

INTERSTITIAL LUNG DISEASE

Inter-observer variation between pathologists in diffuse parenchymal lung disease

A G Nicholson, B J Addis, H Bharucha, C A Clelland, B Corrin, A R Gibbs, P S Hasleton, K M Kerr, N B N Ibrahim, S Stewart, W A H Wallace, A U Wells

Thorax 2004;59:500–505. doi: 10.1136/thx.2003.011734

See end of article for authors' affiliations

Correspondence to:
Dr A G Nicholson,
Department of
Histopathology, Royal
Brompton Hospital,
Sydney Street, London
SW3 6NP, UK;
a.nicholson@rbh.nthames.
nhs.uk

Received 20 June 2003
Accepted
11 November 2003

Background: There have been few inter-observer studies of diffuse parenchymal lung disease (DPLD), but the recent ATS/ERS consensus classification provides a basis for such a study.

Methods: A method for categorising numerically the percentage likelihood of these differential diagnoses was developed, and the diagnostic confidence of pathologists using this classification and the reproducibility of their diagnoses were assessed.

Results: The overall kappa coefficient of agreement for the first choice diagnosis was 0.38 (n = 133 biopsies), increasing to 0.43 for patients (n = 83) with multiple biopsies. Weighted kappa coefficients of agreement, quantifying the level of probability of individual diagnoses, were moderate to good (mean 0.58, range 0.40–0.75). However, in 18% of biopsy specimens the diagnosis was given with low confidence. Over 50% of inter-observer variation related to the diagnosis of non-specific interstitial pneumonia and, in particular, its distinction from usual interstitial pneumonia.

Conclusion: These results show that the ATS/ERS classification can be applied reproducibly by pathologists who evaluate DPLD routinely, and support the practice of taking multiple biopsy specimens.

Inter-observer variation studies in pulmonary pathology are generally concerned with neoplasia,^{1–7} and few have been applied to the reproducibility of non-neoplastic disease.^{8–10}

A consensus classification has recently been published by a joint American Thoracic Society/European Respiratory Society (ATS/ERS) sponsored committee comprising clinicians, radiologists and pathologists,¹¹ and there is evidence that recognition of these patterns provides significant prognostic data.^{8–12–17} However, unlike neoplasia where the pathology is typically definitive, the diagnosis of interstitial lung diseases is seldom unequivocal and diagnostic confidence levels vary. Furthermore, this variation is difficult to capture numerically for analytical purposes for two reasons. Firstly, the wealth of published terminology has historically been interpreted in different ways by clinicians and pathologists and, secondly, studies have concentrated on observer variation between first choice diagnoses and have not evaluated variation in the designation of diagnostic probabilities.

The ATS/ERS consensus classification therefore provides an opportunity to develop a numerical system for assessing pathological diagnoses of diffuse parenchymal lung disease (DPLD), and to use this system to assess the reproducibility of the diagnosis for such cases.

METHODS

Cases referred to the Royal Brompton Hospital for the clinical assessment of DPLD between January 1996 and December 1997, in which a surgical lung biopsy specimen was taken, were retrieved and haematoxylin and eosin stained sections were circulated to 10 members of the UK Interstitial Lung Disease panel. These reviewers were provided only with knowledge of the age and sex of the patient and the site of biopsy and were asked to categorise the histopathological pattern as follows: usual interstitial pneumonia (UIP), desquamative interstitial pneumonia (DIP), respiratory bronchiolitis, diffuse alveolar damage (DAD), organising pneumonia (OP), cellular and fibrotic non-specific interstitial pneumonia (NSIP), lymphoid interstitial pneumonia (LIP),

follicular bronchiolitis (FB), extrinsic allergic alveolitis (EAA), sarcoidosis, end stage lung disease, normal, non-diagnostic, unclassifiable, and "other" (box 1). The observers were instructed to assume that infection had been excluded and to ignore pleural disease. They were asked to specify the differential diagnosis for each lobe, and for each patient when more than one biopsy had been performed. For each stated diagnosis the percentage likelihood was specified to the nearest 5% (censored at 5% and summing to 100% in each lobe and for each patient). A first choice diagnosis was specified (even when the two most likely diagnoses were assigned the same percentage likelihood). If more than one pattern was recognised (for example, follicular bronchiolitis superimposed upon fibrotic NSIP), the differential diagnosis

