

# High-Resolution Computed Tomography in Idiopathic Pulmonary Fibrosis

## Diagnosis and Prognosis

David A. Lynch, J. David Godwin, Sharon Safrin, Karen M. Starko, Phil Hormel, Kevin K. Brown, Ganesh Raghunath, Talmadge E. King, Jr., Williamson Z. Bradford, David A. Schwartz, and W. Richard Webb, for the Idiopathic Pulmonary Fibrosis Study Group

Department of Radiology, and Division of Pulmonary Sciences and Critical Care Medicine, University of Colorado Health Sciences Center; National Jewish Medical and Research Center, Denver, Colorado; Department of Radiology, and Division of Pulmonary and Critical Care Medicine, University of Washington Medical Center, Seattle, Washington; Clinical Science and Biometrics Departments, InterMune, Inc., Brisbane; Department of Medicine, San Francisco General Hospital, and Department of Radiology, University of California–San Francisco, San Francisco, California; and Department of Medicine, Duke University Medical Center, Durham, North Carolina

**Rationale:** High-resolution computed tomography (HRCT) is an integral aspect of the evaluation of patients with suspected idiopathic pulmonary fibrosis (IPF). However, few studies have evaluated its use in a large cohort. **Objectives:** To describe HRCT features in patients with mild to moderate IPF, compare diagnostic evaluations by a radiology core (three thoracic radiologists) with those by study-site radiologists, correlate baseline clinical and physiologic variables with HRCT findings, and evaluate their association with mortality. **Methods:** We assessed HRCT scans from patients with IPF ( $n = 315$ ) enrolled in a randomized controlled study evaluating IFN- $\gamma$ 1b. **Measurements and Main Results:** There was concordance between study-site and core radiologists regarding the diagnosis of IPF in 86% of cases. Diffusing capacity of carbon monoxide ( $D_{LCO}$ ) was the physiologic characteristic most highly correlated with HRCT findings. Multivariate analysis identified three independent predictors of mortality: a higher extent of fibrosis score increased the risk of death ( $p < 0.0001$ ), whereas a higher percent-predicted  $D_{LCO}$  ( $p = 0.004$ ) and treatment assignment to IFN- $\gamma$ 1b rather than placebo ( $p = 0.04$ ) reduced the risk of death. **Conclusions:** A study-site diagnosis of IPF on HRCT was regularly confirmed by core radiologists. Extent of reticulation and honeycombing on HRCT is an important independent predictor of mortality in patients with IPF.

**Keywords:** high-resolution computed tomography; idiopathic pulmonary fibrosis; mortality; prognosis

Idiopathic pulmonary fibrosis (IPF) is a discrete clinical and histopathologic entity with a uniformly poor prognosis (1). Identification of usual interstitial pneumonia (UIP) on surgical lung biopsy has been considered the gold standard for diagnosis (1). However, when assessed by expert clinicians and radiologists, the presence of typical clinical and high-resolution computed tomography (HRCT) features is sufficient to allow a confident diagnosis of IPF in more than 50% of suspected cases and may eliminate the need for surgical lung biopsy in these patients (2). Thus, HRCT has become an integral part of evaluation of patients with idiopathic interstitial pneumonia (2–4).

Despite the increasingly important role of HRCT in the diagnosis and follow-up of patients with IPF, little data exist regarding the HRCT features of patients with mild to moderate physiologic

impairment, the reliability of the HRCT interpretation among radiologists, and the usefulness of HRCT findings as a predictors of mortality in such patients (5, 6). The present study sought to address these issues by using data from a prospective, randomized, double-blind, placebo-controlled clinical trial evaluating the use of IFN- $\gamma$ 1b in patients with IPF (7). We describe the HRCT characteristics of scans in patients with IPF, compare the interpretations of study-site radiologists with a central core group of thoracic radiologists, and examine the correlation between baseline HRCT characteristics at study entry and selected clinical, physiologic, and pathologic parameters. Also, we evaluate the relationship between HRCT findings and selected clinical and physiologic variables with subsequent mortality over a median follow-up period of 58 weeks. Some of the results of this study have been previously reported in abstract form (8).

## METHODS

In a previously published phase 3 trial in patients with IPF (10), 162 patients were randomly assigned to receive IFN- $\gamma$ 1b (200  $\mu$ g subcutaneously three times weekly) and 168 received matched placebo. Patients were recruited from a total of 58 medical centers (39 academic, 19 community-based) in the United States, Europe, Canada, and South Africa. Criteria for enrollment have been previously published.

A total of 160 (98.8%) IFN- $\gamma$ 1b patients and 166 (98.8%) placebo patients had a baseline HRCT scan available for evaluation. HRCT evaluations were included in the analysis only if the baseline HRCT was performed within 60 days before the first dose of study drug in the phase 3 trial. Of these, the scans of 11 (7 IFN- $\gamma$ 1b and 4 placebo) patients were excluded from the analysis due to the timing of the HRCT (i.e., either more than 60 days before the first study drug dose [3 patients] or after the first dose [6 patients] or the inability of either core radiologist to evaluate the image [2 patients]). Thus, 315 baseline HRCT scans (153 IFN- $\gamma$ 1b, 162 placebo) were analyzed.

