for carbon monoxide
ERS	European Respiratory Society
FVC	Forced vital capacity
GAP	Gender, age, and pulmonary function
HRCT	High-resolution computed tomography
HZ	Hazard ratio

Published online: 26 August 2019

IPF	Idiopathic pulmonary fibrosis
JRS	Japanese Respiratory Society
LF	Lung fibrosis
PFS	Progression-free survival
PFT	Pulmonary function test
QILD	Quantitative interstitial lung disease
QLF	Quantitative lung fibrosis
SD	Standard deviation
SE	Standard error
TLC	Total lung capacity
UIP	Usual interstitial pneumonia

Introduction

Idiopathic pulmonary fibrosis (IPF) is a rare and fatal interstitial lung disease (ILD) with a short median survival of 2–5 years after the initial diagnosis [1]. Therapeutic agents show stabilization or relatively slow down the rate of decline in the lung function in 1 year. These changes are compared with those in the placebo group [2, 3]. Progression-free survival (PFS) using forced vital capacity (FVC) changes has shown to be associated with mortality, which we infer that the relative stabilization in FVC leads to reduction in mortality [4–6]. However, individualized treatment and timing of treatments are important due to the nature of different rates of disease progression [1, 7]. Up to date, it is uncertain when a patient with IPF will progress in the longitudinal follow-ups. In the course of disease, physicians make important decisions regarding the types of therapeutic treatment, referral to a clinical trial for an investigational drug, and timing of recommending a lung transplant. A reliable method to predict the rate of progression in the near future would be beneficial to patients' health and clinicians in making decisions. Although high-resolution computed tomography (HRCT) plays an indispensable role in the diagnosis of IPF and prediction of survival based on the prognostic factors of the distribution of lung abnormality or usual interstitial pneumonia (UIP) patterns, a short longitudinal interval of the structural change on HRCT is uncertain in relation to progression in lung function [8, 9].

One of rationales is that local structural worsening in parenchyma with repetitive injury or acute lung disease leads to lung functional deterioration. It is a challenge to find a metric or threshold to determine the local decline [10, 11]. Recently, one study showed that early changes in lung function were not predictive of progression in lung function in the subsequent 6–12 months, bringing attention to the need for a better predictor of declining lung function [12, 13]. Although gender, age, and pulmonary function (GAP) stage and the loss of FVC are one of the prognostic predictors of survival in patients with IPF, we have not yet investigated whether the early changes on GAP scores predict progression-free survival [5–7, 13–15]. The UIP pattern with both HRCT and histological diagnosis at a single time point has been shown to predict a shortened survival in patients with IPF compared with atypical HRCT findings and estimate 3.1–3.3 years of median follow-up, but UIP patterns are unlikely changed over time [16].

Quantitative interstitial scores have been developed and implemented in the past decade [17–20]. Robust HRCT technical factors and variability of texture feature have been identified to obtain quantitative scores from multi-center and multi-manufactured HRCT scans [21]. Automated quantitative analysis of radiological patterns showing the prediction of survival by disease severity at baseline demonstrates the usefulness of extent of fibrotic patterns or total interstitial lung disease with the median of 2.4 years or mean 2-year follow-up [17, 18]. In addition, changes in visual assessment and quantitative changes were associated with predicting survival or changes in FVC [18, 22]. Although the prognostic factors and the changes in quantitative scores demonstrated worsening in function in the later time points, we have not yet systematically shown a threshold-based approach or minimal radio-graphically important difference with respect to visual assessment to predict rapid progression of IPF. Our study is to evaluate the early changes in imaging for predicting clinical progression in patients with IPF using a threshold from visual confirmation.

Methods

Subjects

Pilot set

Volumetric HRCT of 3 sequential scans 6 months apart were collected from subjects who were diagnosed with IPF by the American Thoracic Society (ATS)/European Respiratory Society (ERS)/Japanese Respiratory Society (JRS)/Latin American Thoracic Society (ALAT) criteria and enrolled as part of clinical trials [1]. The anonymized scans were obtained in thin slices (≤ 2 mm) and all in prone position. Two radiologists locked-sequentially assessed the longitudinal changes visually (i.e., 6 months and 12 months) from baseline HRCT scans with the ordinal scales of better, same, or worse in each lobe. Adjudication was performed if the assessment of changes was not agreed upon in worse and not-worse fibrotic reticulation. Visual assessment of a paired scan is defined as stable if the assessments of all the lobes are stable or better. Otherwise, visual assessment is defined as worse if at least one lobe is worse. HRCT scans from 35 subjects were chosen from the retrospective dataset: (1) visual assessments of changes were stable at 6 months and the visual assessments at 12 months were available; and (2) longitudinal percent predicted FVC from pulmonary function tests (PFT) were

available, whose changes were used for determining the progression [2].