Box 1 Permitted histopathological patterns

- 1 Usual interstitial pneumonia (UIP)
- 2a Non-specific interstitial pneumonia (NSIP), cellular
- 2b Non-specific interstitial pneumonia, fibrotic
- 3 Desquamative interstitial pneumonia (DIP)
- 4 Respiratory bronchiolitis (RB)
- 5 Diffuse alveolar damage (DAD)
- 6 Organising pneumonia (OP)
- 7 Lymphocytic interstitial pneumonia (LIP)
- 8 Follicular bronchiolitis (FB)
- 9 Extrinsic allergic alveolitis (EAA)
- 10 Sarcoidosis (Sarc)
- 11 Non-diagnostic (Non-diag)
- 12 Within normal limits (Normal)
- 13 Other: specify disease (e.g. giant cell interstitial pneumonia, eosinophilic pneumonia, amyloidosis, Langerhans' cell granulomatosis, lymphangioleiomyomatosis, bronchopneumonia, bronchiolitis)
- 14 End stage lung (ESL)
- 15 Unclassifiable (Unclass)

	Site	Age	Sex	First choice	Confidence	Differential diagnosis/es for each lobe	Differential diagnosis/es for each patient (% likelihood)
					1 = pathogn. (>95%) 2 = probable (70-95%) 3 = possible (>70%)	(% likelihood)	
1	NK			2B	2	2B(85) 1(15)	2B(80) 1(10) 11(10)
	NK			2B	3	2B(50) 1(10) 11(40)	
2	RLL			1	1	1(100)	1(100)
	RUL			2B	3	2B(60) 1(40)	
3	RLL			4	3	4(50) 11(50)	4(50) 11(50)
	RUL			4	3	4(50) 11(50)	
4	NK			1	1	1(100)	1(100)
5	LLL			9	2	9(70) 2A(20) 6(10)	9(80) 2A(20)
	LUL			2A	3	2A(50) 9(50)	
6	RLL			9	3	9(60) Bronchiolitis (40)	9(80) 1(20)
	RUL			1	3	1(50) 9(50)	
7	RML			9	3	9(60) 2B(30) 3(10)	9(50) 2B(35) 1(15)
	RLL			2B	2	2B(70) 1(30)	
8	LLL			9	1	9(95) 10(5)	9(95) 10(5)
	LUL			9	1	9(95) 10(5)	
9	NK			2B	2	2B(80) 1(20)	2B(80) 1(20)
10	LL			2B	2	2B(70) 1(30)	2B(70) 1(30)
11	LUL			2B	3	2B(50) 1(50)	2B(55) 1(45)
	LLL			2B	3	2B(60) 1(40)	
12	RML			2B	2	2B(70) 1(20) 3(10)	2B(70) 1(30)
	RLL			2B	2	2B(70) 1(30)	
13	LL			13	1	Amyloid(100)	Amyloid(100)
14	RML			5	1	5(90) 2A(10)	5(90) 2A(10)
	RLL			5	1	5(90) 2A(10)	
15	RUL			13	1	Lymphangioleiomyomatosis(100)	Lymphangioleiomyomatosis(100)
	RML			13	1	Lymphangioleiomyomatosis(100)	

RML = right middle lobe, RLL = right lower lobe, RUL = right upper lobe, LUL = upper lobe, LLL = left lower lobe, NK = not known, LL = left lung, UIP = usual interstitial pneumonia, NSIP = non-specific interstitial pneumonia, DIP = desquamative interstitial pneumonia, RB = respiratory bronchiolitis, DAD = diffuse alveolar damage, OP = organising pneumonia, LIP = lymphocytic interstitial pneumonia, FB = follicular bronchiolitis, EAA = extrinsic allergic alveolitis, SARC = sarcoidosis, Non-Dx = non-diagnostic, Norm = normal, OLD = orphan lung disease, ESL = end-stage lung, UNCL = unclassifiable.