The study protocol required thin collimation (1–1.5 mm) images to be obtained through the lung using standard high-resolution technique at 2-cm intervals in supine and prone positions. The minimum tube exposure was 200 milliamperes-seconds. Volumetric CT was not performed. Digital scan data were sent to a central processing site in a Digital Imaging and Communications in Medicine (DICOM)-compatible, anonymized format.

## Radiologic Assessment by Study-Site Radiologists

Using defined criteria, radiologists at each investigational site (“study-site radiologists”) were asked to determine if either “definite” or “probable” IPF was present. A radiographic diagnosis of “definite IPF” required all three of the following criteria: (1) presence of reticular abnormality and/or traction bronchiectasis with basal and peripheral predominance; (2) presence of honeycombing with basal and peripheral predominance; and (3) absence of atypical features, such as micronodules, peribronchovascular nodules, consolidation, isolated (nonhoneycomb) cysts, extensive ground glass attenuation, or extensive mediastinal adenopathy.

(Received in original form December 31, 2004; accepted in final form May 9, 2005)

Supported by InterMune, Inc. (Brisbane, CA).

Correspondence and requests for reprints should be addressed to David A. Lynch, M.D., University of Colorado Health Sciences Center, 4200 East Ninth Avenue, Box A030, Room 2233 Denver, CO 80262. E-mail: david.lynch@uchsc.edu

Am J Respir Crit Care Med Vol 172, pp 488–493, 2005

Originally Published in Press as DOI: 10.1164/rccm.200412-1756OC on May 13, 2005

Internet address: www.atsjournals.org

The presence of the first and third criterion only qualified as “probable” IPF (i.e., honeycombing was not present). Clinical information may have been provided and the study-site radiologists knew that IPF was a consideration.

### Radiologic Assessment by Core Radiologists

After the completion of the trial, a core panel of three thoracic radiologists (“core radiologists”) was convened to independently review the baseline HRCT scans. The radiologists were blinded to clinical data and treatment group assignment; however, they knew that the patients had met nonradiologic inclusion criteria for the study, and that a study-site radiologist had interpreted the HRCT scan as at least probable IPF on the basis of predefined criteria.

Two core radiologists independently scored the baseline HRCT on a standardized form. The HRCT image was assessed for the presence and extent of ground glass attenuation, reticulation, honeycombing, decreased attenuation, centrilobular nodules, other nodules, consolidation, and emphysema. The extent of these abnormalities and the overall extent of fibrosis (i.e., the extent of reticulation and honeycombing) were determined for each entire lung using a 4-point scale (0 = no involvement, 1 = 1–25% involvement, 2 = 26–50% involvement, 3 = 51–75% involvement, and 4 = 76–100% involvement). The presence or absence of upper or lower lobe volume loss, traction bronchiectasis, crazy paving, tree in bud, bronchiolectasis, and mosaic attenuation was also assessed, and the predominant pattern (i.e., ground glass/reticulation/honeycombing vs. nodules/mosaic attenuation/emphysema/other) was determined. Each HRCT was classified by at least two core radiologists as typical IPF, atypical IPF, or inconsistent with IPF using usual diagnostic evaluation processes without prespecified criteria for the study. A third core radiologist evaluated the scan if the first two readers did not agree, and the consensus diagnosis was based on agreement of at least two readers. Only two readers were used for pattern extent scores, including honeycombing. Neither discussion nor adjudication was used for any result. In the event of disagreement between the readers, the result was recorded as missing.

### Data Analysis

Data are given as mean  $\pm$  SD or as patient count and percentage. A final pattern extent score was derived as the mean of up to four scores for each scan (i.e., right and left lung scores from two core radiologists). Honeycombing, for example, was defined as a mean pattern extent score exceeding 0. For the diagnosis of IPF, the categories of “typical IPF” and “atypical IPF” were pooled as “consistent with IPF” and analyzed as a dichotomous variable. The simple  $\kappa$  coefficient, ranging from  $-1$  to  $+1$ , was used to assess the degree of interrater agreement in specific comparisons (9). The Wilcoxon rank sum, Spearman rank order correlation, and Fisher’s exact tests were used for statistical comparisons of selected clinical, histologic, and HRCT characteristics, as appropriate.

Univariate analyses of the association between baseline HRCT, clinical, and physiologic characteristics with mortality used a two-tailed  $\chi^2$  test, with statistical significance defined as  $p < 0.05$ . Clinical and physiologic variables included age, sex, smoking status, use of supplemental oxygen, percent-predicted FVC, percent-predicted diffusing capacity of carbon monoxide (DL<sub>CO</sub>), alveolar–arterial (A-a) oxygen gradient, baseline dyspnea index (10), University of California–San Diego Shortness of Breath Questionnaire (11), the modified Medical Research Council scale (12), and treatment assignment (i.e., IFN- $\gamma$ 1b or placebo). HRCT variables included overall fibrosis, honeycomb, reticulation, and emphysema pattern extent scores, predominant pattern (see above), consensus diagnosis of IPF, and the coexisting diagnosis of emphysema. Thereafter, a stepwise logistic regression model was built using variables with univariate  $p$  values of less than 0.20, entering variables at the 0.05 significance level and removing them at the 0.10 level.

