ERJ Express. Published on December 5, 2007 as doi: 10.1183/09031936.00063706

Original article

Accuracy of Diagnosis of Idiopathic Pulmonary Fibrosis

Michiel Thomeer[1,2], Maurits Demedts[1], Juergen Behr[3], Roland Buhl[4], Ulrich Costabel[5], Christopher D.R. Flower[6], Johny Verschakelen[1], François Laurent[7], Andrew G. Nicholson[8], Eric K.Verbeken[1], Frédérique Capron[9], Marco Sardina[10], Giuseppina Corvasce[10], Ida Lankhorst[10], and the IFIGENIA study group*

*members of the IFIGENIA study group are listed in appendix

The IFIGENIA study is sponsored by the Zambon Group, Milan, Italy

1. Universitaire Ziekenhuizen, Katholieke Universiteit Leuven, Belgium
2. Ziekenhuis Oost Limburg, Genk, Belgium
3. Medizinische Klinik I, Klinikum Grosshadern der Ludwig-Maximilians-Universität, Munich, Germany
4. Medizinische Klinik III, Klinikum der Johannes-Gutenberg-Universität, Mainz, Germany
5. Abteilung Pneumologie/Allergologie, Ruhrlandklinik, Essen-Heidhausen, Germany
6. Evelyn Hospital, Cambridge, UK
7. Hôpital Cardiologique, CHU de Bordeaux, France
8. Royal Brompton Hospital, London, UK
9. Hôpital Antoine-Béclère, Clarmart, France
10. Zambon Group, Bresso, Milan, Italy

Address for correspondence:
Dr. Michiel Thomeer
UZ Leuven, Afdeling Longziekten
Herestraat 49
B - 3000 Leuven, Belgium
tel: +32 16 34 68 00, fax +32 16 34 68 03
Email: michiel.thomeer@scarlet.be

Running head: inter-observer agreement for the diagnosis of idiopathic pulmonary fibrosis

Keywords:
idiopathic pulmonary fibrosis, kappa coefficient, radiology, lung biopsy

Abbreviations

95% CI: 95% confidence interval

FVC: forced vital capacity in % predicted

HRCT: high resolution computed tomography scan of the lung

IIP: idiopathic interstitial pneumonias

IPF: idiopathic pulmonary fibrosis

κ_w: weighted kappa

OLB: open lung biopsy

TBB: Transbronchial lung biopsy

TLB: thoracoscopic lung biopsy

UIP: usual interstitial pneumonia

Abstract

The purpose of this study is to evaluate the accuracy of the diagnosis of idiopathic pulmonary fibrosis by respiratory physicians in 6 European countries and to calculate the inter-observer agreement between HRCT reviewers and histology reviewers in the diagnosis of IPF. After the diagnosis was assessed by the local investigator, following the ATS/ERS consensus statement, the diagnosis of usual interstitial pneumonia was confirmed when a minimum 2 of 3 expert reviewers of each expert panel agreed with the diagnosis. The level of agreement between the readers within each expert panel was calculated by weighted kappa. The diagnosis of UIP was confirmed in 87.2% of the cases by the expert panels. A total of 179 HRCT scans were independently reviewed and an inter-observer agreement of 0.40 was found. In 97 patients an open or thoracoscopic biopsy was performed, 82 of these could be reviewed by the expert committee. The weighted kappa between the histology readers was 0.30. We conclude that although the level of agreement between readers within each panel is only fair to moderate, the overall accuracy of a clinical diagnosis of IPF in expert centers is good (87,2 %).

Introduction

Idiopathic pulmonary fibrosis (IPF) is a specific form of a chronic fibrosing interstitial pneumonia limited to the lung and is typically characterized by the histologic appearance of usual interstitial pneumonia on (open or thoracoscopic) lung biopsy [1]. The clinical diagnosis of IPF is based on the exclusion of known causes of interstitial lung disease, a restrictive lung function pattern with an impaired gas exchange and the presence of a typical pattern of bibasilar reticular abnormalities with minimal ground glass opacities on high resolution computed tomography scans (HRCT) [1].