Figure 1 An example of scoring for one reviewer using the numbered system from box 1.

was to be weighted according to the predominant pattern (for example, follicular bronchiolitis (70%); NSIP (30%)) An example of scoring from one reviewer is shown in fig 1.

Data analysis

The first choice diagnosis was assigned a confidence rating of 1 (>95% = pathognomonic), 2 (diagnostic likelihood 70–95% = high confidence), or 3 (diagnostic likelihood <70% = low confidence) for each lobe (n = 133) and for each patient (n = 83). For the purposes of data analysis, cellular and fibrotic NSIP were combined. Confidence ratings for each biopsy specimen were summed for the 10 observers; overall diagnostic confidence for each biopsy specimen was sub-categorised as high (summed confidence rating ≥ 13 , n = 74) or low (summed confidence rating <13, n = 59).

Inter-observer variation was quantified using the unadjusted kappa coefficient of agreement (κ), with the generation of kappa values for (a) the first choice lobar diagnosis and (b) the first choice diagnosis for each patient. In addition, observer variation for the first choice diagnosis was computed for each individual condition.

Inter-observer variation in the estimation of the probability of each condition was then quantified. For each biopsy specimen the percentage likelihood of each individual diagnosis was converted into a 5 point semi-quantitative scale according to clinically useful diagnostic probabilities: grade 1 = pathognomonic (100%); grade 2 = high probability (70–95%); grade 3 = intermediate probability (30–65%);

grade 4 = low probability (5–25%); grade 5 = condition not included in the differential diagnosis. A weighted kappa coefficient of agreement using quadratic weighting (kw)¹⁸ was then generated for each condition for the whole cohort; weighted kappa computation must be performed between paired observers and, thus, weighted kappa values are expressed as median values with ranges for the 45 possible combinations of 10 observers—that is, observer 1 versus observer 2; observer 1 versus observer 3 etc.

Observer agreement was categorised as poor, fair, moderate, good, or excellent according to kappa values of <0.20, 0.20–0.39, 0.40–0.59, 0.60–0.79, and >0.80, respectively.¹⁹ Statistical analyses were performed using STATA data analysis software (Computing Resource Center, Santa Monica, CA, USA).

Data were then further interrogated to identify the cardinal sources of inter-observer variation. Biopsy specimens in which divergent diagnoses were made by a minimum of two observers were tabulated—for example, UIP, n = 7; NSIP, n = 2; EAA, n = 1 was categorised as diagnostic discordance between UIP and NSIP. This analysis was then repeated with the requirement that divergent diagnoses were made by a minimum of three observers.

RESULTS

As shown in fig 2, the median prevalence of biopsy specimens considered to be pathognomonic by the 10 pathologists was 39% (range 29–57%). The median prevalence of first choice

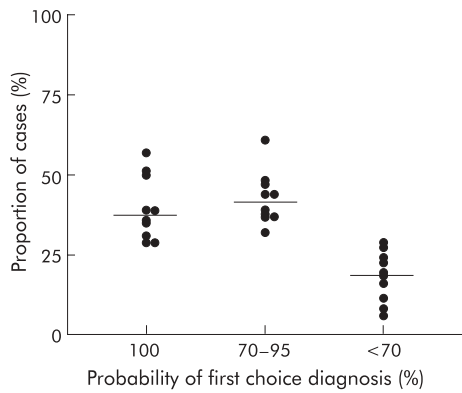


Figure 2 Diagnostic confidence of first choice diagnoses. The median prevalence of biopsy specimens considered to be pathognomonic was 39% (range 29–57%), the median prevalence thought to be pathognomonic or made with high confidence (>70% likelihood) was 82% (range 73–94%), and the median prevalence for diagnoses with low confidence (<70% likelihood) was 18% (range 6–27%).

diagnoses thought to be pathognomonic or made with high confidence (>70% likelihood) was 82% (range 73–94%). Pathologists made first choice diagnoses with low confidence (<70% likelihood) in a median of 18% of cases (range 6–27%).