Dataset for validation

Subjects (mean ± SD age 69.3 ± 7.4 years) were diagnosed with IPF by the ATS/ERS/JRS/ALAT criteria and enrolled as part of clinical trials from February 2011 to October 2015 in either investigational treatments or placebo, but not as part of either pirfenidone (Esbriet, Genentech) or nintedanib (Ofev, Boehringer Ingelheim) [1–3]. Two hundred fifty-five anonymized paired HRCT scans were imported. Out of 255, 157 subjects have met two conditions: (a) at least one follow-up HRCT scan; (b) available data of PFT after 6-month HRCT follow-up. The duration of median (± interquartile ranges) follow-up was approximately 1.1 (± 1.0) years after the second CT scan. Paired HRCT scans were volumetric with thin-slice thickness (≤ 2 mm) and high-contrast reconstruction mostly in prone position (91.7% of scans) (of note that 5.1% of scans were in the supine position for all time points and 3.2% were scanned in prone position at baseline and supine position at 6 months). The duration of follow-up was limited prior to the lung transplantation or changes in treatments. FVC measurements were used to determine the progression when the percent predicted FVC reduced by 10% or more from baseline [2].

Among anonymized paired HRCT inspiration (total lung capacity (TLC)) scans from 255 subjects (189 males, 75%), the mean (±SD) QLF score was 17.2% (± 9.6) in the whole lung and 31.2% (± 16.5) at the most severe lobe at baseline. Mean (± SD) age was 69.3 years (± 7.4) for 185 subjects who had available PFT: mean (± SD) was 67.8% (± 13.0) for the percent predicted FVC and 45.6% (± 13.3) for the percent predicted diffusing capacity of the lungs for carbon monoxide (DLCO), respectively. In total, 157 subjects had no missing GAP score or QLF data who are mostly male 70 years old with mean 70% predicted FVC, 46% of DLCO, and 16% QLF in the whole lung (Table 2). Out of 157, 23 subjects' data could not be used who had progression by FVC at the 1st follow-up HRCT scan. In this final cohort, 134 subjects were used who had not progressed at the 1st HRCT follow-up. Using the landmark of 6 months, the duration of median follow-up was 19 months (inter-quartile range > 12 months), which led to approximately 24-month follow-ups.

CT image analysis

We have previously developed the quantitative lung fibrosis (QLF) and quantitative ILD (QILD) patterns with the classification rate of 95% and 94%, respectively, with the repeatability of coefficient of 0.4% in the whole lung and 2% at the most severe lobe by sampling the neighboring points in a grid [19, 21]. The pipeline process includes (1) lobar segmentation, (2) adaptive denoise, (3) classification using texture features, (4) summation of the voxels with classified LF and ILD patterns, and (5) production of a ratio (%) of the counts of voxels that were classified as LF or ILD to the total counts of voxels in the whole lung [20]. Adaptive denoise is implemented to deal with different noise characteristics across conventional CT for multi-regional studies [19]. We calculated the quantitative scores (%) in the whole lung, as well as in each of the 5 lobes, from baseline and the first follow-up HRCT scans. The most severe lobe (i.e., the highest QLF score) at baseline frequently occurred in the lower lobes (for details, at 50.64% in right lower, 25.64% in left lower, 10.26% in right middle, 9.62% in left upper, and 3.85% in right upper lobes). In addition, we also calculated the quantitative score in volume (ml) by multiplying the QLF score (%) and the volume of the whole lung (ml).

Statistical analysis

Part (1) in the evaluation, descriptive statistics in the mean change in QLF scores were reported by two groups of visual assessments (a) stable for both 6 and 12 months in all lobes and (b) stable at 6 months followed by worsening at least in one lobe at 12 months. In this pilot dataset, we estimated a threshold based on the mean of 6-month QLF changes for the subjects whose visual assessments were stable at 6 months and worse in lobes at 12 months. C-indexes were reported for the fitted Cox models to estimate the prediction performance of the models based on the threshold of 4% QLF changes, where the progression was based on 10% or more reduction in FVC.

Part (2) in the evaluation, the primary outcome was the early quantitative radiological changes in the most severe lobe as a dichotomized scale using the threshold from the pilot dataset. The secondary outcome was a dichotomized QLF changes in the whole lung using the same threshold. In addition, exploratory outcomes were GAP scores, and quantitative imaging outcome. The primary analysis was to compare two groups whose QLF changes were more than or less than the threshold in predicting the progression, based on 10% or more reduction FVC, using the landmark Cox regressions with a covariate of GAP stage at the most severe lobe [14]. C-index was reported to estimate the concordance of prediction rate. Secondary analysis was to perform using the same test in the whole lung. As part of exploratory analyses, the Kaplan-Meier curves were graphed and log-rank tests and univariate Cox regression were performed in all groups. In addition, sensitivity analyses were run for mild and moderate groups (GAP scores 0–5) for checking consistency [14]. C-indexes were reported for the fitted Cox models to estimate the prediction performance based on the threshold of 4% QLF changes and other covariates. Additionally, summary statistics in the changes of QLF in volumes (ml) were reported with 95%

confidence interval. The family-wise error rate was controlled to a significance level of 0.05 for the primary analysis. Statistical analyses were performed by Stata (version 14.0; StataCorp, College Station, TX).