Patients with IPF have a worse survival than patients with other types of idiopathic interstitial pneumonias (IIP) [2-4]. Because the diagnosis of IPF depends on the expertise of the pathologist and the radiologist, it is important for the clinician to know the accuracy of diagnosis in usual interstitial pneumonia (UIP) on HRCT and on lung biopsy. Different studies have calculated the accuracy of HRCT in fibrotic lung diseases [5-7], evaluated the inter-observer agreement for the diagnosis of different HRCT patterns (e.g. ground glass and reticular pattern) [8,9] on HRCTs from patients with a biopsy proven non-specific interstitial pneumonia or UIP [3], or on different forms of interstitial lung diseases [10,11]. Studies on inter-observer agreement amongst pathologists are sparse [12] and only one study with a multicenter prospective design addressed the issue of accuracy of diagnosis of UIP in relation to both radiologist and pathologist [7]. No study addressed this issue in view of the new ATS/ERS criteria[1]. Therefore, the aim of this study was to evaluate the accuracy of the diagnosis of IPF by respiratory physicians, and to calculate the inter-observer agreement between HRCT reviewers and histology reviewers in the diagnosis of UIP.

Methods

Patients

All patients presented in this study were included in the IFIGENIA trial [13]. The IFIGENIA trial is a European prospective, double-blind, placebo-controlled trial studying the effect of high dose N-acetylcysteine in combination with standard therapy (prednisone and azathioprine) in patients with idiopathic pulmonary fibrosis (IPF). Following the judgment of the local investigator, patients were included if the diagnosis of IPF was based on the international consensus criteria [1], patients were aged between 18 and 75 years. Newly diagnosed (< 6 months) as well as previously diagnosed (> 6 months) patients were considered for the study. The IFIGENIA trial was approved by local ethical committee of the participating centres and every patient signed informed consent.

HRCT scanning protocol

HRCT of the thorax was performed in supine position during breath holding at full inspiration with 1 mm or 1.5 mm thick sections at 1 cm intervals throughout the entire thorax. Images were reconstructed with high frequency algorithm at window levels appropriate for pulmonary parenchyma (mean -500 to -700 Hounsfield units; width 1400 to 2000 Hounsfield units). No intravenous contrast was administered.

Review by the radiology committee

The local investigator provided copies of the original HRCT scan and sent these to the international trial coordinator (G.C.). One copy was sent to the 3 members of the radiology committee (C.F., F.L. and J.V.). The copies of the HRCT scans were reviewed independently without knowledge of clinical, physiologic or pathologic parameters. The international trial coordinator took care that the 3 members of the radiology committee (Reviewers A, B and C) were unaware of the patients' identity. Each member of the

committee confirmed the diagnosis of UIP on the HRCT based on the criteria of the international consensus statement [1]. The degree of confidence in the diagnosis was recorded as very suggestive, probable or unlikely for the diagnosis. The diagnosis of UIP on HRCT was confirmed if it was scored as “very suggestive” or “probable” for UIP, and rejected if it was scored as “unlikely”. If a disagreement occurred in the diagnosis of UIP between the three members of the radiology committee, the diagnosis agreed by the majority of the three members was accepted as definite.

Review of lung biopsies by a histology committee

The diagnosis of usual interstitial pneumonia (UIP) using the criteria of the ATS/ERS Consensus Classification [1], was assessed by an independent panel of 3 pathology experts (A.G.N., E.K. and F.C.). The local investigator sent slides from open (OLB) or thoracoscopic (TLB) lung biopsies to the international trial coordinator, who blinded the cases and sent them to 2 members of the pathology review committee (Reviewers D and E). All slides were graded as very suggestive, probable or unlikely for the diagnosis of UIP. For each observer, the diagnosis of UIP on lung biopsy was confirmed if it was scored as “very suggestive” or “probable”, and rejected if scored as “unlikely”. If the 2 reviewers disagreed on the diagnosis of UIP, the slides were sent to the third member of the pathology committee and assessed in an identical fashion (Reviewer F). The diagnosis agreed by the majority of the 3 members was accepted as final. The slides were reviewed independently without knowledge of clinical, physiologic or parameters.

Definite diagnosis of UIP

The diagnosis of UIP was rejected when one or both committees did not confirm a diagnosis of UIP.

Statistics

Weighted kappa coefficients (κ_w) were used to measure the level of inter-observer agreement. The weighted kappa coefficients were calculated using a method recommended for comparing level of agreement with categorical data [14]. κ_w and respective 95% confidence intervals were calculated using SAS v 6.12 software (SAS Institute Inc, North Carolina, USA).

Results

A total of 36 local investigators from 6 European countries included (Table 1) leaving 179 HRCT's and 82 open or thoracic biopsies for review.