## RESULTS

### HRCT Diagnosis of IPF by Study-Site versus Core Radiologists

As was required for entry into the phase 3 trial, using prespecified criteria, the study-site radiologists at each participating medical

center confirmed the diagnosis of IPF on HRCT in all enrolled patients ( $n = 315$ ).

On the basis of the findings by the first two readers among the core radiologists, the scans of 256 (81.3%) patients were considered to be consistent with the diagnosis of IPF (Table 1). Overall, there were concordant interpretations by the first two readers in 271 (86.0%) scans ( $\kappa = 0.33$ ; 95% confidence interval [CI], 0.18–0.48).

A consensus regarding the diagnosis of IPF (i.e., agreement between at least two core radiologists) was reached in 313 (99.4%) scans: 283 (89.8%) as consistent with IPF (Figure 1) and 30 (9.5%) as inconsistent (Figure 2 and Table 1). In two scans, no consensus could be reached (e.g., one core radiologist scored as consistent, another as inconsistent, and the third as unevaluable). Of the 263 scans that were interpreted by the site radiologists as “definite IPF,” 245 (93.2%) were believed to be consistent with IPF by core radiologist consensus, as were 37 (75.5%) of 49 scans read by site radiologists as “probable IPF” ( $p < 0.001$ ; data not shown).

We compared the diagnosis of IPF by the study-site radiologist and the core radiologists according to whether the investigational site was academic ( $n = 39$ ) or community-based ( $n = 19$ ). The diagnosis of IPF was agreed on by consensus of core radiologists in 206 (90.4%) of 228 scans from academic sites and in 76 (90.5%) of 84 scans from community-based sites ( $p = 1.0$ ; data not shown).

### Comparison of UIP on Biopsy with HRCT Scan Pattern

Histologic confirmation of UIP on surgical lung biopsy within the 30 months before study entry was reported in 205 (65%) of the 315 patients. Of these, 181 (88.3%) baseline HRCT scans were interpreted as consistent with IPF by core radiologist consensus, whereas 24 scans (11.7%) were considered inconsistent with IPF. Of the 110 patients who did not have a biopsy, 102 (93%) had scans consistent with IPF, two were indeterminate, and six (5%) were inconsistent with IPF.

### Identification of Specific HRCT Findings

**Honeycombing.** Honeycombing was considered to be present by at least one of two core radiologists in 287 (91.1%) of 314 scans (Table 2). Comparison of the first two readers revealed agreement regarding the presence or absence of honeycombing in 223 (71.7%) scans and disagreement in 88 (28.3%) scans ( $\kappa = 0.21$ ; 95% CI, 0.09–0.32). There were four scans in which there was a lack of two evaluable readings regarding honeycomb.

Comparison of interpretation by study-site radiologists and core radiologists revealed that honeycombing was considered to be present in 263 (83.8%) versus 287 (91.4%) scans, respectively ( $\kappa = 0.31$ ; 95% CI, 0.16–0.45; Table 2). The presence of honeycombing as assessed by study-site radiologists was corroborated by core radiologists in 251 (95.4%) of 263 instances. However,

**TABLE 1. IMAGING DIAGNOSIS OF IDIOPATHIC PULMONARY FIBROSIS BY CORE RADIOLOGISTS**

	Diagnosis of the First Two Readers*	Consensus Diagnosis Based on Up to Three Reviews
Consistent with IPF	256 (81.3%)	283 (89.8%)
Inconsistent with IPF	15 (4.8%)	30 (9.5%)
Lack of agreement	44 (14.0%)	2 (0.6%)

*Definition of abbreviation:* IPF = idiopathic pulmonary fibrosis.

\* For comparison of the interpretation by the first two readers using two categories (i.e., consistent with IPF, inconsistent with IPF), agreement = 271 (86.0%),  $\kappa = 0.33$ , 95% confidence intervals = 0.18–0.48.



**Figure 1.** High-resolution computed tomography (HRCT) image through the lower lungs shows peripheral predominant, basal predominant reticular abnormality, with honeycombing and traction bronchiectasis and bronchiolectasis. These appearances are typical for idiopathic pulmonary fibrosis (IPF).

honeycombing was found by core radiologists in 36 (70.6%) of the 51 scans in which study-site radiologists considered honeycombing to be absent.

**Other CT features.** Of the nine individual CT features for which presence and extent were evaluated, only two had a mean pattern extent score of more than 1: reticulation (mean,  $1.7 \pm 0.6$ ) and overall extent of fibrosis (mean,  $1.9 \pm 0.6$ ). Decreased



**Figure 2.** HRCT image through the lower lungs shows ground glass abnormality, with centrilobular nodularity. There is no reticular abnormality or honeycombing. These appearances are inconsistent with IPF.