Kappa coefficients for first choice diagnoses are shown in table 1. The overall kappa coefficient for agreement of diagnosis was only fair ($\kappa = 0.38$), with higher kappa values for patterns with relatively specific histological parameters such as UIP ($\kappa = 0.42$), OP ($\kappa = 0.57$) and sarcoidosis ($\kappa = 0.76$), and lower kappa values for patterns in which multiple features are required such as EAA ($\kappa = 0.36$) or in which the diagnosis is made, in part, by exclusion such as NSIP ($\kappa = 0.29$). Observer agreement was higher in biopsy specimens for which diagnostic confidence (as judged by summed confidence ratings) was high ($\kappa = 0.50$) than in those for which diagnostic confidence was low ($\kappa = 0.22$).

Biopsy specimens were taken from multiple sites in 48 patients (98 biopsies: two sites, $n = 46$; three sites, $n = 2$). In view of the reduction in sample size, kappa computation for first choice diagnoses was confined to 10 diagnostic categories recognised in a minimum of five cases. As shown in table 2, the kappa coefficient for the first choice diagnosis

Table 2 Kappa coefficients (κ) between lobar and final diagnoses in patients undergoing multiple biopsies

Diagnosis	Lobar diagnosis (n = 98)	Final diagnosis (n = 48)
UIP	0.4	0.49
NSIP	0.32	0.32
DIP	0.67	0.71
OP	0.59	0.67
EAA	0.39	0.35
Sarcoidosis	0.76	0.82
Normal	0.07	N/A
Other	0.36	0.49
ESL	0.28	N/A
Unclass	0.04	N/A
Overall	0.39	0.43

For definitions of abbreviations see box 1.

was marginally higher for patient diagnoses ($\kappa = 0.43$) than for lobar diagnoses ($\kappa = 0.39$).

Weighted kappa coefficients were also limited to five categories where the prevalence of the histological patterns was considered sufficient for diagnosis (table 3). These show that, for lobar diagnoses, there was a mean improvement in the kappa coefficient from 0.48 to 0.58. Furthermore, when the final diagnosis was compared with the lobar diagnosis, taking into account the combined assessment of multiple biopsies in 48 patients, all weighted kappa coefficients were either satisfactory or good (range 0.40–0.76, mean 0.58).

The sources of inter-observer variation are shown in table 4 (divergent diagnoses made by two or more pathologists) and table 5 (divergent diagnoses made by three or more pathologists). Disagreement most often centred upon a diagnosis of NSIP (made in over 50% of cases with divergent diagnoses), with the discrimination between NSIP and UIP a particular difficulty. The distinctions between end stage lung disease and both UIP and NSIP were also important sources of disagreement between observers.

Inter-observer variation within the category “other” is shown in table 6. Diagnoses here included Langerhans’ cell granulomatosis, eosinophilic pneumonia, lymphangiomatosis, pulmonary haemosiderosis, amyloid and lymphangiomyomatosis. There were also diagnoses more appropriate to bronchiolar than parenchymal disease, although terminology differed between individuals.

Table 1 Kappa coefficients of agreement of diagnostic categories of diffuse parenchymal lung disease

Disease category	Lobar specimens (n = 133)	Lobar specimens: high confidence (n = 74)	Lobar specimens: low confidence (n = 59)
UIP	0.42	0.58	0.19
NSIP	0.29	0.31	0.21
DIP	0.56	0.58	0.51
RB	0.09	0.08	0.1
DAD	0.5	0.69	0.11
OP	0.57	0.74	0.36
LIP	0.09	0.1	0.07
FB	0.33	0.39	0.22
EAA	0.36	0.5	0.15
Sarcoidosis	0.76	0.86	0.55
Non-diag	0.21	0.27	0.12
Normal	0.08	0.09	0.00
Other	0.41	0.60	0.13
ESL	0.31	0.35	0.19
Unclass	0.04	0.04	0.03
Overall	0.38	0.50	0.22

For definitions of abbreviations see box 1.