Radiology reviewer A reviewed 178 HRCT (1 HRCT never reviewed), reviewer B 176 (2 HRCT judged as not interpretable and 1 never reviewed) and reviewer C 176 (2 HRCT judged as not interpretable and 1 never reviewed) (Figure 1). After combining the observations of all 3 radiologists, 532 HRCT observations were judged as unlikely in 67 (12.6%), probable in 203 (38.2%) and very suggestive in 258 (48.5%) for the diagnosis of UIP. In 4 observations (0.8%) the HRCT was judged as not interpretable because of lack of quality. A total of 238 HRCT observations could be correlated with the results of lung biopsy: When the HRCT was judged as unlikely for UIP, 67.5% of the corresponding lung biopsy was positive for UIP, 84.4% when judged as probable and 91.7% when judged as very suggestive (Figure 2).

All 82 biopsies (44 OLB and 38 TLB) were sent to the international trial co-ordinator for review by pathology reviewers D and E. After combining the observations of the 3 histology reviewers, 178 OLB/TLB observations were judged as unlikely in 33 (18.5%), probable in 66 (37.1%) and very suggestive in 76 (42.7%) for the diagnosis of UIP. In 3 observations (1.7%) the biopsy slide was judged as not interpretable. Reviewer D reviewed all 82 OLB/TLB, reviewer E 79 (3 judged as not interpretable). Histology reviewer F was solicited to review 14 biopsy slides (Figure 1).

In 12.8% of the patients the diagnosis of UIP was rejected (Table 1, Figure 1) by at least one review committee. The diagnosis of UIP was confirmed by the pathology review committee in 84% of the 82 OLB/TLB. The diagnosis of UIP on HRCT was confirmed in 92.7% of the 165 HRCT (Figure 1).

Table 2a summarizes the level of agreement between the 3 different HRCT reviewers. The κ_w ranged from 0.33 to 0.46. No important differences in κ_w are seen within the different subgroups. Table 2b summarizes the level of agreement between the 2 pathology reviewers: a κ_w of 0.30 (95% CI 0.12-0.48) was calculated. The level of agreement was 0.84 (%95 CI 0.55-1.14) in the subgroup of those patients in whom the diagnosis of UIP on HRCT was not confirmed. When the severity of lung function impairment (FVC more or less than 60% predicted) was taken into account, no difference in level of agreement was observed.

Discussion

Two salient findings emerge from our study. First, the diagnosis of IPF proposed by a respiratory specialist was rejected in 12.8% after reviewing histology and HRCT by expert committee. Second, the mean level of agreement between the 3 different HRCT reviewers was 0.40 and for the two pathology reviewers 0.30.

The accuracy of the diagnosis of IPF by a pulmonary physician in relation to the ATS/ERS diagnostic criteria [1] has not been established. A confident diagnosis of IPF proposed by a clinician was confirmed in our study in 87.2%. The rejection of the diagnosis was not based on the clinical criteria, but on the findings that HRCT and/or lung biopsy were not compatible with the diagnosis of UIP. Hunninghake et al found a probability that a patient has given a confident diagnosis by the referring clinician of 81% [7], which is similar to our number. Although the study of Hunninghake et al published the first prospective multicentre study about the level of agreement between clinicians, radiologists and pathologists on the diagnosis of IPF, it is not clear from their study on which clinical grounds the diagnosis of IPF was made, because no clinical or radiological criteria were provided for the diagnosis of IPF.

The study addressed also the question of agreement between histology reviewers in the diagnosis of UIP in view of the new pathological classification [1]. The inter-observer agreement between the histology reviewers was low with a mean κ_w of 0.30, a level scored as 'fair' following the proposed interpretation of kappa scores by Brennan et al [14]. The kappa score is a score between zero and 1, zero indicating only chance agreement and 1 indicating perfect agreement [14]. Nicholson et al. [15] and Cherniack et al. [16] studied levels of agreement between pathologists for individual histologic parameters (e.g. extent of fibrosis) in biopsies showing UIP and found a kappa score rang-

ing from 0.56-0.76 and –0.06-0.30 respectively. However, because the aim for which their kappa values were calculated was different from ours, these results are not comparable with the present study. Nicholson et al. [17] presented another study examining the prognostic significance of histologic patterns of IIP. Slides of 37 lung biopsies with UIP, 28 with NSIP and 13 with desquamative interstitial pneumonia or respiratory bronchiolitis interstitial lung disease were reviewed independently by 2 pulmonary histopathologists. They found an overall kappa score of 0.49, but the level of agreement was 0.26 in distinguishing between UIP and NSIP. This last number is comparable, in view of the selection of patients, with the present study. It suggests that distinguishing a UIP or NSIP pattern on histology is difficult and perhaps more difficult with the knowledge of lobar histopathologic variability in UIP and NSIP in the same lung [18]. This finding is confirmed in a recent study [12] where an observer agreement on UIP of 0.49 and on NSIP of 0.32 as final histological diagnosis was found.