Results

Pilot set

Mean QLF scores (± standard error (SE)) in the most severe lobe increased by 4.33% (± 2.62) at 6 months and 14.83% (± 4.18) at 12 months, when the visual assessments were stable at 6 months and then worse at 12-month scans. In contrast, the mean changes of QLF (± SE) were 0.74% (± 1.08) at 6 months and 1.22% (± 1.38) at 12 months for the stable group by visual assessments at both 6 and 12 months (see Table 1). Stable changes were defined as < 4% changes at the most severe lobe, and rapid worsening group was defined as ≥ 4% QLF changes at the first follow-up. The 4% changes of QLF at 6 months were considered an important threshold for predicting progression by PFT in patients with IPF (p = 0.0433 by log-rank test; HZ = 3.9, p = 0.017, C-index = 0.66). Mean changes in QLF scores (± SE) in the whole lung were 1.1% (± 1.71) at 6 months and 7.67% (± 2.39) at 12 months for the subjects with more than 4% changes in the worst lobe at 12 months. To use a conservative threshold, we applied 4% thresholds in the most severe lobe and whole lung.

In addition, mean volume (± SE) derived from HRCT scan was 3790 ml (± 148) at baseline. By group, the mean volume (± SE) at baseline was 3415 ml (± 167) for subjects stable at 6 months and worse at 12 months, and 3986 ml (± 271) for subjects stable for 12 months. Mean QLF in the whole lung volume (ml) (± SE; 95% CI) increased by 41 ml (± 39; [− 45, 128 ml]) at 6 months and 177 ml (± 51; [65, 291 ml]) at 12 months for the subjects who were stable at 6 months and worse at 12 months, whereas the changes in the mean (± SE; 95% CI) were modest as 1.0 ml (± 26; [−52, 54 ml]) and 9.0 ml (± 27; [−48, 65 ml]) at 6 and 12 months, respectively, for the subjects who were stable for 12 months.

Table 1 QLF score at baseline and changes by visual assessments at baseline, 6 months, and 12 months in pilot study

	Visual assessment from baseline	
	6 months → 12 months Same → Same (N = 23)	6 months → 12 months Same → Worse (N = 12)
Most severe lobe	QLF*: mean (± SE)	QLF*: mean (± SE)
Baseline	29.43% (± 3.96)	38.42% (± 3.77)
	Changes in QLF (%)* Mean (± SE)	Changes QLF (%)* Mean (± SE)
6 months–baseline	0.74% (± 1.08)	4.33% (± 2.62)
12 months–baseline	1.22% (± 1.38)	14.83% (± 4.18)

*Quantitative lung fibrosis score (QLF) at the most severe lobe at baseline

Evaluation study

Increased QLF score of more than 4% in the most severe lobe had approximately 6-fold higher risk of progression than < 4% changes for subjects with IPF with adjusting the GAP stage (HR = 5.92; p = 0.001, C-index = 0.71; Table 3). QLF changes in the whole lung using 4% threshold also show a significant difference in predicting progression by multivariate analysis (HR = 3.32; p = 0.012, C-index = 0.68; Table 3).

In the exploratory analyses, subjects with ≥ 4% change in QLF at the most severe lobe had significantly greater risk of progression compared with patients with < 4% changes (p = 0.0001 by log-rank test), and findings were similar based on changes in the whole lung (p = 0.0088 by log-rank test) (Fig. 1). The median PFS of subjects with ≥ 4% changes was 11.9 months, whereas that of subjects with < 4% QLF changes was > 18 months. Subjects with increased QLF score of ≥ 4% in the most severe lobe had 4–5-fold risk of progression in the next incident comparing with subjects with < 4% change (HR = 4.81; p < 0.001, C-index = 0.66) by univariate analyses. GAP index shows a consistent trend in predicting progression where GAP scores were 5 or less (Table 3). Two examples of cases were reported in Figs. 2 and 3. The first example is a stable case over 18 months (Fig. 2a–d). Changes in 6 months were within 1% and maintained at 12% of QLF scores. The latter case was progressed after 1.5 years whose QLF score increased 7% in the most severe lobe and 5% in the whole lung, respectively, from baseline to 6 months (Fig. 3a–c for 328 HRCT images and Fig. 3d–f for annotated images).

In summary, mean volume (± SE) in the whole lung from HRCT was 3887 ml (± 78) in TLC and 605 ml (± 26) in QLF at baseline. At 6 months, mean volume in the whole lung (± SE) reduced by 137 ml (± 33) in TLC and increased by 63 ml (± 13) in QLF. By group, mean volume in the whole lung (± SE; 95% CI) decreased by 328 ml (± 47; [− 420, − 234 ml]) and QLF volume increased by 176 ml (± 15; [144, 207 ml]) for the subjects who increased ≥ 4% QLF score in the most severe lobe. In contrast, mean changes in (± SE; 95% CI) the whole lung and QLF volume were modest as 12 ml (± 40; [− 67, 91 ml]) and − 25 ml (± 12; [− 49, 0 ml]), respectively, for the subjects whose QLF changes were < 4% in the most severe lobe.

Discussion

Local structural changes of the fibrotic reticulation in parenchyma are one of the predictive explanations for functional progression in the next follow-up in subjects with IPF.