**TABLE 2. HONEYCOMBING: COMPARISON OF STUDY-SITE RADIOLOGIST AND CORE RADIOLOGIST ASSESSMENTS**

Interpretation of Study-Site Radiologist	Interpretation of Core Radiologist*		Total
	Present	Absent	
Present	251 (79.9%)	12 (3.8%)	263
Absent	36 (11.5%)	15 (4.8%)	51
Total	287	27	314

\* Honeycombing considered present if mean score  $> 0$  (i.e., either expert thoracic reader observed honeycombing). For the comparison of the first two expert thoracic reader interpretations, agreement = 223 (71.7%),  $\kappa = 0.21$ , 95% confidence intervals = 0.09–0.32. For the comparison of medical center radiologist versus expert thoracic reader interpretation, agreement = 266 (84.7%),  $\kappa = 0.31$ , 95% confidence intervals = 0.16–0.45.

attenuation, centrilobular or other nodules, and consolidation were rarely noted (mean scores, 0.01–0.04); bronchiolectasis was commonly identified (seen in 97% of scans).

#### Comparison of Patients with Scans Consistent and Inconsistent with IPF

The CT findings on 283 scans considered by consensus to be consistent with IPF were compared with the 30 scans believed to be inconsistent (Table 3). Scans considered consistent with IPF were significantly more likely to show honeycombing, traction bronchiectasis and bronchiolectasis, and lower lobe volume loss, and less likely to show ground glass attenuation, decreased attenuation, mosaic attenuation, and centrilobular nodules than scans not consistent with IPF. Neither crazy paving nor tree-in-bud patterns were identified in any HRCT image in either subset of patients.

Table 4 compares demographic features, baseline physiology, and survival status in patients with scans consistent and inconsistent with IPF. Patients with scans consistent with IPF were younger and more likely to be male than those with inconsistent scans. Their mean  $DL_{CO}$  was lower, but FVC and A-a oxygen gradient were not significantly different. Although the mortality rate was lower in those with inconsistent scans (3.3 vs. 15.2%), this difference was not statistically different.

#### Association of HRCT Characteristics with Physiologic Findings

Assessment of the association of selected baseline physiologic characteristics ( $DL_{CO}$ , FVC, and A-a gradient) with key baseline HRCT features showed that percent-predicted  $DL_{CO}$  was the physiologic characteristic most consistently associated with consensus diagnosis of IPF (Table 5). In addition, there was a robust inverse association between baseline percent-predicted  $DL_{CO}$  and each of three selected radiographic characteristics (presence of honeycombing, honeycomb pattern extent score, and overall extent of fibrosis score). Baseline percent-predicted FVC was significantly inversely associated only with the extent of fibrosis score, whereas A-a gradient was positively associated with the honeycomb pattern extent score and the extent of fibrosis score but not with consensus diagnosis of IPF or presence of honeycombing.

We also compared specific physiologic characteristics in patients with consistent versus inconsistent scans for IPF. Mean baseline percent-predicted  $DL_{CO}$  was significantly lower in patients with HRCT scans that were consistent with IPF:  $36.1 \pm 9.9$  versus  $42.5 \pm 11.4\%$ ,  $p = 0.001$ . There were no marked differences in baseline percent-predicted FVC ( $64.2 \pm 11.1$  vs.  $61.0 \pm 11.1\%$ ,  $p = 0.1$ ) or A-a gradient ( $25.4 \pm 22.7$  vs.  $22.7 \pm 11.4$  mm Hg,  $p = 0.2$ ) in patients with IPF consistent versus inconsistent scans, respectively.

**TABLE 3. HIGH-RESOLUTION COMPUTED TOMOGRAPHY CHARACTERISTICS IN SCANS CONSISTENT VERSUS INCONSISTENT WITH IDIOPATHIC PULMONARY FIBROSIS**

Characteristics	Total* (n = 313)	Consistent with IPF (n = 283)	Inconsistent with IPF (n = 30)	p Value†
Honeycomb	287 (91.7)	270 (95.4)	17 (56.7)	< 0.001
Ground glass attenuation	94 (30.0)	66 (23.3)	28 (93.3)	< 0.001
Decreased attenuation	25 (8.0)	12 (4.2)	13 (43.3)	< 0.001
Centrilobular nodules	10 (3.2)	6 (2.1)	4 (13.3)	0.010
Reticulation	313 (100.0)	283 (100.0)	30 (100.0)	N/A
Emphysema	101 (32.3)	94 (33.2)	7 (23.3)	0.311
Nodules, not centrilobular	3 (1.0)	3 (1.1)	0 (0.0)	> 0.999
Consolidation	9 (2.9)	8 (2.8)	1 (3.3)	0.601
Bronchiolectasis	294 (93.9)	275 (97.2)	19 (63.3)	< 0.001
Mosaic attenuation	25 (8.0)	13 (4.6)	12 (40.0)	< 0.001
Lower lobe volume loss	224 (71.6)	210 (74.2)	14 (46.7)	0.003
Traction bronchiectasis	290 (92.7)	266 (94.0)	24 (80.0)	0.015
Upper lobe volume loss	49 (15.7)	43 (15.2)	6 (20.0)	0.439
Overall extent of fibrosis, mean score ± SD	1.90 ± 0.61	1.88 ± 0.61	2.01 ± 0.65	0.39‡

Definition of abbreviation: IPF = idiopathic pulmonary fibrosis; N/A = not applicable.