Overall coefficient levels are compared with those where diagnostic confidence was considered high ($\geq 70\%$) or low (<70% confidence).

Table 3 Comparison of kappa and weighted kappa coefficients of agreement

Diagnosis	Kappa (lobe)	Weighted kappa (lobe)	Weighted kappa (final)
UIP	0.42	0.53 (0.41, 0.60)	0.59 (0.24, 0.65)
NSIP	0.29	0.35 (0.26, 0.41)	0.40 (0.14, 0.45)
OP	0.57	0.68 (0.54, 0.75)	0.70 (0.20, 0.75)
EAA	0.36	0.46 (0.35, 0.51)	0.47 (0.23, 0.50)
Sarcoidosis	0.76	0.86 (0.78, 0.91)	0.75 (0.50, 0.80)
Mean	0.48	0.58	0.58

Median values are shown with 25th and 75th percentiles in parentheses.
For definitions of abbreviations see box 1.

DISCUSSION

This study was undertaken to quantify inter-observer variation in DPLD, an area of considerable diagnostic complexity where pathologists often offer a differential diagnosis. The devised method enabled us to see how pathologists come to a particular diagnosis by analysing their differential diagnoses. Our data show, for the first time, the diagnostic confidence and degree of observer variation among a large group of pulmonary pathologists dealing with DPLD. They also allow us to identify areas in which diagnostic problems are most frequently encountered.

The first analysis considered the degree of diagnostic confidence and showed that a 100% (pathognomonic) confidence level for a single diagnosis was only reached in 39% of cases, with 18% of first choice diagnoses made with low confidence. This is probably because recognition of these patterns requires the presence of several independent histological parameters in the sample—for example, fibroblastic foci, patchy subpleural or paraseptal fibrosis, and honeycombing in UIP. The presence of fibroblastic foci is particularly important in distinguishing UIP and NSIP, which may explain the frequent discordance in patients who have multiple biopsies.¹⁶

With regard to first choice lobar diagnoses, the overall kappa value was 0.38. This lies slightly below the “clinically acceptable” level of 0.40, but rose to 0.43 when diagnoses were made from multiple biopsies. We therefore also quantified inter-observer variation in the estimation of diagnostic probability. In order to make these analyses clinically meaningful, percentage likelihood estimations were converted to broad categories of diagnostic probability, applicable to clinical practice. This approach allowed us to address an important deficiency in statements of inter-observer variation based upon first choice diagnoses. When the diagnosis is a close call, two pathologists often concur in

their clinically meaningful conclusions—that is, they agree that two specified diagnoses are similarly likely—but may be absolutely in disagreement on the first choice diagnosis. Our data show a reduction in inter-observer variation when diagnostic probabilities are evaluated; when diagnoses were based upon multiple biopsies, weighted kappa values indicated a satisfactory level of agreement for all evaluated patterns. This is particularly significant in relation to NSIP, in which a single biopsy specimen may provide a low degree of confidence whereas a second specimen showing similar features would probably increase the diagnostic confidence and lower the inter-observer variability. Our data therefore indicate that, in the hands of experienced pulmonary pathologists, the ATS/ERS consensus classification is sufficiently reproducible for routine diagnostic usage.

We also analysed the sources of inter-observer variation by identifying the frequency with which particular differential diagnoses were offered (tables 5 and 6). Our data indicate that over half the observer noise related to a single pattern—namely, NSIP. This was true even without dividing NSIP into its cellular and fibrotic subtypes. Indeed, there were problems distinguishing NSIP from every other category apart from normal lung and sarcoidosis, but distinction from UIP was particularly problematic, these two patterns being frequently offered as alternative diagnoses. This is possibly because some cases of fibrotic NSIP represent early or inactive idiopathic pulmonary fibrosis so that fibroblastic foci are absent or scarce.^{16–20} Table 5 also highlights the problems when a pathologist is confronted with a biopsy specimen showing end stage lung disease (category 14) and attempts to tease a diagnosis of either fibrotic NSIP or UIP from such a sample. Previous studies have shown that coefficients of agreement fall to unacceptable levels in relation to end stage lung disease,⁸ and our analysis shows that pathologists are often unable to make a firm diagnosis because of inappropriate

