Idiopathic Pulmonary Fibrosis: The Association between the Adaptive Multiple Features Method and Fibrosis Outcomes

Margaret L. Salisbury¹, David A. Lynch², Edwin J. R. van Beek³, Ella A. Kazerooni⁴, Junfeng Guo⁵, Meng Xia⁶, Susan Murray⁶, Kevin J. Anstrom^{7,8}, Eric Yow⁸, Fernando J. Martinez⁹, Eric A. Hoffman⁵, and Kevin R. Flaherty¹; for the IPFnet Investigators

¹Division of Pulmonary and Critical Care Medicine, Department of Medicine, ⁴Department of Radiology, and ⁶Department of Biostatistics, University of Michigan, Ann Arbor, Michigan; ²Department of Radiology, National Jewish Health, Denver, Colorado; ³Centre for Inflammation Research, Clinical Research Imaging Centre, University of Edinburgh, Edinburgh, United Kingdom; ⁵Department of Radiology, University of Iowa, Iowa City, Iowa; ⁷Department of Biostatistics and Bioinformatics, Duke University, Durham, North Carolina; ⁸Duke Clinical Research Institute, Duke Medical Center, Durham, North Carolina; and ⁹Division of Pulmonary and Critical Care Medicine, Department of Medicine, Cornell Medical College, New York, New York

ORCID ID: 0000-0001-8217-6955 (M.L.S.).

Abstract

Rationale: Adaptive multiple features method (AMFM) lung texture analysis software recognizes high-resolution computed tomography (HRCT) patterns.

Objectives: To evaluate AMFM and visual quantification of HRCT patterns and their relationship with disease progression in idiopathic pulmonary fibrosis.

Methods: Patients with idiopathic pulmonary fibrosis in a clinical trial of prednisone, azathioprine, and N-acetylcysteine underwent HRCT at study start and finish. Proportion of lung occupied by ground glass, ground glass–reticular (GGR), honeycombing, emphysema, and normal lung densities were measured by AMFM and three radiologists, documenting baseline disease extent and postbaseline change. Disease progression includes composite mortality, hospitalization, and 10% FVC decline.

Measurements and Main Results: Agreement between visual and AMFM measurements was moderate for GGR (Pearson's correlation $r = 0.60$, $P < 0.0001$; mean difference = -0.03 with 95% limits of

agreement of -0.19 to 0.14). Baseline extent of GGR was independently associated with disease progression when adjusting for baseline Gender-Age-Physiology stage and smoking status (hazard ratio per 10% visual GGR increase = 1.98 , 95% confidence interval [CI] = 1.20 – 3.28 , $P = 0.008$; and hazard ratio per 10% AMFM GGR increase = 1.36 , 95% CI = 1.01 – 1.84 , $P = 0.04$). Postbaseline visual and AMFM GGR trajectories were correlated with postbaseline FVC trajectory ($r = -0.30$, 95% CI = -0.46 to -0.11 , $P = 0.002$; and $r = -0.25$, 95% CI = -0.42 to -0.06 , $P = 0.01$, respectively).

Conclusions: More extensive baseline visual and AMFM fibrosis (as measured by GGR densities) is independently associated with elevated hazard for disease progression. Postbaseline change in AMFM-measured and visually measured GGR densities are modestly correlated with change in FVC. AMFM-measured fibrosis is an automated adjunct to existing prognostic markers and may allow for study enrichment with subjects at increased disease progression risk.

Keywords: idiopathic pulmonary fibrosis; multidetector computed tomography; prognosis

(Received in original form July 6, 2016; accepted in final form October 18, 2016)

Supported by National Institutes of Health/NHLBI grants U10 HL080371, R01 HL091743, T32 HL007749-21, and K24 HL111316.

Author Contributions: D.A.L., E.J.R.v.B., E.A.K., J.G., F.J.M., E.A.H., and K.R.F. conceptualized the work and participated in data analysis and acquisition; M.L.S., M.X., S.M., K.J.A., and E.Y. participated in data analysis and acquisition; M.L.S. and K.R.F. drafted the manuscript; D.A.L., E.J.R.v.B., E.A.K., J.G., M.X., S.M., K.J.A., E.Y., F.J.M., and E.A.H. critically revised the manuscript; all authors approved the final version to be published and take accountability for the work.

Correspondence and requests for reprints should be addressed to Margaret L. Salisbury, M.D., Division of Pulmonary and Critical Care Medicine, Department of Medicine, University of Michigan, 1500 E. Medical Center Drive, 3916 Taubman Center, Ann Arbor, MI 48109. E-mail: msalisbu@med.umich.edu

This article has an online supplement, which is accessible from this issue's table of contents at www.atsjournals.org

Am J Respir Crit Care Med Vol 195, Iss 7, pp 921–929, Apr 1, 2017

Copyright © 2017 by the American Thoracic Society

Originally Published in Press as DOI: 10.1164/rccm.201607-1385OC on October 21, 2016

Internet address: www.atsjournals.org

At a Glance Commentary

Scientific Knowledge on the

Subject: Software capable of automated characterization of lung findings on computed tomography scans have been developed but are not widely used in practice, for a variety of reasons.

What This Study Adds to the

Field: In this study, we evaluate the ability of adaptive multiple features method software to accurately characterize interstitial findings on computed tomography scan in patients with idiopathic pulmonary fibrosis. Adaptive multiple features method measurement of fibrosis extent is similar to that of expert radiologist measurements and is associated with increased hazard of disease progression (>10% decline in FVC), hospitalization, or death in patients with idiopathic pulmonary fibrosis. Patients with 10% or more of total lung volume occupied by fibrosis at baseline have substantially worse 60-week event-free survival compared to those with less than 10%.

Idiopathic pulmonary fibrosis (IPF) is a fibrosing interstitial lung disease (ILD) with poor prognosis. High-resolution computed tomography (HRCT) is integral in diagnosis and management of patients with IPF, and may be more sensitive to changes in disease status than clinical or physiologic parameters (1). Radiologists' visual interpretation of disease extent and change over time is hindered by variability, particularly in nonexpert observers (2–4). Thus, there is value in development of automated software that can reliably and objectively characterize type and extent of disease present on HRCT. Validation of software measures relies on comparison with a gold standard, as well as physiologic measures and clinical events. Previous studies have documented a wide range of agreement between software and visual measures (5–7). Histogram-based software measures of interstitial fibrosis correlate with pulmonary function measures in patients with IPF, but assessment for association of visual and software measures with outcomes such as mortality show mixed results in retrospective studies (6, 8, 9).

