

Interstitial lung disease: time to rethink the snapshot diagnosis?

Danielle Antin-Ozerkis,¹ Martin Kolb²

An accurate and early diagnosis of idiopathic pulmonary fibrosis (IPF) is critically required for patients and care providers because it dictates very specific management decisions that include referral to transplant, access to new approved drugs, avoidance of immunosuppression and potential referral to palliative care.¹ While the original diagnosis of IPF was highly dependent on patterns observed on histology, in 2002 the American Thoracic Society (ATS)/European Respiratory Society (ERS) guideline altered the approach to diagnosis so as to include a clinical–radiological–pathological multidisciplinary diagnosis.² A careful history, searching for subtle evidence of connective tissue disease, exposures and other known causes for interstitial lung disease (ILD), was emphasised. A radiographic pattern of usual interstitial pneumonia (UIP) on high-resolution CT (HRCT) was described, characterised by basal-predominant fibrosis with peripheral reticular markings, traction airway change, architectural distortion and honeycombing.

With this new classification system, it was proposed, characteristic HRCT findings could lead to a confident diagnosis of IPF without the need for a biopsy. The advantage of using HRCT patterns as a surrogate for pathological findings was particularly appealing given data describing increased risk for acute exacerbation and death following lung biopsy.³ Support for such an approach was increased by studies demonstrating agreement between radiographic and pathological findings. Raghu *et al*⁴ found that the specificity of HRCT for UIP was 90%. Flaherty *et al*⁵ found that the HRCT interpretation of definite UIP in biopsy-proven UIP and non-specific interstitial pneumonia (NSIP) had 100% specificity. Hunninghake *et al*⁶ found that in a blinded prospective evaluation of patients with surgical lung

biopsies, a confident HRCT diagnosis of UIP was 95% specific for the pathological finding of UIP. HRCT appearance was also predictive of survival: a definite UIP pattern on HRCT (having all the features described above) was associated with a lower survival than for those patients with radiographic findings that were indeterminate or suggestive of another diagnosis.⁵ Despite this apparently high predictive power, however, these same studies reported low agreement regarding a specific diagnosis (kappa of 0.54) on HRCT and other authors have also found only modest agreement (kappa of 0.48) for the finding of honeycombing, suggesting that even this key finding is not easily interpreted.^{6–8} Community physicians are more likely to assign a diagnosis of IPF and to be in agreement on it than physicians at tertiary referral centres.⁸

Subsequent iterations of the ATS/ERS guidelines have attempted to refine the definition of IPF, and, in particular, the radiographic features of UIP. In 2011, ATS consensus statement specifically defined three categories of radiographic criteria: UIP pattern, possible UIP pattern and inconsistent with UIP pattern.¹ The final diagnosis of IPF rested on a combination of radiographic and pathological features (when lung biopsy was obtained), which, in combination with multidisciplinary discussion, would lead to one of several possible diagnostic statements: IPF, probable IPF, possible IPF or not IPF. This classification scheme was designed to allow for the uncertainty that is often present in pathological and radiographic interpretations and to permit less-than-confident definitions of IPF. However, these categories were never validated, either retrospectively or prospectively, nor were studies performed to assess the reproducibility of these diagnostic assessments in actual practice.

In *Thorax*, Walsh *et al*⁹ address this latter issue in an elegant manner. The authors designed a two-part study. CTs from consecutive patients in a tertiary referral centre were obtained. All patients carried a multidisciplinary team diagnosis of idiopathic fibrotic lung disease, chronic hypersensitivity pneumonitis (HP) or fibrotic lung disease associated with a connective tissue disease, excluding

sarcoidosis. All CTs were performed at full inspiration as high-resolution scans. Observers were invited to participate from a variety of radiographic societies. In the first stage of the study, an internet-based viewing application allowed radiologists to view 15 sections from each CT and to assign a diagnostic category: UIP, possible UIP or inconsistent with UIP. Readers were asked to score honeycombing, traction bronchiectasis and emphysema as definitely present, possibly present or absent. In the second stage of the study, chest radiologists were randomly selected from the initial participant group, one subset with >20 years of experience and another with <10 years. These thoracic radiologists were given the full thin-section CTs of a new cohort of patients and scored CTs with the same criteria. Kappa values were calculated for interobserver agreement for diagnostic category and for honeycombing. Surprisingly, agreement on overall diagnosis category was only moderate (ranging from 0.48 for general radiologists to 0.52 for chest radiologists). The agreement between radiologists did not even improve if the approach was simplified into a binary comparison of definite/possible UIP versus inconsistent with UIP. When interpreting the presence of honeycombing by use of a scoring system, the agreement scores ranged from 0.56 to 0.65, though the higher agreement scores were found in less experienced observers (fellows), possibly suggesting that there was agreement but not necessarily accuracy.