Table 4 Noise analysis in which a minimum of two observers gave the same patterns as divergent diagnoses for a biopsy specimen

	UIP	NSIP	DIP	RB	DAD	OP	LIP	FB	EAA	Sarc	Non-diag	Normal	OLD	ESL	Unclass
UIP		29	1		3	7		1	4		3			14	1
NSIP			5	1	4	2	1	2	6		8		4	12	5
DIP				1							2			1	
RB												2			
DAD						4			1				2		
OP									1		1			1	
LIP								2						2	
FB									2	1	4			3	
EAA										3				1	
Sarcoidosis														1	
Non-diag												2	1		2
Normal															
OLD															4
ESL															
Unclass															

The number of occasions on which a differential diagnosis occurred is shown.
OLD=orphan lung disease. For definitions of other abbreviations see box 1.

Table 5 Noise analysis in which a minimum of three observers gave the same patterns as divergent diagnoses for a biopsy specimen

	UIP	NSIP	DIP	RB	DAD	OP	LIP	FB	EAA	Sarc	Non-diag	Normal	OLD	ESL	Unclass
UIP		14	1			1			1		1				7
NSIP			3					2	3		3				3
DIP											2			1	
RB												1			
DAD						1					1				
OP															
LIP								1							
FB										1			2		
EAA										1					
Sarcoidosis															
Non-diag												1			
Normal															
OLD															
ESL															
Unclass															

The number of occasions on which a differential diagnosis occurred is shown. OLD=orphan lung disease. For definitions of other abbreviations see box 1.

sampling or because the biopsy specimen shows only non-specific advanced disease.

Table 5 also illustrates problems that arise with particular histological patterns. For example, in EAA one might expect the most common differential diagnosis to be sarcoidosis, given that these are both granulomatous in nature. However, sarcoidosis proved to be a less common differential diagnosis than UIP and NSIP, suggesting that the fibrotic component of chronic EAA (versus UIP and fibrotic NSIP) and the relative bronchiolocentricity of the inflammation (subacute EAA versus cellular NSIP) provide greater problems for pathologists than the relative frequency of granulomas. Highlighting these areas therefore indicates areas that require further study and refinement of diagnostic criteria. To this end, a workshop specifically focused on NSIP is currently being sponsored by the ATS.

Our last analysis was of the "other" category which was primarily included to allow assessment of rare diffuse parenchymal lung diseases. These included individual cases of Langerhans' cell granulomatosis, eosinophilic pneumonia, primary pulmonary hypertension, lymphangioleiomyomatosis, amyloidosis, lymphangiomatosis/lymphangectasia, and pulmonary haemosiderosis, all of which were diagnosed by more than one observer. Both lobar and overall kappa coefficients were satisfactory for this group, but levels of 0.41 and 0.49 are perhaps lower than expected for disorders with quite distinct histological features. However, sub-

analysis shows that most observer disagreement centred on pathologists classifying six out of the 13 such biopsy specimens as showing some form of bronchiolocentric lung injury. This is of interest in view of a recent report describing idiopathic bronchiolocentric pneumonia.²¹ It also suggests an overlap between small airways disease and diffuse parenchymal disease that needs clarification.

One potential limitation to this study is the fact that the participating pathologists all have a special interest in pulmonary pathology. General histopathologists may be less confident in dealing with DPLD. However, most surgical lung biopsies in the UK are reported by pathologists with a designated interest in lung disease, so we believe that this study reflects current practice and is likely to be internationally applicable.