The adaptive multiple features method (AMFM) is an automated HRCT analysis program available through the University of Iowa's Department of Radiology (Iowa City, IA) Pulmonary Analysis Software Suite. Development for use on two- and three-dimensional images (using multidetector row CT highly spatially resolved volumetric images during a single breath hold) has previously been described (10–13). The three-dimensional AMFM method uses the best combination of 5–6 of 26 different mathematical features describing regional density patterns, along with a Bayesian classifier, and can be trained to recognize and quantify the volume occupied by a variety of radiologic patterns. This software has not been assessed with regard to agreement with visual analysis or association with pulmonary function and clinical outcomes in patients with IPF.

In this study, we used prospectively collected HRCT and clinical data from patients enrolled in the National Institutes of Health–sponsored IPFnet clinical therapeutic PANTHER-IPF (Prednisone, Azathioprine, and N-Acetylcysteine: a Study that Evaluates Response in IPF) trial (14, 15). We tested the relationship between semiquantitative visual scores and quantitative, automated AMFM scores, and determined how visual and AMFM scores at baseline and completion of the study relate to changes in pulmonary physiology and clinical outcomes, including pulmonary function decline, hospitalizations, and mortality. Some of the results of this study have been previously reported in the form of an abstract (16).

Methods

Subjects enrolling in the PANTHER-IPF trial were approached for participation in this ancillary study, and a separate written informed consent was obtained. The University of Michigan (Ann Arbor, MI) Institutional Review Board approved the protocol (HUM 24386). Any subject enrolling in the PANTHER-IPF study (14, 15) was eligible to participate, provided willingness to submit to an HRCT performed at study enrollment and again at completion of the study, and allow investigators access to PANTHER-IPF study data (patient characteristics, treatments administered, serial pulmonary

function measures, and clinical events, including hospitalizations, death, and confirmed declines in FVC > 10%). Pulmonary function testing was done locally at study sites, per American Thoracic Society guidelines, at screening, baseline (Week 0), and visits in Weeks 4, 15, 30, 45, and 60 as part of the PANTHER-IPF study (17, 18).

Image Acquisition and Visual Analysis

HRCT images were acquired locally on a variety of CT scanners; baseline scan was required at or within 3 months preceding PANTHER-IPF screening, and the follow-up scan was done at the end of the study period. Thin-slice inspiratory HRCT scans with contiguous images and without contrast were encouraged, but a more specific protocol (i.e., volumetric, edge enhancement, etc.) was not required. Summary characteristics of HRCTs (slice thickness and interval, scanner make and model, etc.) included in this study are available in Table E1 in the online supplement.

Three radiologists (E.J.R.v.B., E.A.K., and D.A.L.) independently interpreted each HRCT scan, blinded to patient characteristics, timing of the scan, and AMFM scores; data were recorded via a Web-based form. Each radiologist provided a semiquantitative extent of honeycombing, ground glass (GG), GG-reticular (GGR), and emphysema as: 0 (0% involvement of the lung with the feature); 1 (1–10% involvement); 2 (11–20%); 3 (21–30%); 4 (31–40%); 5 (41–50%); 6 (51–60%); 7 (61–70%); 8 (71–80%); 9 (81–90%); and 10 (91–100%) (19). The remainder was noted as normal. The midpoint of each category was arbitrarily used to convert the semiquantitative (0–10) to quantitative (0–100%) scores. For each HRCT, the percent involvement of each category for the right and left lungs was averaged to obtain a total lung score, and the three radiologists' total lung scores were averaged to obtain total volume occupied by each feature (the gold standard in this study). Scan quality was documented (excellent, good, poor, or uninterpretable); poor or uninterpretable scans and those with large gaps or missing images were excluded from analysis.

AMFM Software Analysis

The AMFM software training used HRCTs from patients with ILD enrolled in the Lung Tissue Research Consortium (www.ltrcpublic.com) and STEP-IPF

(Sildenafil Trial of Exercise Performance in Idiopathic Pulmonary Fibrosis) studies (20). Three expert radiologists independently assigned labels of honeycombing, GG, GGR, bronchovascular, emphysema, and normal to unlabeled volumes of interest (VOIs; *see* Figure 1 for GGR example and Figure E1 for examples of GG, normal, and honeycombing). AMFM software was trained using VOIs with complete independent agreement, as these were felt to represent the clearest examples of each pattern (rather than adjudicate patterns by consensus). Training and testing accuracy on a separate dataset of VOIs for software labeling of VOIs was 0.89 or greater (full train full test, half train half test, 10-fold cross-validation, leave-one-out cross-validation).

Trained AMFM software was then applied to each PANTHER-IPF CT having a compatible imaging protocol (in particular, contiguous image slices). Software output is a quantitative measure of the proportion of lung volume occupied by each feature based on the number of VOIs assigned to each feature category. Right and left lung feature scores were averaged to obtain a percent total lung involvement. Muscle kernel (MK) series AMFM data were used for all analyses; bone kernel series data were also evaluated and, due to excellent agreement in scores with MK (*see* Table E2), MK was used.

Statistical Analysis

Analyses were performed using SAS version 9.4 (SAS Institute, Cary, NC) and R version 3.2.1 (The R Foundation, Vienna, Austria). Baseline patient characteristics are shown

as number and percent for categorical variables, with continuous variables shown as mean (SD). Kaplan-Meier estimators of failure probability by 60 weeks are shown for important clinical events (death, hospitalization, physiologic decline defined as >10% FVC drop, and a composite of these events). Agreement between the three radiologists regarding extent of interstitial features is given via generalized weighted κ (sometimes called the S statistic), with κ less than 0.40 noted as poor, κ from 0.40 to 0.60 as intermediate, κ greater than 0.60 to 0.75 as good, and κ greater than 0.75 as excellent (21). Relationship between visual and AMFM scores is provided by Pearson's correlation coefficients (r values) and Bland-Altman plots with mean difference (visual minus AMFM score for each subject, averaged for all subjects) and 95% limits of agreement (22). Pearson's correlation is interpreted as follows: $0 < |r| < 0.3$ is weak, $0.3 < |r| < 0.7$ moderate, and $|r| \geq 0.7$ strong. Cox proportional hazards assess association between baseline visual and AMFM total-lung HRCT feature scores, clinical variables, and the disease progression composite outcome, as well as these outcomes individually. Existing literature suggests that age, sex, FVC % predicted, diffusing capacity of the lung for carbon monoxide (DL_{CO}) % predicted, the gender-age-physiology (GAP) stage (a score simultaneously accounting for patient age, sex, FVC, and DL_{CO} , with higher GAP stage associated with increased risk of death), and smoking history may be important clinical predictors of outcome and were thus screened for inclusion in a multivariable prediction

model (23, 24). The index of concordance (C statistic) was used to compare model discrimination between various models. Kaplan-Meier plots show event-free survival for subjects with visual and AMFM-measured GGR less than 10% and 10% or greater. For all subjects having a baseline and final scan, postbaseline change (final scan minus baseline scan) was calculated for each feature, with paired t test comparing differences between visual and AMFM measures of change. Pearson's correlation coefficient describes the relationship between postbaseline trajectory in HRCT scores and postbaseline FVC trajectory (calculated using linear regression slopes for each patient, incorporating all FVC data available to the point of the final HRCT scan).