The level of agreement between the HRCT readers was fair to moderate [14]. This kappa-score is comparable with the scores from MacDonald et al. [9] and from Flaherty et al [3] (table 3), which compared the inter-observer agreement for HRCT of patients with NSIP and UIP. Hunninghake et al found an inter-observer agreement of 0.54 [7]. Because in this last study the radiologic criteria for the diagnosis of UIP on HRCT were not mentioned, we cannot explain the difference of their kappa score and of MacDonald et al, Flaherty et al and the current study. Others [8,10,19,20] however, have found a higher level of agreement but the study population and aim for which the observer variability was calculated, differed significantly from the present study (table 3). Aziz et al found an observer agreement on the first choice diagnosis of a cohort of 131 patients with diffuse parenchymal lung disease of 0.48 [21], a kappa score in the IPF cohort was 0.50.

It is important to emphasize that radiologists with differing levels of experience and expertise can interpret radiographic images differently. The radiologists in this study are specialists in thoracic imaging and have extensive expertise in the interpretation of HRCT scans. Each reader was blinded from the clinical parameters and the reading was performed separately, so that the different readers could not influence each other. The kappa coefficient is used to evaluate observer variability in order to remove the component of agreement attributable to chance. While this method of statistical analysis allows a more accurate assessment of observer variability than does unadjusted data, a kappa value may underestimate a high level of agreement [14]. In the present study the level of agreement between the different readers was unexpectedly low. Could this be due to the high prevalence of the disease in the study population or could it be by observer variation bias? The interpretation of κ_w depends on the prevalence of the disease which was high in this study (0.84) [14]. The prevalence of the disease was high because the HRCT and lung biopsy slides were from patients selected by a local investigator who confirms the diagnosis of IPF conform the ATS/ERS criteria beforehand. The higher prevalence of the disease in our study population is a possible explanation of the conflicting finding that 67% histology confirmed UIP whose HRCT are reported as unlikely. This is not astonishing however, since a recent publication reported that 59 % of patients with a definite or probably NSIP on CT had a histology diagnosis of UIP [4]. The current authors assume that many of the CT scans that had been reported as being unlikely UIP would fulfil the CT criteria for NSIP.

The results of the presented study may evoke a concern about the diagnostic accuracy of IPF. This form of lung fibrosis is a rare disease and no single accurate test for the diagnosis IPF exists. Studies of accuracy of diagnosis of IPF are performed mostly in tertiary referral centres. Even in these studies, an important interobserver variability exists.

In most of these studies as in ours, prior knowledge of the presence of a form of interstitial lung disease exists, which may incite an observer bias and therefore influence the results on the diagnostic accuracy. The incidence of IPF is low in a general pulmonary practice. Diagnostic accuracy (i.e. sensitivity and specificity) depends also from the prevalence of the disease. A lower prevalence of disease results in a higher number of false positive and false negative diagnosis. If in the future very costly therapeutic options will be on the market, the only way to ensure a as high as possible accurate diagnosis of IPF is to refer to centres with expertise in the pulmonary histology, thoracic imaging and clinical experience in IPF [22].

In summary we have shown that the accuracy of clinical diagnosis of IPF is 87.2%. Given that IPF has such a poor prognosis [2] in relation to other forms of IIP [3] we conclude that the use of an independent HRCT and histology panel to assure an accurate diagnosis of IPF, as performed in the IFIGENIA study [13], is extremely valuable and helps minimize bias. The study demonstrates that using a reviewer panel for radiology and histology in IPF trials is feasible. For the clinician it is important to know that an accurate diagnosis of IPF requires specific expertise which is available in tertiary referral centers in close collaboration of histopathologists, radiologists and clinicians.

References

1. American Thoracic Society (ATS) atERSE. Idiopathic pulmonary fibrosis: diagnosis and treatment. International consensus statement. *Am J Respir Crit Care Med* 2000; 161: 646-664.