Historically, the use of surgical biopsy specimens has been viewed as the diagnostic standard in diffuse lung disease. Our data call into question this premise since, in two-thirds of cases, the histological appearances were not pathognomonic, with major uncertainty in 18% of cases. Furthermore, we have quantified major observer variation among pathologists even though the overall level of agreement was clinically acceptable. These findings highlight the importance of integrating histopathological data with HRCT scanning and clinical findings, as recently recommended by expert groups.¹¹ Indeed, the reconciliation of histopathological and HRCT data has recently been shown to refine the

Table 6 Cases in which a diagnosis of "other" was made by more than one observer (n = 10)

Diagnosis (no of times made)	No of patients	No of biopsies	Other diagnoses
LCG (4)	1	1	DIP, DAD (2), Non-diag, NSIP
EPN (5)	1	1	DAD (2), OP (2), DIP (1)
LS/E (6)	1	1	UNCL (2), FB (1), RB (1)
IPH (6)	1	1	NSIP (2), DIP (1), FB (1)
Amyloid (10)	1	1	-
LAM (17)	1	2	EMPH (2), Normal (1)
CB (6)	1	3	UNCL (9), NSIP (11), Non-diag (2), EAA (2)
BO (1), CB (1), PPH (1)	1	1	FB (7)
BPN (3) EPN (2) BO (1)	1	1	NSIP (2), UIP (1), EAA (2), UNCL (1)
BO (1) LCG (1)	1	1	OP (5), UIP (1), NSIP (2)

LCG = Langerhans' cell granulomatosis; DIP = desquamative interstitial pneumonia; DAD = diffuse alveolar damage, non-diag = non-diagnostic; NSIP = non-specific interstitial pneumonia; EPN = eosinophilic pneumonia; OP = organising pneumonia; LS/E = lymphangiomatosis/lymphangectasia; UNCL = unclassifiable; FB = follicular bronchiolitis; RB = respiratory bronchiolitis, Pulm haemosid = pulmonary haemosiderosis; EAA = extrinsic allergic alveolitis; UIP = usual interstitial pneumonia; LAM = lymphangioleiomyomatosis; EMPH = emphysema; CB = chronic bronchiolitis; BO = bronchiolitis obliterans; PPH = primary pulmonary hypertension; BPN = bronchopneumonia. Figures in parentheses indicate the number of occasions that an individual diagnosis was made (maximum of 10).

prognostic evaluation in fibrotic idiopathic interstitial pneumonias.²²

In conclusion, this study shows that the ATS/ERS consensus classification is sufficiently reproducible when used by pathologists who routinely evaluate lung biopsy specimens in diffuse lung disease. It also shows that much of the disagreement among pathologists centres on the category of NSIP. Bronchiolocentric fibrosis is another area worthy of further study.

Authors' affiliations

A G Nicholson, B Corrin, Department of Histopathology, Royal Brompton Hospital, London, UK
B J Addis, Department of Histopathology, Southampton General Hospital, Southampton, UK
H Bharucha, Department of Histopathology, Royal Victoria Hospital, Belfast, UK
C A Clelland, Department of Histopathology, John Radcliffe Infirmary, Oxford, UK
A R Gibbs, Department of Histopathology, Llandough Hospital, Penarth, South Glamorgan, UK
P S Hasleton, Department of Histopathology, Wythenshawe Hospital, Manchester, UK
K M Kerr, Department of Histopathology, Aberdeen Royal Infirmary, Aberdeen, UK
N B N Ibrahim, Department of Histopathology, Frenchay Hospital, Bristol, UK
S Stewart, Department of Histopathology, Papworth Hospital, Cambridge, UK
W A H Wallace, Department of Histopathology, Royal Infirmary of Edinburgh, Edinburgh, UK
A U Wells, Department of Medicine, Royal Brompton Hospital, London, UK