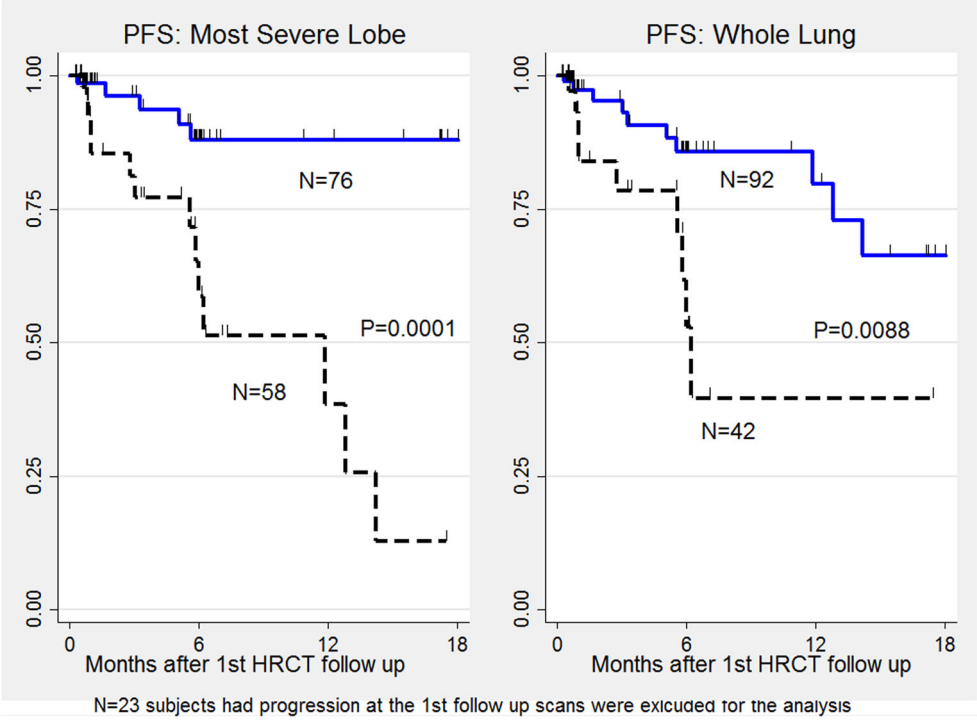

Fig. 1 The Kaplan-Meier progression-free survival based on the quantitative lung fibrosis (QLF) scores (QLF < 4% vs. ≥ 4% changes at the 1st follow-up) at the most severe lobe (**a**) at baseline and (**b**) in the whole lung ($N = 134$). Out of 157 subjects, 23 had progression defined by pulmonary function test at the first imaging follow-up. $^{*}p < 0.0001$ and $^{\#}p < 0.0088$ representing the association of early changes in QLF between subjects with ≥ 4% and < 4% changes in the worst lobe and in the whole lung, respectively

Functional change is a global measurement that determines progression, whereas radiological outcome derived from texture features can locally identify specific patterns of disease. In the previous studies, the robust structural worsening on HRCT can play a role of surrogate outcomes to support the stability or progression, although prior PFT functional changes were not predicative for the next follow-up visits [12, 13]. Most patients with IPF have ongoing changes in the lung that lead to a persistent decline in lung function over time [2, 3]. The result of our study also supports that IPF is a progressive-fibrosing ILD with declining lung functions, also appearing at 6 months [23]. The mean of lung volume by HRCT decreased by 138 ml (3.5%), while the mean of QLF volume increased by 63 ml (10.4%). Interestingly, the changes of accelerated activation are likely to occur in the worst areas of the IPF lung. Thus, the most structurally severe lobe, probably the worst functional lobe, develops subtle architectural changes such as alveolar collapse and increasing reticulation, which can be measured sensitively by a quantitative change to predicting the disease progression. The concordance index was numerically higher in the most severely affected lobe than in the whole lung (C-index = 0.71 vs. 0.68, respectively in multivariate analyses). This study demonstrated that early quantitative changes on HRCT can predict the functional progression in the next visit.

Physiological changes are highly predictive to survival [5, 6]. However, the remaining challenge is to optimize the therapeutic window for a subject with IPF. Our study supports that early physiological changes on HRCT during the initial therapy can provide the confirmation of staying treatment or information of changes in treatment planning. The efficacy of recently approved active treatment with or without honeycombing patterns in patients with IPF was not different in progression [24]. Owing to the benefit from therapeutic options and the limited window of treatment of 2–5 years, we want to expand the comprehensive outcome measurements of physiological lung function changes, the acute exacerbation, clinical information, and laboratory tests for an informed decision for a patient's next clinical visits [2, 3, 25].