2. Bjoraker JA, Ryu JH, Edwin MK, et al. Prognostic significance of histopathologic subsets in idiopathic pulmonary fibrosis. *Am J Respir Crit Care Med* 1998; 157: 199-203.

3. Flaherty KR, Thwaite EL, Kazerooni EA, et al. Radiological versus histological diagnosis in UIP and NSIP: survival implications. *Thorax* 2003; 58: 143-148.

4. Flaherty KR, Toews GB, Travis WD, et al. Clinical significance of histological classification of idiopathic interstitial pneumonia. *Eur Respir J* 2002; 19: 275-283.

5. Tung KT, Wells AU, Rubens MB, et al. Accuracy of the typical computed tomographic appearances of fibrosing alveolitis. *Thorax* 1993; 48: 334-338.

6. Lynch DA, Newell JD, Logan PM, et al. Can CT distinguish hypersensitivity pneumonitis from idiopathic pulmonary fibrosis? *AJR Am J Roentgenol* 1995; 165: 807-811.

7. Hunninghake GW, Zimmerman MB, Schwartz DA, et al. Utility of a lung biopsy for the diagnosis of idiopathic pulmonary fibrosis. *Am J Respir Crit Care Med* 2001; 164: 193-196.

8. Collins CD, Wells AU, Hansell DM, et al. Observer variation in pattern type and extent of disease in fibrosing alveolitis on thin section computed tomography and chest radiography. *Clin Radiol* 1994; 49: 236-240.

9. MacDonald SL, Rubens MB, Hansell DM, et al. Nonspecific interstitial pneumonia and usual interstitial pneumonia: comparative appearances at and diagnostic accuracy of thin-section CT. *Radiology* 2001; 221: 600-605.

10. Grenier P, Valeyre D, Cluzel P, et al. Chronic diffuse interstitial lung disease: diagnostic value of chest radiography and high-resolution CT. *Radiology* 1991; 179: 123-132.

11. Mathieson JR, Mayo JR, Staples CA, et al. Chronic diffuse infiltrative lung disease: comparison of diagnostic accuracy of CT and chest radiography. *Radiology* 1989; 171: 111-116.

12. Nicholson AG, Addis BJ, Bharucha H, et al. Inter-observer variation between pathologists in diffuse parenchymal lung disease. *Thorax* 2004; 59: 500-505.

13. Demedts M, Behr J, Buhl R, et al. High-dose acetylcysteine in idiopathic pulmonary fibrosis. *N Engl J Med* 2005; 353: 2229-2242.

14. Brennan P, Silman A. Statistical methods for assessing observer variability in clinical measures. *BMJ* 1992; 304: 1491-1494.

15. Nicholson AG, Fulford LG, Colby TV, et al. The relationship between individual histologic features and disease progression in idiopathic pulmonary fibrosis. *Am J Respir Crit Care Med* 2002; 166: 173-177.

16. Cherniack RM, Colby TV, Flint A, et al. Quantitative assessment of lung pathology in idiopathic pulmonary fibrosis. The BAL Cooperative Group Steering Committee. *Am Rev Respir Dis* 1991; 144: 892-900.

17. Nicholson AG, Colby TV, du Bois RM, et al. The prognostic significance of the histologic pattern of interstitial pneumonia in patients presenting with the clinical entity of cryptogenic fibrosing alveolitis. *Am J Respir Crit Care Med* 2000; 162: 2213-2217.

18. Flaherty KR, Travis WD, Colby TV, et al. Histopathologic variability in usual and nonspecific interstitial pneumonias. *Am J Respir Crit Care Med* 2001; 164: 1722-1727.

19. Wells AU, Rubens MB, du Bois RM, et al. Serial CT in fibrosing alveolitis: prognostic significance of the initial pattern. *AJR Am J Roentgenol* 1993; 161: 1159-1165.

20. Kazerooni EA, Martinez FJ, Flint A, et al. Thin-section CT obtained at 10-mm increments versus limited three-level thin-section CT for idiopathic pulmonary fibrosis: correlation with pathologic scoring. *AJR Am J Roentgenol* 1997; 169: 977-983.

21. Aziz ZA, Wells AU, Hansell DM, et al. HRCT diagnosis of diffuse parenchymal lung disease: inter-observer variation. *Thorax* 2004; 59: 506-511.

22. Thomeer M. External validity of randomised controlled trials in idiopathic pulmonary fibrosis. *Eur Respir J* 2006; 27: 1072.