Characteristics are given as number (percentage) of patients with scans that had the finding identified.

\* Total number of patients with a diagnosis consistent with IPF (283) or inconsistent with IPF (30).

† p values derived by two-tailed Fisher's exact test.

‡ Wilcoxon rank-sum test.

### Assessment of Risk Factors for Mortality

Forty-four (13.3%) patients died during the study (7). Univariate analysis of baseline radiologic variables revealed the following to be associated with a statistically significant increase in the risk of mortality (i.e.,  $p < 0.05$ ): overall extent of fibrosis score, reticulation pattern score, honeycomb pattern score, and the presence of reticulation as the predominant HRCT pattern (Table 6). Baseline variables not significantly associated with mortality on univariate analysis included ground glass pattern score, predominant pattern of ground glass, and consensus diagnosis of IPF. Other baseline variables not associated with mortality ( $p > 0.1$ ) were emphysema pattern score, predominant pattern of honeycomb, and coexisting diagnosis of emphysema.

In addition, univariate analysis of baseline clinical variables revealed the following to be associated with a statistically significant increase in the risk of mortality (i.e.,  $p < 0.05$ ): A-a gradient, percent-predicted  $DL_{CO}$ , use of supplemental oxygen, and percent-predicted FVC (Table 6). Baseline variables not significantly associated with mortality on univariate analysis ( $p$  values between 0.05 and 0.08 for each comparison) included treatment group assignment (i.e., IFN- $\gamma$ 1b or placebo), age, sex, University of California–San Diego Shortness of Breath Questionnaire score, baseline dyspnea index, and modified Medical Research Council score.

Multivariate analysis identified three independent predictors

of mortality at entry into the study: a higher overall extent of fibrosis pattern score (hazard ratio [HR], 2.71; 95% CI, 1.61–4.55;  $p < 0.0001$ ) was associated with an increased the risk of death, whereas both a higher  $DL_{CO}$  (HR, 0.94; 95% CI, 0.90–0.98;  $p = 0.004$ ) and treatment assignment to IFN- $\gamma$ 1b rather than placebo (HR, 0.53; 95% CI, 0.28–0.99;  $p = 0.04$ ) were associated with a decreased risk of death (Table 7).

### DISCUSSION

HRCT has become a central component of the diagnostic evaluation of patients with suspected IPF (2), and guidelines for HRCT evaluation have been generated (1, 2, 13, 14). However, it is commonly recommended that interpretation be performed by expert thoracic radiologists rather than by less experienced observers (2, 15, 16). The present study, using data derived from a prospective, multinational trial comprising 58 medical centers, shows that HRCT interpretation of IPF by study-site radiologists (using predefined criteria) was confirmed by core radiologists in 90% of 315 baseline scans, and agreement on the presence or absence of honeycombing was found in 85% of the scans. Thus, the ability of study-site radiologists to diagnose IPF by HRCT in this prospective study was quite good and better than expected based on previous data (13).

There are several possible explanations for this finding. A potential bias inherent in the study design is that both the study-site radiologists and the core radiologists knew that the patients

**TABLE 4. COMPARISON OF CLINICAL FEATURES IN PATIENTS WITH SCANS CONSISTENT AND INCONSISTENT WITH IDIOPATHIC PULMONARY FIBROSIS**

	Total (n = 313)	Consistent with IPF (n = 283)	Inconsistent with IPF (n = 30)	p Value
Age, yr (mean ± SD)	63.59 ± 8.59	64.09 ± 8.43	58.90 ± 8.81	0.002*
Male sex, no. (%)	215 (68.7)	201 (71.0)	14 (46.7)	0.012†
Percent-predicted FVC (mean ± SD)	63.90 ± 11.06	64.21 ± 11.10	61.00 ± 10.44	0.131*
Percent-predicted $DL_{CO}$ (mean ± SD)	34.74 ± 10.23	36.13 ± 9.93	42.47 ± 11.43	0.001*
A-a gradient (mean ± SD)	25.17 ± 10.98	25.43 ± 10.92	22.67 ± 11.42	0.190*
Survival, no. (%) of deaths	44 (14.1)	43 (15.2)	1 (3.3)	0.096*

Definition of abbreviations: A-a gradient = alveolar–arterial oxygen gradient; IPF = idiopathic pulmonary fibrosis.

\* p value derived from two-sample t test.

† p value derived from Fisher's exact test.