This study demonstrates that the problem of inter-reader reliability, which has long been a concern in the study of IPF, has not been fully addressed by the recent ATS/ERS consensus revision. Reliably phenotyped subjects are key to interpreting data from clinical trials and translational studies. Similarly, clinical decision-making regarding the use of novel antifibrotic medications and immunosuppressive therapies increasingly hinges on CT interpretation as fewer patients are undergoing surgical lung biopsy.¹⁰

Several possible approaches could be considered to address these issues. The first is to further refine the current criteria by better identifying the reasons for uncertainty in CT interpretation. One of the limitations to this study is that we do not know why the radiologists called certain scans inconsistent. Are these discrepancies due to areas of lucency that are variably interpreted as emphysema or honeycombing? Would addition of expiratory images help improve overall results by highlighting air trapping or would they

¹Section of Pulmonary Critical Care and Sleep Medicine, Department of Medicine, Yale University, New Haven, Connecticut, USA; ²Division of Respiratory, Department of Medicine and Pathology/Molecular Medicine, McMaster University Hamilton, Ontario, Canada

Correspondence to Dr Danielle Antin-Ozerkis, Section of Pulmonary Critical Care and Sleep Medicine, Department of Medicine, Yale University, New Haven, CT 06520-8057, USA; danielle.antin-ozerkis@yale.edu

add to the complexity and unreliability of interpretation?

It is worth looking at the information given in some of the images in detail. In figure 1,⁹ HRCT cuts are shown from a patient with known rheumatoid arthritis. Among the observers, there was a wide range of interpretations (definite UIP 20%, possible UIP 36.5% and inconsistent with UIP 42.6%), and only 62.6% identified definite honeycombing. In contrast, figure 2⁹ represents images from a patient with UIP on biopsy. There was much less variability (definite UIP 73.9%, possible UIP 21.7% and inconsistent with UIP 4.3%) and fully 91.3% of observers graded definite honeycombing. It is possible that these data are pointing towards something we frequently recognise clinically: that CT scans from patients with known causes of lung disease, such as connective tissue disease, are the most difficult to categorise, but that when these clinical diagnoses are incorporated, overall interpretation improves. It is well known that patients with connective tissue-related ILD often have multiple histological patterns of disease. The CT may be mirroring this reality, making it difficult for even experienced radiologists to easily classify scans. Therefore, it may not be realistic or clinically important to hold these patients to the same standards of radiographic classification as diagnostic and therapeutic decisions do not hinge on this interpretation. In fact, this is the situation where multidisciplinary discussion plays the key role. CT interpretation should not be a stand-alone test for all patients with fibrotic lung disease but should primarily be used to assess idiopathic interstitial pneumonias. As the authors suggest, a valuable follow-up study would be to assess inter-rater reliability when only HRCTs from suspected idiopathic disease are included.

A simultaneous, practical approach to the issue of inter-rater reliability may revolve around improved reference standards and training. The ILD community and chest radiologists, in particular, would benefit from using an agreed-upon set of reference images that could be used as training exercises for radiologists at all levels of experience and in various specialties and clinical settings. Further work could assess whether intensive training on CT interpretation improves inter-reader agreement. It appears that training of radiologists in these criteria can lead to some uniformity of diagnosis, though whether this leads to sufficient levels of agreement is not clear.¹¹ If we start by becoming more precise in our CT interpretation, we

would then be better able to assess whether we are also accurate in our diagnoses.

Using a system of binary outcomes for patterns (definite/possible UIP vs inconsistent with UIP) could be useful, but may not be the best way to classify diseases (IPF vs non-IPF, IPF vs chronic HP). It is well known that histology in these patients exists along a continuum, with chronic HP merging into a UIP pattern in some patients. Establishing a clear radiological cut-off may, therefore, be overly simplistic and does not reflect the dynamics and nature of the underlying disease. If a continuous scoring system is used, there may be more general agreement in describing the images. In addition to improving reproducibility, this may also focus attention on the reality that more severe fibrotic lung disease tends to merge with each other in clinical outcome regardless of its pathogenesis. There is evidence that the extent of fibrosis in NSIP is predictive of poor prognosis, and that once HP has reached a fibrotic appearance similar to IPF, survival is likewise low.^{12 13} Future studies on this subject should consider including follow-up CT in the assessment as a significant number of scans that are initially uncertain may change over time and thereby also improve interobserver agreement.