Tables

N (%)		Belgium	France	Germany	Italy	Nether-lands	Spain	Total
Diagnosis of IPF by HRCT n=179	No IPF	2 (10.0)	4 (10.0)	3 (4.8)	1 (5.9)	2 (10.5)	2 (9.5)	**14 (7.3)**
	Yes IPF	18 (90.0)	36 (90.0)	59 (95.2)	16 (94.1)	17 (89.5)	19 (90.5)	**165 (92.7)**
Diagnosis of UIP by OLB/TLB n=82	No UIP	1 (8.3)	4 (26.7)	5 (29.4)	1 (14.3)	0 (0.0)	3 (25.0)	**14 (16.0)**
	Yes UIP	11 (91.7)	11 (73.3)	12 (70.6)	6 (85.7)	19 (100.0	9 (75.0)	**68 (84.0)**
Definite diagnosis n=179	**No IPF**	**2 (10.0)**	**6 (15.0)**	**8 (12.9)**	**1 (5.9)**	**2 (10.5)**	**4 (19.0)**	**23 (12.8)**
	Yes IPF	**18 (90.0)**	**34 (85.0)**	**54 (87.1)**	**16 (94.1)**	**17 (89.5)**	**17 (81.0)**	**156 (87.2)**

Table 1: Diagnosis of IPF in different subgroups. The definite diagnosis is defined when histology and radiology committees agreed with the diagnosis of IPF when a OLB/TLB and HRCT were available, or when the radiology committee agreed with the diagnosis of IPF when only a HRCT was available.
OLB: open lung biopsy, TLB: thoracoscopic lung biopsy, HRCT: high resolution computed tomography of the thorax, IPF: idiopathic pulmonary fibrosis, UIP: usual interstitial pneumonia.

HRCT reviewers	mean κ_w	A vs B κ_w (95% CI)	A vs C κ_w (95% CI)	B vs C κ_w (95% CI)
all	**0.40**	**0.40 (0.29-0.52)**	**0.33 (0.23-0.44)**	**0.46 (0.36-0.56)**
OLB/TLB: no UIP	**0.41**	0.47 (0.03-0.92)	0.24 (-0.21-0.69)	0.52 (0.19-0.86)
OLB/TLB: yes UIP	**0.35**	0.35 (0.16-0.54)	0.21 (0.03-0.40)	0.49 (0.33-0.66)
OLB/TLB yes+no UIP	**0.40**	0.41 (0.24-0.58)	0.27 (0.11-0.43)	0.53 (0.38-0.67)
TBB/ no biopsy	**0.40**	0.40 (0.24-0.56)	0.40 (0.26-0.54)	0.39 (0.26-0.53)
FVC < 60%	**0.40**	0.36 (0.16-0.56)	0.33 (0.15-0.51)	0.50 (0.34-0.67)
FVC > 60%	**0.39**	0.42 (0.28-0.56)	0.33 (0.20-0.46)	0.43 (0.30-0.56)

Table 2a: weighted kappa scores between HRCT reviewers A, B and C. Each reviewer scored a HRCT as unlikely, probable or very suggestive for the diagnosis of UIP. κ_w: weighted kappa, CI: confidence interval, FVC: forced vital capacity in % predicted HRCT: high resolution computed tomography of the thorax, OLB: open lung biopsy, TBB: transbronchial biopsy, TLB: thoracoscopic lung biopsy, UIP: usual interstitial pneumonia.

Histology reviewers	D vs E κ_w (95% CI)
all	**0.30 (0.12-0.48)**
HRCT: no UIP	0.84 (0.55-1.14)
HRCT: yes UIP	0.16 (-0.03-0.36)
FVC<60%	0.33 (0.05-0.61)
FVC>60%	0.28 (0.06-0.50)

Table 2b: weighted kappa scores between histology reviewers D and E. Every reviewer scored each OLB or TLB as unlikely, probable or very suggestive for the diagnosis of UIP. κ_w: weighted kappa, CI: confidence interval, FVC: forced vital capacity in % predicted, HRCT: high resolution computed tomography of the thorax.