This study suggests that early changes from quantitative fibrotic reticulation using a threshold-based approach can be used as a personalized tool for precision at the

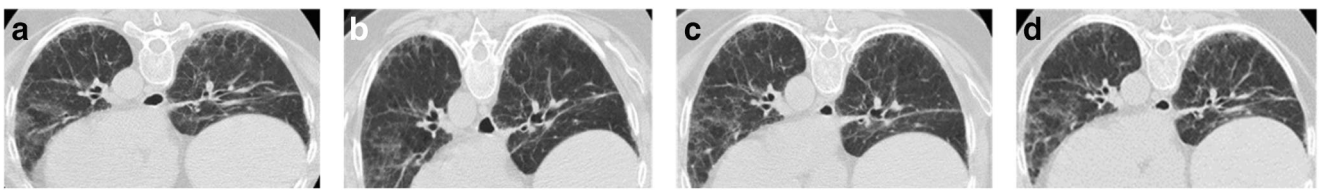

Fig. 2 A stable case by PFT for 18 months from validation set: **a**–**d** a 65-year-old female with a whole-lung QLF score of 12% at baseline and no changes for 18 months; (**a**) baseline, (**b**) 6 months, (**c**) 1 year, and (**d**) 1.5 years of follow-up HRCT scans

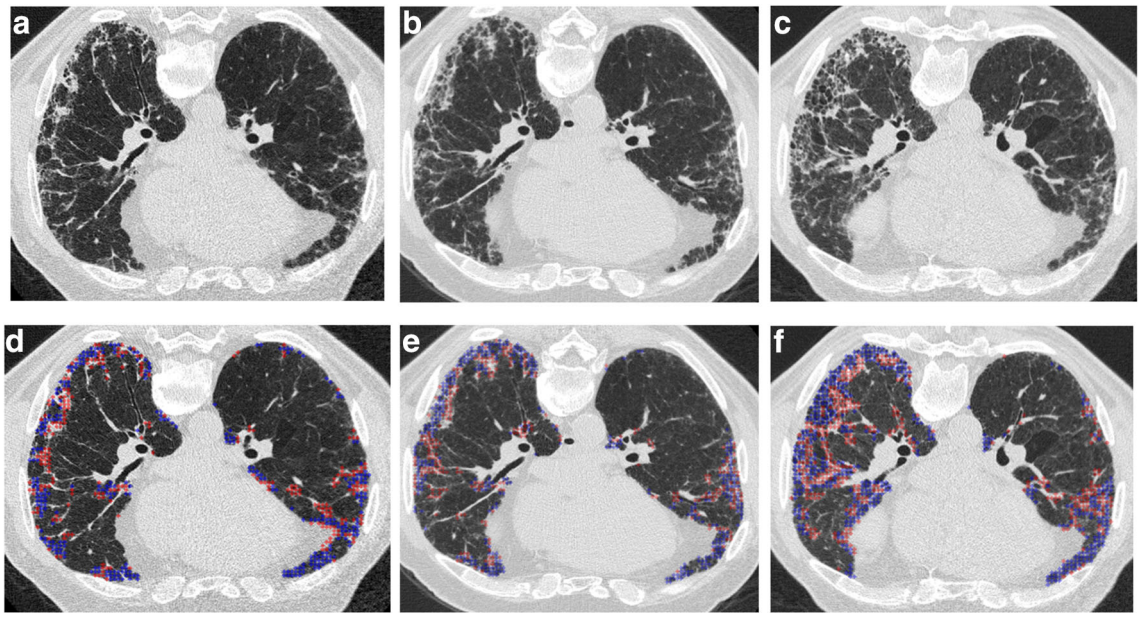

Fig. 3 A case *from validation set* with increase of 5% QLF at 6 months and progression by PFT at year 1: **a**–**c** a 74-year-old male at baseline, 6-month, and 1-year follow-up whose percent predicted FVC was 80%, 78%, and 65% at baseline, 6 months, and 12 months, respectively; (**a**) baseline, (**b**) 6 months, and (**c**) 1 year; **d**–**f** annotated images of **a**–**c**, respectively, with overlaying QLF classification; blue + red circles = QLF; (**d**) 17% QLF score at baseline in the whole lung, (**e**) 22% QLF score at 6 months in the whole lung, and (**f**) 29% QLF score at year 1

individual patient level to enable more informed and timely management decisions. The initial pilot study shows the sensitive quantitative scores based on imaging features in patients with IPF demonstrating that early changes at 6 months differentiate the worsening of fibrosis by visual assessment at 12 months (Table 1). Using 4% threshold from visual assessment, the study in evaluation cohort shows that early changes in quantitative fibrosis scores at 6 months can predict the PFS at the later time points (Table 2, Fig. 1). Subjects with increasing ≥ 4% in quantitative fibrosis score at 6 months had 3–5 folds higher likelihood of experience progression with the median of 18 months compared with subjects < 4% changes (Table 3). By adjusting the clinical information of age, gender, and FVC, the concordance of prediction improved (Table 3). The threshold of 4% can be considered a conservative threshold referring to two published studies: (1) the reproducibility of coefficient in the grid sampling (2.77*standard deviation) was 2% at the most severe lobe and < 1% in the whole lung [21], and (2) a minimal clinically important difference (MCID) from two scleroderma ILD studies is reported as 2% at the most severe lobe [26].