The overarching question is not only whether we can but also whether we should create a system of high interobserver agreement. Walsh *et al* clearly show us that radiology cannot deliver it. Pathological interpretation of biopsies is subject to the same issues of inter-reader reliability. Thus, unlike many other disease areas, no true gold standard exists. Current best practice uses multidisciplinary discussions to incorporate all information, including clinical and functional data to come to a 'consensus diagnosis'. However, in real life 'consensus' is often the opinion of the strongest voice in the panel and not necessarily true agreement. We have to keep in mind that all fibrotic lung diseases, and IPF in particular, are dynamic and unpredictable. Any diagnosis is only the diagnosis made at a certain time and based on the information present at that time. Radiological patterns are undoubtedly one of the crucial pieces in the puzzle, but it would be unwise to believe that it would reflect the underlying biology of the disease, no matter how good inter-reader agreement is. Research must also focus on molecular biomarkers, which can help to identify patients at risk for developing disease, as well as to assist

in diagnosis and prognosis, identify targets for new drug therapeutics and predict response to therapy. Ultimately, we need to be aware that guidelines on diagnosis and treatment of such a complex group of disorders need to be flexible and allow for constant change. The recent classification of idiopathic interstitial pneumonias made an important step in the right direction as it includes clinical disease behaviour.¹⁴ Disease behaviour and peripheral blood biomarkers that correlate with biological subtypes in a meaningful way may even be more important than a firm CT diagnosis and allow for a 'dynamic' diagnosis of fibrotic lung disease rather than a 'snapshot' in time.

Acknowledgements Thanks to Robert Homer, MD PhD, for helpful discussions.

Competing interests None declared.

Provenance and peer review Commissioned; internally peer reviewed.



CrossMark

To cite Antin-Ozerkis D, Kolb M. *Thorax* 2016;**71**:5–7.



► <http://dx.doi.org/10.1136/thoraxjnl-2015-207252>

Thorax 2016;**71**:5–7.
doi:10.1136/thoraxjnl-2015-208148

REFERENCES

- 1 Raghu G, Collard HR, Egan JJ, *et al*. An official ATS/ERS/IRIS/ALAT statement: idiopathic pulmonary fibrosis: evidence-based guidelines for diagnosis and management. *Am J Respir Crit Care Med* 2011;**183**:788–824.
- 2 American Thoracic Society, European Respiratory Society. American Thoracic Society/European Respiratory Society International Multidisciplinary Consensus Classification of the Idiopathic Interstitial Pneumonias. This joint statement of the American Thoracic Society (ATS), and the European Respiratory Society (ERS) was adopted by the ATS board of directors, June 2001 and by the ERS Executive Committee, June 2001. *Am J Respir Crit Care Med* 2002;**165**:277–304.
- 3 Lettieri CJ, Veerappan GR, Helman DL, *et al*. Outcomes and safety of surgical lung biopsy for interstitial lung disease. *Chest* 2005;**127**:1600–5.
- 4 Raghu G, Mageto YN, Lockhart D, *et al*. The accuracy of the clinical diagnosis of new-onset idiopathic pulmonary fibrosis and other interstitial lung disease: a prospective study. *Chest* 1999;**116**:1168–74.
- 5 Flaherty KR, Thwaite EL, Kazerooni EA, *et al*. Radiological versus histological diagnosis in UIP and NSIP: survival implications. *Thorax* 2003;**58**:143–8.
- 6 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**:193–6.

- 7 Sundaram B, Gross BH, Martinez FJ, *et al.* Accuracy of high-resolution CT in the diagnosis of diffuse lung disease: effect of predominance and distribution of findings. *AJR Am J Roentgenol* 2008;191:1032–9.
- 8 Aziz ZA, Wells AU, Hansell DM, *et al.* HRCT diagnosis of diffuse parenchymal lung disease: inter-observer variation. *Thorax* 2004;59:506–11.
- 9 Walsh SL, Calandriello L, Sverzellati N, *et al.*, Interobserver agreement for the ATS/ERS/JRS/ALAT criteria for a UIP pattern on CT. *Thorax* 2016;71:45–51.
- 10 Chung JH, Chawla A, Peljto AL, *et al.* CT scan findings of probable usual interstitial pneumonitis have a high predictive value for histologic usual interstitial pneumonitis. *Chest* 2015;147:450–9.
- 11 de Andrade J, Schwarz M, Collard HR, *et al.* The Idiopathic Pulmonary Fibrosis Clinical Research Network (IPFnet): diagnostic and adjudication processes. *Chest* 2015;148:1034–42.
- 12 Park IN, Jegal Y, Kim DS, *et al.* Clinical course and lung function change of idiopathic nonspecific interstitial pneumonia. *Eur Respir J* 2009;33:68–76.
- 13 Churg A, Sin DD, Everett D, *et al.* Pathologic patterns and survival in chronic hypersensitivity pneumonitis. *Am J Surg Pathol* 2009;33:1765–70.
- 14 Travis WD, Costabel U, Hansell DM, *et al.* An official American Thoracic Society/European Respiratory Society statement: update of the international multidisciplinary classification of the idiopathic interstitial pneumonias. *Am J Respir Crit Care Med* 2013;188:733–48.