	kappa	Study population	Comments
1991, Grenier [10]	0.64-0.78	sarcoidosis n = 53 pulmonary fibrosis n = 33 histiocytosis X n = 17 other ILD n = 37	- definition of IPF is not clear - 3 observers
1993, Wells [19]	0.58-0.76	systemic sclerosis n = 35 IPF n = 21	- inter-observer agreement for grading CT appearance, change in extent of disease and for nature of change - 2 observers
1994, Collins [8]	0.48	systemic sclerosis n = 63 IPF n = 63	- inter-observer agreement for pattern type on CT - 4 observers
1997, Kazerooni [20]	0.51-0.83	UIP n=24 desquamative IP n = 1	- inter-observer agreement for pattern type of different lobes - 4 observers
2001, MacDonald [9]	0.40	NSIP n = 21 UIP n = 32	- inter-observer agreement for NSIP and UIP - 4 observers
2001, Hunninghake [7]	0.54	IPF n = 54 non IPF n = 37	- inter-observer agreement for IPF versus non IPF - criteria for diagnosis of IPF not mentioned - 4 observers
2003, Flaherty [3]	0.43	NSIP n = 23 UIP n = 73	- inter-observer agreement for NSIP and UIP - 2 observers
2004, Aziz [21]	0.50	DPILD n = 131	- inter-observer agreement of first choice diagnosis of IPF - 11 observers
Current study	0.40	UIP n = 156 non UIP n = 23	- patients included with IPF following ATS/ERS criteria - inter-observer agreement for IPF versus

			non IPF - 3 observers

Table 3: overview of different studies addressing the inter-observer agreement on CT in different forms of pulmonary fibrosis.
CT: computed tomography of the thorax, DPILD: diffuse parenchymal lung disease, IPF: idiopathic pulmonary fibrosis, UIP: usual interstitial pneumonia, NSIP: non specific interstitial pneumonia.

Figures

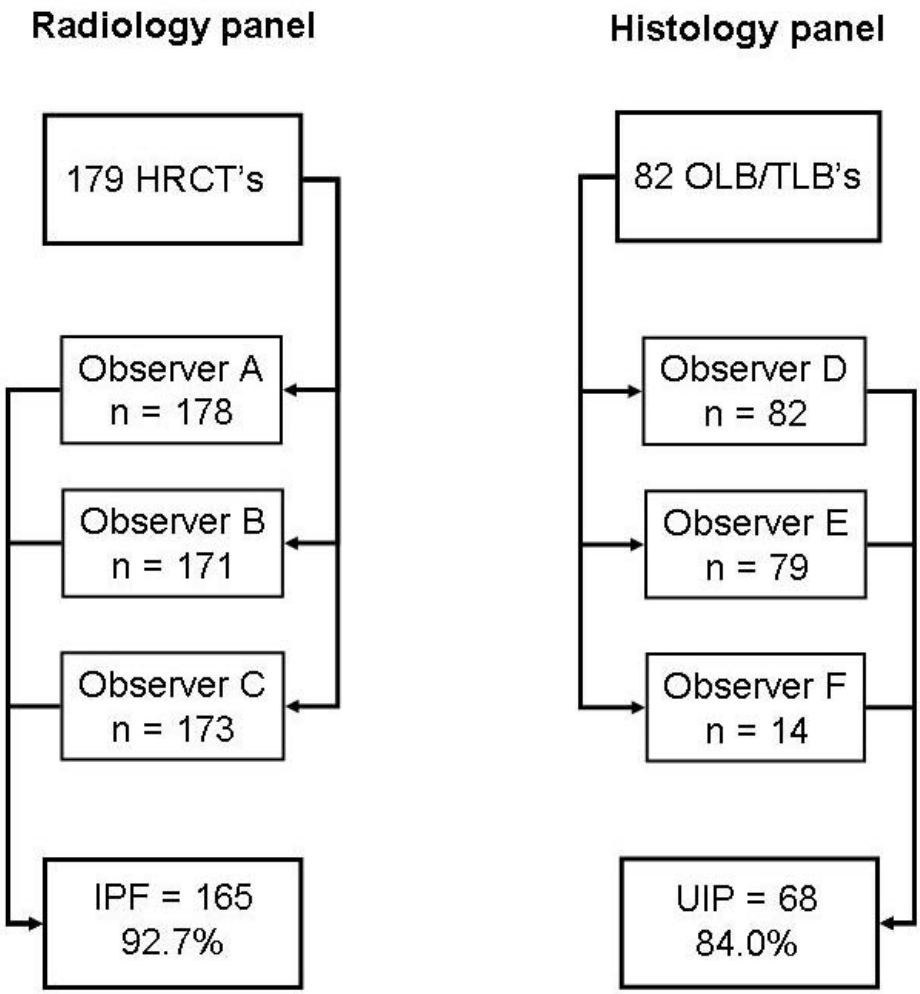

Figure 1: Study profile
Observer F was only solicited when observer D and E disagreed. HRCT: high resolution computed tomography of the thorax, IPF: idiopathic pulmonary fibrosis, UIP: usual interstitial pneumonia, OLB/TLB: open or thoracoscopic lung biopsy