**TABLE 5. ASSOCIATIONS BETWEEN BASELINE HIGH-RESOLUTION COMPUTED TOMOGRAPHY CHARACTERISTICS AND BASELINE CLINICAL CHARACTERISTICS**

Core Radiologist Interpretation	p Value <sup>†</sup>		
	% Predicted DL <sub>CO</sub>	% Predicted FVC	A-a Gradient
Consensus diagnosis of IPF, yes/no	0.002	0.11	0.23
Presence of honeycombing, yes/no	< 0.001	0.82	0.80
Overall extent of fibrosis score*	< 0.0001	0.002	< 0.0001
Honeycomb pattern extent score*	< 0.0001	0.39	< 0.0001

Definition of abbreviations: A-a gradient = alveolar-arterial oxygen gradient; DL<sub>CO</sub> = diffusing capacity of carbon monoxide; IPF = idiopathic pulmonary fibrosis.

\* Analyzed as a continuous variable.

<sup>†</sup> Derived from the Wilcoxon rank-sum test for dichotomous reader interpretations and from the Spearman rank-order correlation test for continuous scores.

were being entered into a clinical treatment trial for IPF, which may have influenced their interpretation of the CT findings. The use of predefined criteria for the assessment of the HRCT scan by the study-site radiologists likely improved the consistency of the evaluations. Also, expertise at the local level likely has improved since earlier studies. In addition, the entry criteria for this study were designed to ensure the likelihood of the recruitment of true cases of IPF, thus inflating the sensitivity of diagnosis. Furthermore, the high proportion (39 of 58) of academic medical centers participating in the trial (with likely access to subspecialized thoracic radiologists), the selection of investigational sites for their capabilities in the area of IPF, and the training provided to study sites regarding the criteria used for the study may render the study-site radiologists to be not truly representative of the general community. In fact, we found no apparent differences in the diagnostic capabilities of academic and nonacademic sites (data not shown).

Importantly, there was poor agreement between the core radiologists regarding the presence or absence of honeycombing—the most critical feature consistent with a definite diagnosis of IPF. Of note, the level of agreement between study-site radiologists and core thoracic radiologists was similar ( $\kappa = 0.31$ ) to that between core radiologists compared with each other ( $\kappa = 0.21$ ). Conceivably, this relatively low level of agreement between core radiologists regarding the presence of honeycombing reflects either the absence of standardized, prespecified

guidelines for HRCT interpretation in this study or the true difficulty with the interpretation of honeycombing in the scans of many patients with IPF. Because the  $\kappa$  value may be artificially reduced when the prevalence of disease is either low or high, this might have accounted in part for the relatively low  $\kappa$  value for most of the comparisons in this study (17). In future studies, the use of standardized CT images for various CT patterns, analogous to those used in the International Labour Organization (ILO) classification system for chest radiographs, may help reduce the discrepancy. Alternatively, computerized, automated methods of disease characterization and quantification may become helpful (18, 19).

Scans considered consistent with IPF were significantly more likely to show honeycombing, traction bronchiectasis and bronchiolectasis, and lower lobe volume loss, and less likely to show ground glass attenuation, decreased attenuation, mosaic attenuation, and centrilobular nodules than scans not consistent with IPF. These findings are not surprising given the established diagnostic criteria for IPF. The male predominance, and slightly older age, in patients with consistent scans reflects the typical demographics of IPF. The lower DL<sub>CO</sub> found in these patients suggests that they have more physiologic impairment than those with scans inconsistent with IPF. A limitation of this analysis is that we did not prospectively define the CT criteria for atypical IPF.

An analysis of the association between the histologic and radiologic diagnosis of IPF (available in 65% of cases) revealed a consensus diagnosis of IPF by core radiologists in 88% of patients with a reported histologic diagnosis of UIP before study entry, confirming the reliability of a typical UIP pattern seen on HRCT in predicting the underlying UIP pathology (15). This degree of concordance is higher than those cited in previous studies (64 and 74% in References 17 and 18, respectively). It is known that histologic confirmation of UIP on surgical lung biopsy can be identified in a subset of patient who do not have HRCT findings consistent with a definite diagnosis of IPF (6, 15). However, the lack of centralized and standardized pathologic

**TABLE 6. UNIVARIATE ANALYSIS OF THE PREDICTORS OF MORTALITY**

Baseline Variable	Hazard Ratio <sup>‡</sup>	95% Confidence Interval of Hazard Ratio		p Value <sup>§¶</sup>
		Ratio	Ratio	
<b>HRCT features</b>				
Overall disease extent score*	3.12	2.00, 4.89		< 0.0001
Reticulation pattern score*	2.69	1.71, 4.23		< 0.0001
Honeycomb pattern score*	3.06	1.75, 5.34		0.0001
Predominant pattern: reticulation <sup>†</sup>	0.41	0.17, 0.99		0.04
<b>Other clinical features</b>				
Percent-predicted DL <sub>CO</sub> *	0.92	0.89, 0.96		0.0001
A-a gradient*	1.06	1.03, 1.09		< 0.0001
Current oxygen use <sup>†</sup>	2.37	1.29, 4.34		0.004
Percent-predicted FVC*	0.97	0.94, 1.00		0.03

Definition of abbreviation: DL<sub>CO</sub> = diffusing capacity of carbon monoxide; HRCT = high-resolution computed tomography.

\* Continuous scale.