Table 2 Baseline characteristic in validation study

	GAP scores 0–3 ($N = 53$) Mean (± SD)	GAP scores 4–5 ($N = 92$) Mean (± SD)	GAP scores 6–7 ($N = 12$) Mean (± SD)	Total ($N = 157$) Mean (± SD)
Age	65.2 (± 7.6)	71.7 (± 6.4)	71.6 (± 3.5)	69.5 (± 7.3)
Male	25 (47.2%)	76 (82.6%)	21 (87.5%)	113 (72.0%)
FVC % predicted	76.8 (± 11.7)	66.7 (± 10.4)	58.5 (± 7.0)	69.5% (± 12.8)
DLCO % predicted[+]	53.1 (± 16.8)	44.7 (± 11.3)	30.7 (± 4.7)	46.1% (± 14.4)
QLF at the most severe lobe (%)	22.1 (± 12.9)	33.6 (± 17.2)	36.8 (± 12.2)	30.0% (± 16.5)
QLF in the whole lung (%)	12.1 (± 7.0)	18.0 (± 9.1)	21.4 (± 6.3)	16.3 (± 8.8)
QILD in the whole lung (%)	30.3 (± 12.7)	38.8 (± 14.2)	45.6 (± 7.1)	36.1 (± 14.1)
Whole lung volume from HRCT (ml)	4022 (± 1200)	3815 (± 872)	3839 (± 614)	3887 (± 980)

+: $N = 123$ ($N = 40$ from GAP scores 0–3, $N = 71$ from GAP scores 4–5, and $N = 12$ from GAP scores 6–7) for non-missing DLCO and $N = 34$ for missing or *unable to perform* DLCO tests; *QLF*, quantitative lung fibrosis; *QILD*, quantitative interstitial lung disease (i.e., total abnormality)

Table 3 Cox regression models from univariate and multivariate analyses

Baseline covariates	All groups ($N = 157$)		
	Hazard ratio (SE)	p value	C-index
GAP index	1.43 (.53)	0.335	0.59
Baseline covariates and 6-month changes	All groups with non-progression at 6 month follow-up ($N = 134$)		
Most severe lobe	Hazard ratio (SE)	p value	C-index
QLF ≥ 4%	4.81 (2.15)	< 0.001	0.66
GAP index	1.33 (0.49)	0.443	0.71
QLF ≥ 4%	5.92 (3.11)	*0.001*	
Whole lung	Hazard ratio (SE)	p value	C-index
QLF ≥ 4%	3.25 (1.53)	0.013	0.66
GAP index	1.47 (0.60)	0.337	0.68
QLF ≥ 4%	3.32 (1.58)	*0.012*	
Baseline covariates	GAP 0–3 and 4–5 ($N = 145$)		
	Hazard ratio (SE)	p value	C-index
GAP index	2.43 (1.37)	0.115	0.63
Baseline covariates and 6-month changes	GAP 0–3 and 4–5 with non-progression at 6-month follow-up ($N = 121$)		
Most severe lobe	Hazard ratio (SE)	p value	C-index
QLF ≥ 4%	5.50 (2.92)	0.001	0.67
GAP index	2.50 (1.42)	0.108	0.73
QLF ≥ 4%	5.77 (3.07)	0.001	
Whole lung	Hazard ratio (SE)	p value	C-index
QLF ≥ 4%	3.20 (1.54)	0.016	0.62
GAP index	2.14 (1.21)	0.183	0.70
QLF ≥ 4%	3.08 (1.48)	0.019	

GAP gender, age, and pulmonary function; *QLF*, quantitative lung fibrosis; *C-index*, concordance index after fitting a Cox regression model; the result of primary analyses were included in the solid box where multivariate Cox models were performed with QLF ≥ 4% adjusting GAP index at the most severe lobe and in the whole lung. The results in italic were from the primary outcomes (most severe lobe) and secondary outcomes (whole lung)

The result should be interpreted within the context of certain limitation. Firstly, the threshold of 4% change reflects the visual assessment of progression in 6 months later (i.e., mean QLF 4% changes at 6 months correspond to stable, but worsening at 12 months by visual assessment). The threshold of 4% can be a clinically important difference, but not necessarily MCID. The threshold is based on the predictability of visual assessment using the mean changes of QLF prior to the visual confirmation of worsening, where radiologists scored at each lobe. The 4% changes in the worst lobe reflect approximately 1–2% changes in the whole lung due to the heterogeneous interstitial distribution in lower lobes. For simplicity, we used a 4% threshold at both of the most severe lobe and whole lung. This score was based on one technique among many other available algorithms [17–19, 27–29]. Secondly, it was retrospectively collected and anonymized HRCT images. Diagnosis of IPF criteria was based on the time of data collection [1, 30]. Although quantitative scores from HRCT were available, other uncollected clinical outcomes could influence progression such as ethnicity, smoking status, treatment options, dyspnea, and 6-min walk. Pulmonary functions were not necessarily collected in uniform time intervals. Subjects with higher GAP scores may have dropped out early for transplantation. This is consistent that GAP score does not predict the rate of changes in PFT [13]. Especially there are limited sample sizes and duration of follow-up times in subjects, where most of subjects had lost follow-up rapidly after the 2nd HRCT scans and may move to other treatments, which are the exclusion criteria (i.e., switching a therapy) in this study. Checking the standardization of full inspiration was a challenge. Consistent scanning position either in supine or in prone is important for changes in ILD assessments. Although there are limited studies regarding the repeatability or reproducibility, it has been known that the gravity effect on HRCT images and the gradient of quantitative values influence the intensities in a normal parenchyma [31]. In this study, most of subjects had consistent table positions. We included 5 subjects who had inconsistent positions. Five subjects underwent in prone position at baseline and followed by the supine position at 6 months. Two subjects had more than 4% QLF changes and the rest had less than 4% change. Additionally, we have not collected symptom scores, which can be valuable to test "subclinical" changes that may precede symptoms or not, although there are no specific symptom scores yet for IPF