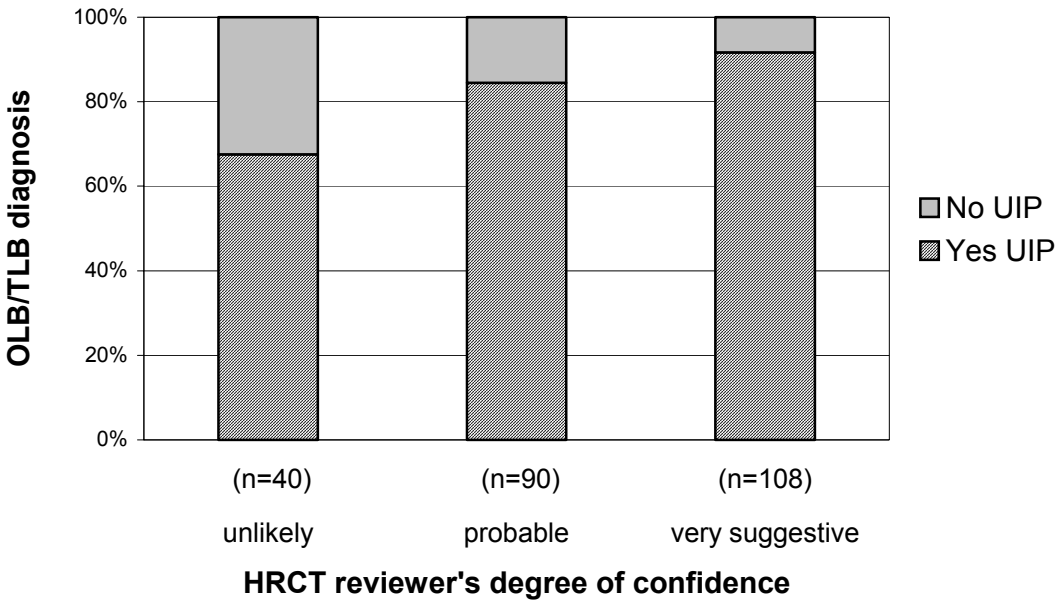

Figure 2: Proportion of UIP diagnosis on OLB/TLB correlated with the degree of confidence of UIP diagnosis on 238 HRCT observations.
HRCT: high resolution computed tomography, OLB: open lung biopsy, TLB: thoracoscopic lung biopsy, UIP: usual interstitial pneumonia.

Appendix: IFIGENIA study group

Steering committee

J.Behr, R.Buhl, U.Costabel, R.Dekhuijzen, M.Demedts (Chairman), H.M.Jansen, W.Macnee, M.Thomeer, B.Wallaert

Country co-ordinators

J.Behr (Germany), J.M.M. van den Bosch (The Netherlands), P.de Vuyst (Belgium), W.MacNee (United Kingdom), E.Rodriguez-Becerra (Spain), S.Petruzzelli (Italy), B.Wallaert (France).

Radiology review committee

C.D.R. Flower, J. Verschakelen, F. Laurent

Histology review committee

A. G. Nicholson, E.K. Verbeken, F. Capron

Local investigators

Belgium: P.De Vuyst, E.Michiels, H.Slabbynck, M.Thomeer

France: J.Cadranel, P.Camus, P.Chanez, P.Delaval, J.F.Muir, B.Wallaert

Germany: U.Costabel, J.Behr, R.Bonnet, R.Buhl, R.Loddenkemper, A.Meyer, J.Müller-Quernheim, T.Welte, H. Worth

Italy: G.Anzalone, G.B.Bottino, G.Bustacchini, M.Dottorini, S.Gasparini, C.Giuntini, A.Rossi, G.Simon

Spain: J.Ancochea Bermudez, L.Callol Sanchez, J.L.Llorente, J.M.Rodriguez-Arias, E.Rodríquez-Becerra

The Netherlands: F.Beaumont, M.Drent, H.M.Jansen, J.M.M. van den Bosch, F.J.J.van den Elshout

Zambon personnel and consultants

A. Ardia (consultant), M. Sardina, G. Corvasce, I Lankhorst (consultant)