Results

Baseline patient characteristics and clinical outcomes for all included subjects and subgroups of interest are shown in Table 1. Overall, the mean age was 67.0 (8.4) years, with 76.9% ($n = 153$) male sex, and 26.8% ($n = 53$) never smokers. In subjects having baseline HRCT scored by both visual and AMFM methods, the composite endpoint of greater than 10% decline in FVC, hospitalization, or death occurred in 49 subjects (60-wk Kaplan-Meier event rate, 34.7%; *see* Table 1, column 2). Figure E2 displays a flow diagram outlining the numbers of patients and HRCTs entering the study, and the numbers used in each analysis.

Agreement among radiologists about extent of the HRCT features is shown in Table E3, and ranged from a κ value of 0.70 (good agreement) for GGR to a κ value of 0.88 (excellent agreement) for emphysema (all $P < 0.0001$). Correlation, Bland-Altman mean difference, and 95% limits of agreement for visual and AMFM scores are shown in Table 2. The mean difference between scoring methods for GGR was -0.03 (95% limits = -0.19 to 0.14), indicating that AMFM measured 3% more (absolute difference) total lung with GGR abnormality on average compared with visual. We observed a tendency for AMFM to measure substantially more GG abnormality (mean difference = -0.19 ; 95% limits = -0.41 to 0.03) and less normal lung (mean difference = 0.26 ; 95% limits = -0.09 to 0.60) compared with visual scoring. Supplemental Figure E3 shows Bland-Altman plots for agreement

Figure 1. Expert radiologist labeling of volumes of interest for adaptive multiple features method training. Shown is an example of ground glass-reticular opacity. The boxed area within the lung parenchyma corresponds to the double-outlined box (upper right corner) in the associated black panel.

Table 1. Patient Characteristics at Baseline and Clinical Outcomes

Characteristic	All (n = 199)	Baseline HRCT (n = 156)	Sequential HRCT (n = 105)
Age, yr, mean (SD)	67.0 (8.4)	67.5 (8.3)	67.2 (8.7)
Male sex, n (%)	153 (76.9)	114 (73.1)	73 (69.5)
White race, n (%)	191 (96.0)	150 (96.2)	101 (96.2)
Never smoker, n (%)	53 (26.8)	46 (29.5)	30 (28.9)
Time since diagnosis, yr, mean (SD)	1.01 (1.1)	1.1 (1.1)	1.1 (1.1)
GAP stage, n (%)			
1	85 (43.2)	66 (42.3)	51 (48.6)
2	86 (43.7)	70 (44.9)	46 (43.8)
3	26 (13.2)	20 (12.8)	8 (7.2)
Coexisting illnesses, n (%)			
CAD	41 (20.6)	30 (19.2)	22 (21.0)
Diabetes	26 (13.1)	21 (13.5)	17 (16.2)
GERD	128 (64.3)	106 (68.0)	71 (67.2)
Lung function, mean (SD)			
FVC % predicted	73.0 (15.5)	73.5 (15.6)	75.0 (15.1)
DL _{CO} % predicted	46.1 (11.5)	45.7 (12.3)	47.1 (11.8)
Room air Pa _{O₂} , mm Hg	80.3 (11.1)	81.0 (11.3)	82.5 (10.1)
6-min-walk distance, m	370.5 (113.0)	370.1 (115.2)	378.2 (98.7)
Patient-reported outcomes scores, mean (SD)			
UCSD Shortness of Breath Questionnaire*	28.2 (19.0)	27.6 (18.0)	26.8 (15.6)
European Quality of Life score*	0.81 (0.2)	0.8 (0.2)	0.8 (0.2)
SF-36 aggregate physical score*	40.2 (9.2)	40.3 (9.2)	40.0 (8.9)
SF-36 aggregate mental score*	53.7 (9.3)	53.7 (9.5)	53.8 (9.1)
Observed outcomes by 60 wk follow-up, n (%) [†]			
All-cause death	7 (4.0)	4 (2.9)	0
FVC drop >10%	43 (25.3)	29 (21.9)	20 (19.7)
Hospitalization	33 (18.1)	28 (19.6)	13 (12.6)
Death, hospitalization, or FVC drop >10%	65 (36.0)	49 (34.7)	28 (27.3)

Definition of abbreviations: CAD = coronary artery disease; DL_{CO} = diffusing capacity of the lung for carbon monoxide; GAP = gender-age-physiology; GERD = gastroesophageal reflux disease; HRCT = high-resolution computed tomography; SF-36 = 36-Item Short Form Survey; UCSD = University of California, San Diego.

*UCSD Shortness of Breath Questionnaire scores range from 0 to 120, with higher scores indicating worse function; European Quality of Life scores range from 0 to 100 with higher scores indicating better quality of life; SF-36 scores range from 0 to 100, with higher scores indicating better function.

[†]Kaplan-Meier estimators of failure probability at 60 weeks.

between AMFM and visual scores for each HRCT feature. Correlation was moderate for GGR and normal lung ($r = 0.60$, 95% confidence interval [CI] = 0.52–0.67, $P < 0.0001$; and $r = 0.30$, 95% CI = 0.19–0.40, $P < 0.0001$, respectively), and weak for the other features.

HRCT Scores and Outcomes

As displayed in Table 3, greater baseline volume occupied by GGR in visual and AMFM assessment was associated with increased hazard for disease progression (hazard ratio [HR] per 10% visual GGR increase = 2.21, 95% CI = 1.44–3.39, $P = 0.0003$; and HR per 10% AMFM GGR increase = 1.37, 95% CI = 1.07–1.75, $P = 0.01$). The standardized HRs (HR per

1-unit change in SD) are also shown to allow direct comparison of the magnitude of disease progression prediction across continuous predictors with differing distributions. Using the C statistic as a measure of model discrimination, GGR variables classify progressing individuals as well or better than other clinical variables (visual GGR C statistic = 0.671 [95% CI = 0.590–0.753] and AMFM GGR C statistic = 0.647 [95% CI = 0.564–0.730]), followed by GAP stage (C statistic = 0.639 [95% CI = 0.565–0.714]), and FVC % predicted (C statistic = 0.638 [95% CI = 0.555–0.721]).