population [5]. Thirdly, due to the frequent and unidentifiable causes of missing DLCO data (e.g., patient's worsening vs. the technical failure), we used the FVC % predicted changes to determine clinical progression. However, in the setting of retrospective study, the cause of missing DLCO is difficult to identify. It was challenging to find comments in PFT reports for distinguishing a patient's worsening from the technical failure. Fourthly, the duration of follow-up of subjects was mostly from 6 months or 7 months which may not happen in the current clinical follow-ups. The longitudinal changes on HRCT images mostly are used for the confirmation of structural changes in the sign of progression.

Conclusion

Using imaging features of quantitative lung fibrosis scores in patients with IPF, the early changes in quantitative fibrosis scores at 6 months can predict the PFS later time points. With a threshold-based approach, the model can be useful at the individual patient level to enable precisely informed and timely management decisions after taking 6 months of treatment. Furthermore, it can be used for future research to define homogeneous cohorts for testing new safe and effective therapies and to elucidate the effects of therapies in patients with biologically heterogeneous disease in the timing of progression.

Acknowledgments The authors gratefully acknowledge the National Heart, Lung, and Blood-National Institutes of Health for the support (NIH-NHLBI R21HL123477-01A1 with the title of Prediction of IPF Progression Using HRCT Imaging Patterns).

Funding This study received support from the National Heart, Lung, and Blood-National Institutes of Health (NIH-NHLBI R21HL123477-01A1 with the title of Prediction of IPF Progression Using HRCT Imaging Patterns). This study is part of R21 studies focused on "Secondary analyses of existing datasets in Heart, Lung, and Blood Disease and Sleep Disorder from NHLBI."

Compliance with ethical standards

Guarantor The scientific guarantor of this publication is Grace Hyun Kim.

Conflict of interest The authors of this manuscript declare no relationships with any companies whose products or services may be related to the subject matter of the article.

Statistics and biometry Grace Hyun Kim is trained in Biostatistics with MS and PhD degree and also an associate professor at the Department of Radiology. Yu Shi, a PhD student at Biostatistics, worked on data management and basic statistical summary statistics. Grace Hyun Kim provided statistical advice for this manuscript.

To our knowledge, no complex statistical methods were necessary for this paper.

Informed consent Written informed consent was not required for this study because this is a retrospective study and secondary date usage.

Approval from the institutional animal care committee was not required because this study is based on imaging only from human.

Ethical approval Institutional Review Board approval was obtained.

Methodology
- Retrospective
- Observational
- Performed at one institution

References

1. Raghu G, Collard HR, Egan JJ et al (2011) An official ATS/ERS/JRS/ALAT statement: idiopathic pulmonary fibrosis: evidence-based guidelines for diagnosis and management. Am J Respir Crit Care Med 183(6):788–824
2. King TE Jr, Bradford WZ, Castro-Bernardini S et al (2014) A phase 3 trial of pirfenidone in patients with idiopathic pulmonary fibrosis. N Engl J Med 370(22):2083–2092
3. Richeldi L, du Bois RM, Raghu G et al (2014) Efficacy and safety of nintedanib in idiopathic pulmonary fibrosis. N Engl J Med 370(22):2071–2082
4. Du Bois RM, Weycker D, Albera C et al (2011) Ascertainment of individual risk of mortality for patients with idiopathic pulmonary fibrosis. Am J Respir Crit Care Med 184(4):459–466
5. Collard HR, King TE Jr, Bartelson BB, Vourlekis JS, Schwarz MI, Brown KK (2003) Changes in clinical and physiologic variables predict survival in idiopathic pulmonary fibrosis. Am J Respir Crit Care Med 168(5):538–542
6. Hanson D, Winterbauer RH, Kirtland SH, Wu R (1995) Changes in pulmonary function test results after 1 year of therapy as predictors of survival in patients with idiopathic pulmonary fibrosis. Chest. 108(2):305–310. https://doi.org/10.1378/chest.108.2.305
7. Ley B, Collard HR, King TE Jr (2011) Clinical course and prediction of survival in idiopathic pulmonary fibrosis. Am J Respir Crit Care Med 183(4):431–440
8. Lynch DA, Godwin JD, Safrin S et al (2005) High-resolution computed tomography in idiopathic pulmonary fibrosis: diagnosis and prognosis. Am J Respir Crit Care Med 172(4):488–493
9. Sumikawa H, Johkoh T, Colby TV et al (2008) Computed tomography findings in pathological usual interstitial pneumonia: relationship to survival. Am J Respir Crit Care Med 177(4):433–439
10. Loomis-King H, Flaherty KR, Moore BB (2013) Pathogenesis, current treatments and future directions for idiopathic pulmonary fibrosis. Curr Opin Pharmacol 13(3):377–385
11. Martinez FJ, Safrin S, Weycker D et al (2005) The clinical course of patients with idiopathic pulmonary fibrosis. Ann Intern Med 21(142):963–967
12. Nathan SD, Albera C, Bradford WZ et al (2016) Effect of continued treatment with pirfenidone following clinically meaningful declines in forced vital capacity: analysis of data from three phase 3 trials in patients with idiopathic pulmonary fibrosis. Thorax. 71(5):429–435
13. Salisbury ML, Xia M, Zhou Y et al (2016) Idiopathic pulmonary fibrosis: gender-age-physiology index stage for predicting future lung function decline. Chest. 149(2):491–498
14. Ley B, Ryerson CJ, Vittinghoff E et al (2012) A multidimensional index and staging system for idiopathic pulmonary fibrosis. Ann Intern Med 156(10):684–691
15. Flaherty KR, Andrei A, Murray S et al (2006) Idiopathic pulmonary fibrosis: prognostic value of changes in physiology and six-minute-walk test. Am J Respir Crit Care Med 174(7):803–809