REFERENCES

- Nicholson AG, Perry LJ, Cury PM, et al. Reproducibility of the WHO/IASLC grading system for pre-invasive squamous lesions of the bronchus: a study of inter-observer and intra-observer variation. *Histopathology* 2001;**38**:202–8.
- Kung IT, Lam WK, Lam TH. Observer variability studies of the WHO classification of lung cancer. *Cancer Treat Res* 1989;**45**:53–69.
- Campobasso O, Andron A, Ribotta M, et al. The value of the 1981 WHO histological classification in inter-observer reproducibility and changing pattern of lung cancer. *Int J Cancer* 1993;**53**:205–8.
- Skov BG, Lauritzen AF, Hirsch F, et al. The histopathological diagnosis of malignant mesothelioma v. pulmonary adenocarcinoma: reproducibility of the histopathological diagnosis *Histopathology* 1994;**24**:553–7.
- Skov BG, Braendstrup O, Hirsch FR, et al. Are pathologists biased by clinical information? A blinded cross-over study of the histopathological diagnosis of mesothelial tumours versus pulmonary adenocarcinoma. *Lung Cancer* 1994;**11**:365–72.
- Raab SS, Thomas PA, Lenel JC, et al. Pathology and probability. Likelihood ratios and receiver operating characteristic curves in the interpretation of bronchial brush specimens. *Am J Clin Pathol* 1995;**103**:588–93.
- Colby TV, Tazelaar HD, Travis WD, et al. Pathologic review of the Mayo Lung Project. Is there a case for misdiagnosis or overdiagnosis of lung carcinoma in the screened group? *Cancer* 2002;**95**:2361–5.
- 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**:2213–7.
- Bouros D, Nicholson AG, Colby TV, et al. Histopathological subsets of fibrosing alveolitis in patients with scleroderma and their relationship to natural history and treated course. *Am J Respir Crit Care Med* 2002;**165**:1581–6.
- Nicholson AG, Fulford LG, Colby TV, et al. The frequency of fibroblastic foci in usual interstitial pneumonia and their relationship to disease progression. *Am J Respir Crit Care Med* 2002;**166**:173–7.
- Travis WD, King TE, Bateman ED, et al. ATS/ERS consensus statement of idiopathic interstitial pneumonias. *Am J Respir Crit Care Med* 2002;**265**:277–304.
- Bjoraker JA, Ryu JH, Edwin MK, et al. Prognostic significance of histopathological subsets in idiopathic pulmonary fibrosis. *Am J Respir Crit Care Med* 1998;**157**:199–203.
- Daniil ZD, Gilchrist FC, Nicholson AG, et al. A histologic pattern of nonspecific interstitial pneumonia is associated with a better prognosis than usual interstitial pneumonia in patients with cryptogenic fibrosing alveolitis. *Am J Respir Crit Care Med* 1999;**160**:899–905.
- Nagai S, Kitaichi M, Itoh H, et al. Idiopathic nonspecific interstitial pneumonia/fibrosis: comparison with idiopathic pulmonary fibrosis and BOOP. *Eur Respir J* 1998;**12**:1010–9.
- Travis WD, Matsui K, Moss J, et al. Idiopathic nonspecific interstitial pneumonia: prognostic significance of cellular and fibrosing patterns. Survival comparison with usual interstitial pneumonia and desquamative interstitial pneumonia. *Am J Surg Pathol* 2000;**24**:19–33.
- Flaherty KR, Travis WD, Colby TV, et al. Histopathologic variability in usual and nonspecific interstitial pneumonias. *Am J Respir Crit Care Med* 2001;**164**:1722–7.
- Flaherty KR, Toews GB, Travis WD, et al. Clinical significance of histological classification of idiopathic interstitial pneumonia. *Eur Respir J* 2002;**19**:275–83.
- Streiner DL, Norman GR. *Health measurement scales: a practical guide to their development and use*. Oxford: Oxford University Press, 2000:94–5.
- 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**:236–40.
- Nicholson AG, Wells AU. Nonspecific interstitial pneumonia—nobody said it's perfect. *Am J Respir Crit Care Med* 2001;**164**:1553–4.
- Yousem SA, Dacic S. Idiopathic bronchiolocentric interstitial pneumonia. *Mod Pathol* 2002;**15**:1148–53.
- Flaherty KR, Thwaite EL, Kazerooni EA, et al. Radiological versus histological diagnosis in UIP and NSIP: survival implications. *Thorax* 2003;**58**:143–8.