To determine if visual and AMFM-measured GGR were independently associated with disease progression after adjustment for relevant covariates,

multivariable models were assessed. Model C statistics were used to compare discrimination across models in subjects having data available on all variables ($n = 156$). As shown in Table 4, models with the best discrimination include GGR (visual and AMFM models shown side by side), GAP stage, and smoking status. After adjusting for covariates, GGR by visual and AMFM assessment were independently associated with elevated hazards of the composite endpoint (HR per 10% visual GGR increase = 1.98, 95% CI = 1.20–3.28, $P = 0.008$; and HR per 10% AMFM GGR increase = 1.36, 95% CI = 1.01–1.84, $P = 0.04$). Model discrimination for the visual GGR and AMFM GGR models were similar (visual C statistic = 0.722, 95% CI = 0.639–0.804; versus AMFM C statistic = 0.720, 95% CI = 0.637–0.802).

All other models tested are shown in the supplement (Tables E4–E6) along with C statistics for comparison to the final model. The final models with visual or AMFM HRCT data had better discriminative ability than similar models without GGR (C statistic without GGR = 0.670, 95% CI = 0.590–0.750). For completeness, models showing individual GAP score components (sex, age, pulmonary function), along with GGR variables, can be found in Tables E5 and E6.

Tables E7 and E8 separate the disease progression composite into FVC decline or death (29 events), and hospitalization or death (20 events), respectively. With four observed deaths in our cohort, we were unable to model mortality alone. Visual and AMFM GGR were independently associated with FVC decline or death, whereas only visual GGR remained independently and significantly associated with hospitalization or death; this may have been due to fewer events and loss of power to detect statistically significant association of AMFM GGR with this subcomposite. Strength and direction of AMFM GGR for hospitalization prediction remained similar to the main model predicting the disease progression composite.

We also assessed the extent of GGR categorically at a threshold of 10% or greater total lung involvement versus less than 10%. This threshold was selected due to a peak in the C statistic between 10 and 20% of lung involvement with GGR by visual and AMFM assessment, and applicability to the greatest number of subjects at the 10%

Table 2. Mean Lung Occupied and Agreement between Visual and Adaptive Multiple Features Method High-Resolution Computed Tomography Feature Scores

HRCT Feature	Mean Feature Score*			Bland-Altman Analysis		Correlation		
	Visual	AMFM	P Value	Mean Difference	95% Limits of Agreement	r	95% CI	P Value†
Ground glass	0.01 (0.03)	0.20 (0.11)	<0.0001	−0.19	−0.41 to 0.03	0.23	0.12 to 0.33	<0.0001
Ground glass reticular	0.12 (0.07)	0.15 (0.10)	<0.0001	−0.03	−0.19 to 0.14	0.60	0.52 to 0.67	<0.0001
Honeycombing	0.04 (0.04)	0.09 (0.14)	<0.0001	−0.05	−0.31 to 0.22	0.29	0.18 to 0.39	<0.0001
Emphysema	0.01 (0.02)	0.01 (0.02)	0.26	0.002	−0.06 to 0.06	0.06	−0.06 to 0.17	0.34
Normal	0.81 (0.11)	0.55 (0.17)	<0.0001	0.26	−0.09 to 0.60	0.30	0.19 to 0.40	<0.0001

Definition of abbreviations: AMFM = adaptive multiple features method; CI = confidence interval; HRCT = high-resolution computed tomography.

Analysis includes all HRCTs scored visually and by AMFM (n = 292).

*Shown is the mean proportion (SD) of total lung having each HRCT feature by visual and AMFM analysis. P value is for the difference between visual and AMFM mean using paired t test.

†P value compares correlation with 0 (no correlation).

threshold. In subjects with AMFM and visually measured GGR scores of 10% or greater, the hazards of FVC decline, hospitalization, or death were 3.37 (95% CI = 1.68–6.75, $P = 0.0006$) and 2.44 (95% CI = 1.35–4.39, $P = 0.003$), respectively. After adjustment for age, sex, baseline FVC % predicted, and smoking status, a baseline

AMFM-measured GGR score of 10% or greater was independently associated with elevated hazard for disease progression (HR = 2.60, 95% CI = 1.24–5.45, $P = 0.01$). Figure 2 shows Kaplan-Meier event-free survival estimates after stratification by baseline visual or AMFM GGR score 10% or greater total lung involvement versus

less than 10%. Subjects with 10% or greater total lung occupied by AMFM GGR had 29.6% lower (absolute) 60-week event-free survival compared with those having less than 10% (60-wk Kaplan-Meier difference, $P < 0.0001$), and those with 10% or greater visual GGR had a 22.3% lower event-free survival at 60 weeks

Table 3. Univariable Cox Proportional Hazards Analysis for Association of Baseline Clinical and High-Resolution Computed Tomography Variables with Disease Progression

Predictor	HR (95% CI)	Standardized HR (95% CI)	P Value	C Statistic (95% CI)
Visual HRCT features				
Ground glass	0.90 (0.31–2.63)	0.97 (0.73–1.30)	0.84	0.501 (0.423–0.578)
Ground glass reticular	2.21 (1.44–3.39)	1.57 (1.23–2.00)	0.0003	0.671 (0.590–0.753)
Honeycombing	1.11 (0.60–2.06)	1.04 (0.82–1.33)	0.74	0.516 (0.435–0.597)
Emphysema	0.74 (0.17–3.21)	0.94 (0.70–1.27)	0.69	0.522 (0.459–0.584)
Normal	0.78 (0.63–0.97)	0.76 (0.60–0.96)	0.02	0.610 (0.528–0.693)
AMFM HRCT features				
Ground glass	1.09 (0.87–1.37)	1.11 (0.85–1.44)	0.45	0.532 (0.449–0.615)
Ground glass reticular	1.37 (1.07–1.75)	1.34 (1.07–1.68)	0.01	0.647 (0.564–0.730)
Honeycombing	1.02 (0.82–1.26)	1.02 (0.79–1.33)	0.87	0.506 (0.423–0.588)
Emphysema	0.93 (0.25–3.38)	0.99 (0.71–1.36)	0.91	0.544 (0.461–0.627)
Normal	0.87 (0.75–1.00)	0.78 (0.61–1.00)	0.054	0.611 (0.528–0.694)
Clinical variables				
Age, yr	1.43 (1.01–2.03)	1.35 (1.01–1.80)	0.046	0.576 (0.494–0.659)
Male sex	0.97 (0.52–1.80)	—	0.92	0.504 (0.440–0.569)
Former smoker	0.82 (0.46–1.48)	—	0.51	0.538 (0.470–0.606)
FVC % predicted	0.73 (0.58–0.88)	0.59 (0.43–0.82)	0.001	0.638 (0.555–0.721)
DL _{CO} % predicted	0.72 (0.55–0.94)	0.67 (0.48–0.92)	0.02	0.608 (0.525–0.691)
GAP stage				
1	Ref.	—	Ref.	0.639 (0.565–0.714)
2	1.68 (0.86–3.28)	—	0.13	—
3	5.42 (2.53–11.63)	—	<0.0001	—

Definition of abbreviations: AMFM = adaptive multiple features method; CI = confidence interval; DL_{CO} = diffusing capacity of the lung for carbon monoxide; GAP = gender-age-physiology; HR = hazard ratio; HRCT = high-resolution computed tomography.