16. Flaherty KR, Thwaite EL, Kazerooni EA et al (2003) Radiological versus histological diagnosis in UIP and NSIP: survival implications. Thorax. 58(2):143–148
17. Iwasawa T, Asakura A, Sakai F et al (2009) Assessment of prognosis of patients with idiopathic pulmonary fibrosis by computer-aided analysis of CT images. J Thorac Imaging 24(3):216–222
18. Maldonado F, Moua T, Rajagopalan S et al (2014) Automated quantification of radiological patterns predicts survival in idiopathic pulmonary fibrosis. Eur Respir J 43(1):204–212
19. Kim HJ, Tashkin DP, Clements PJ et al (2010) A computer-aided diagnosis system for quantitative scoring of extent of lung fibrosis in scleroderma patients. Clin Exp Rheumatol 28:S26–S25
20. Kim HJ, Brown MS, Chong D et al (2015) Comparison of the quantitative CT imaging biomarkers of idiopathic pulmonary fibrosis at baseline and early change with an interval of 7 months. Acad Radiol 22(1):70–80
21. Kim HJ, Brown MS, Elashoff R et al (2011) Quantitative texture-based assessment of one-year changes in fibrotic reticular patterns on HRCT in scleroderma lung disease treated with oral cyclophosphamide. Eur Radiol 21(12):2455–2465
22. Raghu G, Scholand MB, de Andrade J et al (2016) FG-3019 anti-connective tissue growth factor monoclonal antibody: results of an open-label clinical trial in idiopathic pulmonary fibrosis. Eur Respir J 47(5):1481–1491
23. Cottin V, Hirani NA, Hotchkin DL et al (2018) Presentation, diagnosis and clinical course of the spectrum of progressive-fibrosing interstitial lung diseases. Eur Respir Rev 27(150). https://doi.org/10.1183/16000617.0076-2018
24. Raghu G, Wells AU, Nicholson AG et al (2017) Effect of nintedanib in subgroups of idiopathic pulmonary fibrosis by diagnostic criteria. Am J Respir Crit Care Med 195(1):78–85
25. Raghu G, Rochwerg B, Zhang Y (2015) An official ATS/ERS/JRS/ALAT clinical practice guideline: treatment of idiopathic pulmonary fibrosis. An update of the 2011 clinical practice guideline. Am J Respir Crit Care Med 192(2):e3–e19
26. Kafaja S, Clements PJ, Wilhalme H et al (2018) Reliability and minimal clinically important differences of forced vital capacity: results from the Scleroderma Lung Studies (SLS-I and SLS-II). Am J Respir Crit Care Med 197(5):644–652
27. Wu X, Kim GH, Salisbury ML et al (2019) Computed tomographic biomarkers in idiopathic pulmonary fibrosis. The future of quantitative analysis. Am J Respir Crit Care Med 199(1):12–21
28. van Royen FS, Moll SA, van Laar JM, van Montfrans JM, de Jong PA, Mohamed Hoesein FAA (2019) Automated CT quantification methods for the assessment of interstitial lung disease in collagen vascular diseases: a systematic review. Eur J Radiol 112:200–206
29. Lee SM, Seo JB, Oh SY et al (2018) Prediction of survival by texture-based automated quantitative assessment of regional disease patterns on CT in idiopathic pulmonary fibrosis. Eur Radiol 28:1293–1300
30. Wells AU (2018) IPF diagnosis: flexibility is a virtue. Lancet Respir Med 6(10):735–737
31. West JB (1977) State of the art: ventilation-perfusion relationships. Am Rev Respir Dis 116:919–943