<sup>†</sup> Dichotomous variable.

<sup>‡</sup> Hazard ratio calculated by the Cox proportional hazards model, stratifying by smoking status.

<sup>§</sup> p value derived from the Wald  $\chi^2$  test on scores, stratifying by smoking status.

<sup>¶</sup> Statistically significant results (i.e., p < 0.05) presented only.

**TABLE 7. MULTIVARIATE ANALYSIS OF THE PREDICTORS OF MORTALITY**

Variable	Hazard Ratio*	95% Confidence Interval	p Value <sup>†</sup>
Overall extent of fibrosis score	2.71	1.61, 4.55	< 0.0001
Baseline % predicted DL <sub>CO</sub>	0.94	0.90, 0.98	0.004
Treatment assignment to IFN- $\gamma$ 1b	0.53	0.28, 0.99	0.04

Definition of abbreviation: DL<sub>CO</sub> = diffusing capacity of carbon monoxide.

\* Hazard ratio calculated by the Cox proportional hazards model, stratifying by smoking status.

<sup>†</sup> p value derived from the Wald  $\chi^2$  test, stratifying by smoking status.

reading of surgical lung biopsies by consensus of an expert panel of pathologists may have biased this estimate.

We found that several physiologic characteristics correlated well with key HRCT findings at the time of study entry, especially baseline percent-predicted  $DL_{CO}$ . This is of particular interest given that  $DL_{CO}$  was one of only three independent predictors of mortality identified in the multivariate analysis, suggesting that baseline  $DL_{CO}$  remains an important clinical indicator, even when adjusted for HRCT findings. In addition, a higher overall extent of fibrosis on HRCT was associated with a 2.7-fold increased risk of mortality in the multivariate analysis ( $p < 0.0001$ ). These two findings agree with those of Mogulkoc and colleagues (20), who identified percent-predicted  $DL_{CO}$  and the HRCT fibrosis score as the sole two independent predictors of 2-year survival in 115 patients with UIP.

The third independent predictor of mortality in our analysis was randomized assignment to therapy with IFN- $\gamma$ 1b rather than placebo in the phase 3 study, which was associated with a significantly reduced risk of death (HR, 0.53; 95% CI, 0.28–0.99;  $p = 0.04$ ). This finding mirrors the trend toward increased overall survival in patients receiving IFN- $\gamma$ 1b identified in the primary analysis of the trial (HR, 0.6; 95% CI, 0.3–1.1;  $p = 0.08$ ) (10). However, it should be emphasized that the current analysis is exploratory.

Our findings support the usefulness of HRCT as an integral part of the evaluation of patients with suspected IPF and suggest that expertise in radiologic interpretation may be extending to the level of regional medical centers. Also, our results identify that the overall extent of lung fibrosis on HRCT (i.e., combined extent of reticulation and honeycomb change) is a strong independent predictor of mortality in patients with IPF.

**Conflict of Interest Statement:** D.A.L. received \$20,300 in 2002 and \$3,500 in 2003 for reading CT scans, and \$1,500 as a lecture honorarium from InterMune, Inc.; he also received approximately \$3,000 in 2004 for reading CT scans from Actelion, Inc. J.D.G. was paid as a consultant to InterMune for blinded reading of CT scans, amounting to \$23,880 in 2002 and 2003. He was paid an honorarium for a lecture to InterMune guests at the American College of Chest Physicians meeting October 2004. S.S. was employed by InterMune, Inc., at the time of this study analysis. She holds a total of 100 shares of InterMune stock. K.M.S. was an InterMune employee and had stock options; she is no longer an employee of InterMune and has no stock options. P.H. has been paid approximately \$400,000 since January 2003 for biostatistical consulting work for InterMune, Inc. K.K.B. has served as a consultant and/or advisory board member to InterMune, Wyeth, Actelion, Fibrogen, Genzyme, Arizeke, and Encysive. He received research support for the direct costs associated with multicentered clinical research trials from the following companies: InterMune, Wyeth, Actelion, and Fibrogen. G.R. served as a consultant, advisor, and steering committee member for clinical trials, and received \$1,000 (lecture), \$500 (consultation) in 2001; \$1,500 (advisor) and \$3,125 (lecture) in 2002, \$2,500 (advisor), \$500 (consultation) in 2003, and \$2,000 (lecture) in 2003 from InterMune, Inc. He received reimbursement for meeting and travel expenses for \$8,260 in 2002 and \$732.05 in 2003 from InterMune. He also received \$1,000 (consultant, advisor) from Shronogi in 2002. He received honoraria of \$2,000 in 2003 and \$6,000 in 2004 for CME lectures related to idiopathic interstitial pneumonia. T.E.K. has been involved in scientific advisory boards for Actelion (compensation in 2003 = \$11,725; 2004 = \$9,940). He received \$16,000 in 2002, \$21,000 in 2003, and \$15,000 in 2004 from InterMune. He received \$3,250 in 2002, \$7,500 in 2003, and \$5,000 in 2004 from Nektar, and \$12,625 from GlaxoSmithKline in 2004. He has also served as a consultant for Alexza, AstraZeneca, Biogen, Centocor, Fibrogen, Genzyme, Human Genome Sciences, Merck, and CoTherix; none of these consulting fees exceeded \$10,000 per company per year in any one of the 3 years preceding submission of this manuscript. W.Z.B. is an employee of InterMune, owns stock, and has stock options. D.A.S. received \$16,000 over the past 4 years from InterMune for assistance with work for studies involved with IFN- $\gamma$  and pulmonary fibrosis. W.R.W. received \$20,000 in 2002–2003 for the interpretation of approximately 200 paired CT scans and \$2,500 in 2004 for serving on the pulmonary advisory board from InterMune, Inc.