Analysis includes n = 156 patients with both AMFM and visual scores for baseline HRCT and data on all included variables. Of 159 patients with baseline AMFM-scored HRCT, n = 1 subject was missing smoking status and n = 2 subjects were missing DL_{CO}. Baseline characteristics for this subgroup, including number of events/outcomes, are shown in Table 1, column titled “Baseline HRCT.” Unstandardized HR is for each 10% increase in baseline total lung volume occupied by HRCT features and for each 10% increase in baseline % predicted FVC or DL_{CO}, and for 10-year increase in age. Standardized HR is for each standard deviation increase in the continuous predictors. The categorical variables were not standardized. The GAP stage (24) is a mortality prediction model, incorporating age, sex, FVC, and DL_{CO}, with higher GAP stages experiencing the worst outcomes. The outcome of interest is a composite FVC drop >10%, hospitalization, or mortality event.

Table 4. Multivariable Cox Proportional Hazards Models for Association of Baseline Clinical and High-Resolution Computed Tomography Variables with Disease Progression

Predictor	Visual Assessment C Statistic: 0.722 (95% CI = 0.639–0.804)			AMFM Assessment C Statistic: 0.720 (95% CI = 0.637–0.802)		
	HR (95% CI)	Standardized HR (95% CI)	P Value	HR (95% CI)	Standardized HR (95% CI)	P Value
Ground glass reticular	1.98 (1.20–3.28)	1.47 (1.11–1.96)	0.008	1.36 (1.01–1.84)	1.35 (1.01–1.76)	0.04
Former smoker	0.77 (0.42–1.41)	—	0.40	0.66 (0.36–1.22)	—	0.18
GAP stage						
1	Ref.	—	Ref.	Ref.	—	Ref.
2	1.49 (0.74–3.00)	—	0.27	1.65 (0.83–3.29)	—	0.15
3	3.96 (1.73–9.09)	—	0.001	4.66 (2.10–10.31)	—	0.0002

Definition of abbreviations: AMFM = adaptive multiple features method; CI = confidence interval; GAP = gender-age-physiology; HR = hazard ratio. Analysis includes n = 156 patients. Of 159 subjects having baseline AMFM-scored high-resolution computed tomography (HRCT), n = 1 subject was missing smoking status and n = 2 missing diffusing capacity of the lung for carbon monoxide (D_{LCO}) required for GAP stage calculation. Baseline characteristics for this subgroup, including number of events/outcomes, is shown in Table 1 column titled “Baseline HRCT.” Unstandardized HR is for each 10% increase in baseline total lung volume occupied by ground glass reticular. Standardized HR is for each standard deviation increase in ground glass reticular. The categorical variables were not standardized. The GAP stage (24) is a mortality prediction model incorporating age, sex, FVC, and D_{LCO} , with higher GAP stages experiencing the worst outcomes. The outcome of interest is a composite FVC drop >10%, hospitalization, or mortality event.

compared with those having less than 10% (60-wk Kaplan-Meier difference, $P = 0.002$).

Postbaseline HRCT Scores and PFT Change

The baseline HRCT was obtained a median of −4 weeks (range −34.7 to 5) from PANTHER-IPF enrollment, and the final scan was at a median of 60 weeks (range, 35 to 113) from enrollment. There were three outliers in terms of timing of the final scan (i.e., significantly sooner or later than the expected 60-wk mark), and these

deviations were due to early termination of the prednisone-azathioprine-NAC arm of the parent study (15) or deviation from protocol resulting in late final HRCT. Median time between baseline and final scans was 63.9 weeks (range, 37.9–113.7). Table E9 shows the postbaseline change in total lung volume occupied by each HRCT feature, as assessed visually and by AMFM, as well as correlation between change in visual and AMFM scores. On average, little change was observed between scans for any HRCT feature. Visual and AMFM

postbaseline change assessment for GGR was moderately correlated ($r = 0.47$, 95% CI = 0.30–0.60, $P < 0.0001$), but visual and AMFM postbaseline change for the other HRCT features were not significantly correlated ($P \geq 0.06$).

We observed AMFM-measured GGR trajectory to be weakly negatively correlated with postbaseline FVC trajectory ($r = -0.25$, 95% CI = −0.42 to −0.06, $P = 0.01$), such that increasing GGR was associated with decreasing FVC. Visual GGR trajectory was moderately negatively correlated with FVC trajectory ($r = -0.30$, 95% CI = −0.46 to −0.11, $P = 0.002$). As shown in Table 5, postbaseline change in the other visual and AMFM-measured features were not correlated significantly with postbaseline FVC trajectory. In addition to correlation analyses, corresponding linear regression analyses are presented in Table E10 that relate 10% postbaseline increase in HRCT features to postbaseline change in % predicted FVC. Supplemental Figure E4 shows these relationships visually as scatterplots with overlaid LOESS (locally weighted scatterplot smoothing) lines.

Discussion

The AMFM analysis software is capable of automated and objective quantification of lung volume occupied by a variety of HRCT features. We prospectively demonstrate the utility of AMFM software HRCT analysis in patients with IPF in several

Figure 2. Kaplan-Meier event-free survival analysis stratified by visual and adaptive multiple features method (AMFM) ground glass reticular (GGR) density scores of 10% or greater, or less than 10%. Shown are Kaplan-Meier event-free (events of interest are FVC % predicted decline >10%, hospitalization, or death) survival estimates in subjects stratified by visual and AMFM GGR scores of 10% or greater total lung involvement versus less than 10%. The crosses denote censoring and the steps are events. Subjects having 10% or more total lung occupied by GGR densities at baseline have significantly worse event-free survival than those with less than 10% (log-rank visual, $P = 0.002$; AMFM, $P = 0.0003$).

Table 5. Correlation of Postbaseline High-Resolution Computed Tomography Feature Scores Trajectories and Postbaseline FVC % Predicted Trajectory

HRCT Feature	Visual Assessment (n = 105)			AMFM Assessment (n = 105)		
	r	95% CI	P Value	r	95% CI	P Value
Ground glass	−0.18	−0.36 to 0.01	0.07	−0.11	−0.29 to 0.09	0.27
Ground glass reticular	−0.30	−0.46 to −0.11	0.002	−0.25	−0.42 to −0.06	0.01
Honeycombing	−0.13	−0.32 to 0.06	0.17	0.09	−0.10 to 0.28	0.34
Emphysema	0.09	−0.10 to 0.28	0.35	0.12	−0.07 to 0.31	0.21
Normal	0.04	−0.16 to 0.23	0.70	0.12	−0.08 to 0.30	0.24

For definition of abbreviations, see Table 2.

Of 107 patients with baseline and final scans, scan dates were missing on n = 2, precluding calculation of HRCT score trajectory. Baseline characteristics for this subgroup, including number of events/outcomes, is shown in Table 1 column titled “Sequential HRCT.”

ways: (1) AMFM-measured fibrosis extent, as measured by the GGR score, agrees reasonably well with visual measurement by expert radiologists; (2) greater baseline AMFM-measured fibrosis is independently associated with elevated hazard of disease progression (specifically, a composite of >10% FVC decline, hospitalization, or mortality, as well as FVC decline or death) after adjustment for relevant covariates; and (3) postbaseline change in HRCT features measured by AMFM software correlates with change in pulmonary function. The gold standard, a three-expert average of visually measured fibrosis, was similarly associated with disease progression and postbaseline pulmonary function change.

Development of software to aid with interpretation of HRCT images is pursued because of the well-recognized variability in visual interpretation of HRCTs regarding presence and extent of fibrosis and other features (2–4). It is also hypothesized that software may be more sensitive to subtle abnormalities or temporal changes compared with humans. Several important hurdles exist in development of such software. First, there is no true “gold standard” measure to calibrate software and ensure that it is measuring what it is trained to measure. Surrogates include correlation of software measurements with expert radiologist measurements, or testing software measures in predicting clinical events or outcomes. We assess both here, with an average of total lung scores assigned by three expert radiologists after independent HRCT review used as the “gold standard.”

AMFM software was successfully trained, and shows reasonable agreement with the gold standard regarding measurement of fibrosis extent, with

moderate correlation and mean absolute between-method difference of −0.03% for GGR. The between-method difference for GGR did have a relatively wide 95% limit of agreement (−14 to 19%). It is unclear what an acceptable 95% limit of agreement is for visual versus software fibrosis measurement; this is the rationale for assessing both methods of measurement for their association with clinical outcomes.

Both AMFM and visually measured baseline fibrosis extent are independently associated with increased hazard of disease progression. The gold standard three-expert average visual GGR score may slightly outperform AMFM GGR as an individual predictor, based upon univariable C statistics, but we lacked statistical power to conclude that one method was superior to another, as evidenced by overlapping CIs for the univariable C statistics (see Table 3). The GGR variables had higher univariable C statistics than any other individual predictor, including age, GAP stage, FVC, and DL_{CO}. After adjusting for smoking status and baseline GAP stage, both visual and AMFM-measured GGR were independently associated with elevated hazard of disease progression. Model discrimination (as measured by C statistic) of the final models with smoking status, GAP stage, and GGR were similar, regardless of the method of fibrosis measurement (visual model C statistic = 0.722 compared with AMFM model C statistic = 0.722). Addition of either the visual or AMFM GGR variable to multivariable analysis did improve model discrimination compared with models not including GGR variables (see Table E4 [no GGR variables] compared with Table E5 [visual models] and E6 [AMFM models]). Given the variability in visual interpretation

of disease extent in practice, use of software may produce results that are more reliably associated with outcomes than an average radiologist.

Maldonado and colleagues (6) found CALIPER (Computer Aided Lung Informatics for Pathology Evaluation and Rating) software-measured reticular densities and total ILD to be associated with increased hazard of death in a retrospective cohort of patients with IPF. With only four deaths in our cohort, we lacked power to show association of GGR with this outcome specifically, and this is an important area of future study. We suspect that the paucity of deaths is due to our use of clinical therapeutic trial data, leading to inclusion of healthier patients, whereas Maldonado and colleagues used a retrospective cohort of patients with IPF seen at a tertiary referral center.

Association of GGR fibrosis with nonmortality disease progression (specifically, >10% FVC decline or death), after adjustment for relevant covariates, has important clinical utility, as well as relevance with regard to clinical trial design. Risk factors for increased mortality in IPF are well defined, but prior analyses have failed to identify factors strongly associated with disease progression (25–27). With change in FVC adopted as the primary endpoint in recent clinical trials, AMFM measurement of GGR at baseline may prove a useful, automated tool, particularly for cohort enrichment likely to improve power when change in FVC is the primary endpoint (28–34). Prior studies used different methodology in retrospective cohort samples, and found that neither baseline GAP stage nor initial PFT decline were associated with faster subsequent pulmonary function decline (25, 26). Ley and colleagues (27) found baseline FVC, University of California, San Diego Shortness of Breath Questionnaire score, 6-minute-walk distance, and GAP stage 3 (vs. stage 1) to be unadjusted predictors of future FVC decline greater than 10% or death. Multivariable Cox prediction models, including combinations of baseline variables, were assessed with C statistics found to be less than 0.70, and it was thus concluded that FVC decline could not be reliably predicted using the evaluated baseline variables. Park and colleagues (35) retrospectively applied a texture-based automated HRCT scoring system to patients with IPF and found that those with greater than 10% decline in FVC

after 1 year of follow up had significantly more reticular opacities at baseline. We believe that our results are in line with these findings, and together support the idea that more severe baseline disease is associated with greater risk of disease progression. The discriminative capacity of our final models could be considered good, with C statistics of 0.722 and 0.720 for visual and AMFM multivariable models, respectively. Excellent prediction of events, such as disease progression or death, by any particular statistical model is unlikely. Our finding that subjects with 10% or more of total lung having GGR fibrosis at baseline have 20–30% more events compared with those with less than 10% fibrosis is nonetheless useful (see Figure 2). Furthermore, addition of GGR variables to multivariable Cox models results in incremental improvement of model discrimination over models without radiologic variables.

We also found postbaseline change in AMFM and visually measured fibrosis to be weakly and moderately correlated with postbaseline FVC trajectory, respectively. Similar to our findings, Kim and colleagues (36) found that change in a software-measured quantitative lung fibrosis score was correlated with FVC and DL_{CO} change over time in patients with IPF. The weak correlations of postbaseline change in HRCT features with postbaseline FVC trajectory is likely due to the minimal average change for these features. It has been suggested that change in HRCT fibrosis could be used as an endpoint in clinical trials, representing disease progression, where FVC decline is currently used (37–39). We believe that the GGR score is a reasonable candidate marker, but

further study is needed to determine what degree of change in GGR over time is clinically significant, and whether interval change in GGR score is associated with future risk of death. Unfortunately, we have no follow-up data beyond 60 weeks, and are therefore unable to answer this question.

Although AMFM seems to measure GGR fibrotic change well, it was not as reliable in measuring GG, honeycombing, or normal lung. The Bland-Altman analysis indicates a systematic tendency for AMFM to measure more GG than is measured by visual interpretation. It is possible that software interprets artifacts, such as respiratory motion, as GG, or that AMFM is more sensitive to early abnormalities. Baseline extent of GG was not associated with outcomes when measured by visual or AMFM method. In addition, although we found moderate correlation between visual and AMFM scores for GGR, and weak correlation for most other features, there was no correlation for emphysema. The reason for this is likely the overall minimal emphysema present, on average, in our population (total lung volume measurements ranging from 0.7 to 0.9%) as a result of PANTHER-IPF inclusion/exclusion criteria; no HRCT was scored with more than approximately 20% of lung involved with emphysema by either method.