**Acknowledgment:** The authors thank the investigators, staff, and participating patients in the phase 3 study.

## References

1. American Thoracic Society/European Respiratory Society. Idiopathic pulmonary fibrosis: diagnosis and treatment [international consensus statement]. *Am J Respir Crit Care Med* 2000;161:646–664.
2. American Thoracic Society/European Respiratory Society. International multidisciplinary consensus classification of the idiopathic interstitial pneumonias. *Am J Respir Crit Care Med* 2002;165:277–304.
3. Tung KT, Wells AU, Rubens MB, Kirk JME, du Bois RM, Hansell DM. Accuracy of the typical computed tomographic appearances of fibrosing alveolitis. *Thorax* 1993;48:334–338.
4. Mathieson JR, Mayo JR, Staples CA, Müller NL. Chronic diffuse infiltrative lung disease: comparison of diagnostic accuracy of CT and chest radiography. *Radiology* 1989;171:111–116.
5. Gay SE, Kazerooni EA, Toews GB, Lynch JP III, Gross BH, Cascade PH, Spizarny DL, Flint A, Schork MA, Whyte RI, et al. Idiopathic pulmonary fibrosis: predicting response to therapy and survival. *Am J Respir Crit Care Med* 1998;157:1063–1072.
6. Flaherty KR, Thwaite EL, Kazerooni EA, Gross BH, Toews BG, Colby TV, Travis WD, Mumford S, Flint A, Lynch JP III, et al. Radiological versus histological diagnosis in UIP and NSIP: survival implications. *Thorax* 2003;58:143–148.
7. Raghu G, Brown KK, Bradford WZ, Starko K, Noble PW, Schwartz DA, King TE. A placebo-controlled trial of interferon gamma-1b in patients with idiopathic pulmonary fibrosis. *N Engl J Med* 2004;350:125–133.
8. Lynch D, Godwin JD, Webb WR, Brown KK, Bradford WZ, Noble PW, Raghu G, Schwartz DA, Starko KM. HRCT diagnosis of IPF and relationship between HRCT characteristics and histology and clinical findings in patients in a phase 3, randomized, double-blind, placebo-controlled trial of interferon gamma-1b. *Chest* 2004;126:773S.
9. Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics* 1977;33:159–174.
10. Mahler DA, Weinberg DH, Wells CK, Feinstein AR. The measurement of dyspnea: contents, interobserver agreement, and physiologic correlates of two new clinical indices. *Chest* 1984;85:751–758.
11. Eakin EG, Resnikoff PM, Prewitt LM, Ries AL, Kaplan RM. Validation of a new dyspnea measure: the UCSD Shortness of Breath Questionnaire. University of California, San Diego. *Chest* 1998;113:619–624.
12. Mahler DA, Wells CK. Evaluation of clinical methods for rating dyspnea. *Chest* 1988;93:580–586.
13. Grenier P, Chevret S, Beigelman C, Brauner MW, Chastang C, Valeyre D. Chronic diffuse infiltrative lung disease: determination of the diagnostic value of clinical data, chest radiography, and CT with Bayesian analysis. *Radiology* 1994;191:383–390.
14. Bessis L, Callard P, Gotheil C, Biaggi A, Grenier P. High-resolution CT of parenchymal lung disease: precise correlation with histologic findings. *Radiographics* 1992;12:45–58.
15. Hunninghake GW, Zimmerman MB, Schwartz DA, King TE Jr, Lynch J, Hegele R, Waldron J, Colby T, Müller N, Lynch D, et al. Utility of a lung biopsy for the diagnosis of idiopathic pulmonary fibrosis. *Am J Respir Crit Care Med* 2001;164:193–196.
16. Aziz ZA, Wells AU, Hansell DM, Bain GA, Copley SJ, Desai SR, Ellis SM, Gleeson FV, Grubnic S, Nicholson AG, et al. HRCT diagnosis of diffuse parenchymal lung disease: inter-observer variation. *Thorax* 2004;59:506–511.
17. Kundel HL, Polansky M. Measurement of observer agreement. *Radiology* 2003;228:303–308.
18. 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.
19. 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.
20. Mogulkoc N, Brutsche MH, Bishop PW, Greaves SM, Horrocks AW, Egan JJ. Pulmonary function in idiopathic pulmonary fibrosis and referral for lung transplantation. *Am J Respir Crit Care Med* 2001;164:103–108.