Our study has some weaknesses. We excluded HRCTs when scan protocol was not appropriate for AMFM analysis or when poor quality scans were submitted from local centers (see Figure E2). We used expert thoracic radiologists' quality ratings to determine which scans were not suitable for inclusion. Before widespread adoption

of AMFM HRCT analysis, automated or technician-administered quality-control measures are needed to prevent erroneous generation or interpretation of data obtained from poor-quality scans (40). Reassuringly, fewer than 10% of total scans were excluded due to poor quality. We also can't exclude differences between included patients and PANTHER-IPF subjects who declined to participate in this study, but our patient characteristics do not differ dramatically from those seen in either PANTHER-IPF study (14, 15). In addition, although extent of GGR fibrosis at baseline appears to be an important predictor of disease progression in this population of patients enrolled in a clinical therapeutic trial, caution in application of these findings should be taken until results can be externally validated in a separate population.

In conclusion, we show, in a prospective study of patients with IPF enrolled in a clinical therapeutic trial, that AMFM software objectively quantifies the volume of lung occupied by various interstitial abnormalities. AMFM and visual measurement of fibrosis extent had reasonable agreement, and greater volume occupied by AMFM-measured fibrosis was independently associated with increased hazard of disease progression in a manner similar to an averaged visual fibrosis score from three expert radiologists. AMFM-measured fibrosis extent could serve as an objective HRCT biomarker of IPF severity and increased risk of future disease progression. ■

Author disclosures are available with the text of this article at www.atsjournals.org.

References

1. Raghu G, Collard HR, Egan JJ, Martinez FJ, Behr J, Brown KK, Colby TV, Cordier JF, Flaherty KR, Lasky JA, et al.; ATS/ERS/JRS/ALAT Committee on Idiopathic Pulmonary Fibrosis. An official ATS/ERS/JRS/ALAT statement: idiopathic pulmonary fibrosis: evidence-based guidelines for diagnosis and management. *Am J Respir Crit Care Med* 2011;183:788–824.
2. Sumikawa H, Johkoh T, Colby TV, Ichikado K, Suga M, Taniguchi H, Kondoh Y, Ogura T, Arakawa H, Fujimoto K, et al. Computed tomography findings in pathological usual interstitial pneumonia: relationship to survival. *Am J Respir Crit Care Med* 2008;177:433–439.
3. Walsh SL, Calandriello L, Sverzellati N, Wells AU, Hansell DM; UIP Observer Consort. Interobserver agreement for the ATS/ERS/JRS/ALAT criteria for a UIP pattern on CT. *Thorax* 2016;71:45–51.
4. Watadani T, Sakai F, Johkoh T, Noma S, Akira M, Fujimoto K, Bankier AA, Lee KS, Müller NL, Song JW, et al. Interobserver variability in the CT assessment of honeycombing in the lungs. *Radiology* 2013;266:936–944.
5. Kliment CR, Araki T, Doyle TJ, Gao W, Dupuis J, Latourelle JC, Zazueta OE, Fernandez IE, Nishino M, Okajima Y, et al. A comparison of visual and quantitative methods to identify interstitial lung abnormalities. *BMC Pulm Med* 2015;15:134.
6. Maldonado F, Moua T, Rajagopalan S, Karwoski RA, Raghunath S, Decker PA, Hartman TE, Bartholmai BJ, Robb RA, Ryu JH. Automated quantification of radiological patterns predicts survival in idiopathic pulmonary fibrosis. *Eur Respir J* 2014;43:204–212.
7. Rosas IO, Yao J, Avila NA, Chow CK, Gahl WA, Gochoico BR. Automated quantification of high-resolution CT scan findings in individuals at risk for pulmonary fibrosis. *Chest* 2011;140:1590–1597.
8. Best AC, Lynch AM, Bozic CM, Miller D, Grunwald GK, Lynch DA. Quantitative CT indexes in idiopathic pulmonary fibrosis: relationship with physiologic impairment. *Radiology* 2003;228:407–414.
9. Best AC, Meng J, Lynch AM, Bozic CM, Miller D, Grunwald GK, Lynch DA. Idiopathic pulmonary fibrosis: physiologic tests, quantitative CT indexes, and CT visual scores as predictors of mortality. *Radiology* 2008;246:935–940.

10. Uppaluri R, Hoffman EA, Sonka M, Hunninghake GW, McLennan G. Interstitial lung disease: a quantitative study using the adaptive multiple feature method. *Am J Respir Crit Care Med* 1999;159: 519–525.
11. Uppaluri R, Mitsa T, Sonka M, Hoffman EA, McLennan G. Quantification of pulmonary emphysema from lung computed tomography images. *Am J Respir Crit Care Med* 1997;156:248–254.
12. Xu Y, van Beek EJ, Hwanjo Y, Guo J, McLennan G, Hoffman EA. Computer-aided classification of interstitial lung diseases via MDCT: 3D adaptive multiple feature method (3D AMFM). *Acad Radiol* 2006;13:969–978.
13. Uppaluri R, Hoffman EA, Sonka M, Hartley PG, Hunninghake GW, McLennan G. Computer recognition of regional lung disease patterns. *Am J Respir Crit Care Med* 1999;160:648–654.
14. Martinez FJ, de Andrade JA, Anstrom KJ, King TE Jr, Raghu G; Idiopathic Pulmonary Fibrosis Clinical Research Network. Randomized trial of acetylcysteine in idiopathic pulmonary fibrosis. *N Engl J Med* 2014;370:2093–2101.
15. Raghu G, Anstrom KJ, King TE Jr, Lasky JA, Martinez FJ; Idiopathic Pulmonary Fibrosis Clinical Research Network. Prednisone, azathioprine, and N-acetylcysteine for pulmonary fibrosis. *N Engl J Med* 2012;366:1968–1977.
16. Salisbury ML, van Beek EJ, Kazerooni EA, Guo J, Xia M, Murray S, Anstrom KJ, Yow E, Martinez FJ, Hoffman EA, et al. Association of adaptive multiple features method measures of fibrosis with outcomes in idiopathic pulmonary fibrosis [abstract]. International Colloquium on Lung and Airway Fibrosis, Dublin, Ireland, September 24–18, 2016.
17. Macintyre N, Crapo RO, Viegi G, Johnson DC, van der Grinten CP, Brusasco V, Burgos F, Casaburi R, Coates A, Enright P, et al. Standardisation of the single-breath determination of carbon monoxide uptake in the lung. *Eur Respir J* 2005;26:720–735.
18. Miller MR, Hankinson J, Brusasco V, Burgos F, Casaburi R, Coates A, Crapo R, Enright P, van der Grinten CP, Gustafsson P, et al.; ATS/ERS Task Force. Standardisation of spirometry. *Eur Respir J* 2005;26:319–338.
19. Hansell DM, Bankier AA, MacMahon H, McLoud TC, Müller NL, Remy J. Fleischner Society: glossary of terms for thoracic imaging. *Radiology* 2008;246:697–722.
20. Zisman DA, Schwarz M, Anstrom KJ, Collard HR, Flaherty KR, Hunninghake GW; Idiopathic Pulmonary Fibrosis Clinical Research Network. A controlled trial of sildenafil in advanced idiopathic pulmonary fibrosis. *N Engl J Med* 2010;363:620–628.
21. Marasini D, Quatto P, Ripamonti E. Assessing the inter-rater agreement for ordinal data through weighted indexes. *Stat Methods Med Res* 2014;25:1–23.
22. Bland JM, Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet* 1986;1:307–310.
23. Ley B, Elicker BM, Hartman TE, Ryerson CJ, Vittinghoff E, Ryu JH, Lee JS, Jones KD, Richeldi L, King TE Jr, et al. Idiopathic pulmonary fibrosis: CT and risk of death. *Radiology* 2014;273:570–579.
24. Ley B, Ryerson CJ, Vittinghoff E, Ryu JH, Tomassetti S, Lee JS, Poletti V, Buccioli M, Elicker BM, Jones KD, et al. A multidimensional index and staging system for idiopathic pulmonary fibrosis. *Ann Intern Med* 2012;156:684–691.
25. Schmidt SL, Tayob N, Han MK, Zappala C, Kervitsky D, Murray S, Wells AU, Brown KK, Martinez FJ, Flaherty KR. Predicting pulmonary fibrosis disease course from past trends in pulmonary function. *Chest* 2014; 145:579–585.
26. Salisbury ML, Xia M, Zhou Y, Murray S, Tayob N, Brown KK, Wells AU, Schmidt SL, Martinez FJ, Flaherty KR. Idiopathic pulmonary fibrosis: gender-age-physiology index stage for predicting future lung function decline. *Chest* 2016;149:491–498.
27. Ley B, Bradford WZ, Vittinghoff E, Weycker D, du Bois RM, Collard HR. Predictors of mortality poorly predict common measures of disease progression in idiopathic pulmonary fibrosis. *Am J Respir Crit Care Med* 2016;194:711–718.
28. Daniels CE, Lasky JA, Limper AH, Mieras K, Gabor E, Schroeder DR; Imatinib-IPF Study Investigators. Imatinib treatment for idiopathic pulmonary fibrosis: randomized placebo-controlled trial results. *Am J Respir Crit Care Med* 2010;181:604–610.
29. Demedts M, Behr J, Buhl R, Costabel U, Dekhuijzen R, Jansen HM, MacNee W, Thomeer M, Wallaert B, Laurent F, et al.; IFIGENIA Study Group. High-dose acetylcysteine in idiopathic pulmonary fibrosis. *N Engl J Med* 2005;353:2229–2242.
30. King TE Jr, Behr J, Brown KK, du Bois RM, Lancaster L, de Andrade JA, Stähler G, Leconte I, Roux S, Raghu G. BUILD-1: a randomized placebo-controlled trial of bosentan in idiopathic pulmonary fibrosis. *Am J Respir Crit Care Med* 2008;177:75–81.
31. King TE Jr, Bradford WZ, Castro-Bernardini S, Fagan EA, Glaspole I, Glassberg MK, Gorina E, Hopkins PM, Kardatzke D, Lancaster L, et al.; ASCEND Study Group. A phase 3 trial of pirfenidone in patients with idiopathic pulmonary fibrosis. *N Engl J Med* 2014;370: 2083–2092.
32. Noble PW, Albera C, Bradford WZ, Costabel U, Glassberg MK, Kardatzke D, King TE Jr, Lancaster L, Sahn SA, Swartzberg J, et al.; CAPACITY Study Group. Pirfenidone in patients with idiopathic pulmonary fibrosis (CAPACITY): two randomised trials. *Lancet* 2011; 377:1760–1769.
33. Raghu G, Brown KK, Bradford WZ, Starko K, Noble PW, Schwartz DA, King TE Jr; Idiopathic Pulmonary Fibrosis Study Group. A placebo-controlled trial of interferon gamma-1b in patients with idiopathic pulmonary fibrosis. *N Engl J Med* 2004;350:125–133.
34. Raghu G, Brown KK, Costabel U, Cottin V, du Bois RM, Lasky JA, Thomeer M, Utz JP, Khandker RK, McDermott L, et al. Treatment of idiopathic pulmonary fibrosis with etanercept: an exploratory, placebo-controlled trial. *Am J Respir Crit Care Med* 2008;178: 948–955.
35. Park HJ, Lee SM, Song JW, Lee SM, Oh SY, Kim N, Seo JB. Texture-based automated quantitative assessment of regional patterns on initial CT in patients with idiopathic pulmonary fibrosis: relationship to decline in forced vital capacity. *Am J Roentgenol* 2016;207: 976–983.
36. Kim HJ, Brown MS, Chong D, Gjertson DW, Lu P, Kim HJ, Coy H, Goldin JG. Comparison of the quantitative CT imaging biomarkers of idiopathic pulmonary fibrosis at baseline and early change with an interval of 7 months. *Acad Radiol* 2015;22:70–80.
37. Collard HR, King TE Jr, Bartelson BB, Vourlekis JS, Schwarz MI, Brown KK. Changes in clinical and physiologic variables predict survival in idiopathic pulmonary fibrosis. *Am J Respir Crit Care Med* 2003;168: 538–542.
38. Flaherty KR, Andrei AC, Murray S, Fraley C, Colby TV, Travis WD, Lama V, Kazerooni EA, Gross BH, Toews GB, et al. Idiopathic pulmonary fibrosis: prognostic value of changes in physiology and six-minute-walk test. *Am J Respir Crit Care Med* 2006;174:803–809.
39. Latsi PI, du Bois RM, Nicholson AG, Colby TV, Bisirtzoglou D, Nikolakopoulou A, Veeraraghavan S, Hansell DM, Wells AU. Fibrotic idiopathic interstitial pneumonia: the prognostic value of longitudinal functional trends. *Am J Respir Crit Care Med* 2003;168:531–537.
40. Hoffman EA, Reinhardt JM, Sonka M, Simon BA, Guo J, Saba O, Chon D, Samrah S, Shikata H, Tschirren J, et al. Characterization of the interstitial lung diseases via density-based and texture-based analysis of computed tomography images of lung structure and function. *Acad Radiol* 2003;10:1104–1118.