Unclassifiable-interstitial lung disease: outcome prediction using CT and functional indices

Joseph Jacob[1], Brian J. Bartholmai[2], Srinivasan Rajagopalan[3], Ryoko Egashira[4], Anne Laure Brun[1], Maria Kokosi[5], Arjun Nair[6], Simon L.F. Walsh[7], Ronald Karwoski[3], Andrew G. Nicholson[8], David M. Hansell[1], Athol U. Wells[5].

[1]Department of Radiology, Royal Brompton Hospital, Royal Brompton and Harefield NHS Foundation Trust, London, UK.

[2]Division of Radiology, Mayo Clinic Rochester, Rochester, Minnesota, USA.

[3]Department of Physiology and Biomedical Engineering, Mayo Clinic Rochester, Rochester, Minnesota, USA.

[4]Department of Radiology, Faculty of Medicine, Saga University, Saga city, Japan.

[5]Interstitial Lung Disease Unit, Royal Brompton Hospital, Royal Brompton and Harefield NHS Foundation Trust, London, UK.

[6]Department of Radiology, Guys and St Thomas' NHS Foundation Trust, London, UK.

[7]Department of Radiology, Kings College Hospital NHS Foundation Trust, London, UK.

[8]Department of Histopathology, Royal Brompton Hospital, Royal Brompton and Harefield NHS Foundation Trust, and National Heart and Lung Institute, Imperial College, London, UK.

Corresponding author: Joseph Jacob, 10 Wolsey road, Northwood, Middlesex, UK HA6 2HW.

Email: joseph.jacob@nhs.net

Authors contributions

JJ, MK, RE, ALB, AN, SLFW, AGN, AUW, DMH were involved in either the acquisition, or analysis or interpretation of data for the study.

JJ, AUW and DMH were also involved in the conception and design of the study.

BJB, RK and SR invented and developed CALIPER. They were involved in processing the raw CT scans and in generation of figures but were not involved with the analysis or interpretation of the data in the study.

All authors revised the work for important intellectual content and gave final approval for the version to be published. All authors agree to be accountable for the all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved.

Ethics committee approval

Approval for this study of clinically indicated CT and pulmonary function data was obtained from the Institutional Ethics Committee of the Royal Brompton Hospital and Mayo Clinic and informed patient consent was not required.

Keywords

Quantitative CT; unclassifiable interstitial lung disease; longitudinal analysis

ABBREVIATION LIST

Ao	aorta
BAL	broncho-alveolar lavage
CI	confidence interval
CPI	composite physiologic index
CT	computed tomography
DA	decreased attenuation
DLco	diffusing capacity for carbon monoxide
FEV1	forced expiratory volume in one second
FVC	forced vital capacity
GGO	ground glass opacity
HR	hazard ratio
ILD	interstitial lung disease
IPF	idiopathic pulmonary fibrosis
Kco	carbon monoxide transfer coefficient
MDT	multi-disciplinary team
PA	pulmonary artery
PFT	pulmonary function test
PVV	pulmonary vessel volume
RVSP	right ventricular systolic pressure
TLC	total lung capacity
TxBx	traction bronchiectasis
uILD	unclassifiable interstitial lung disease
UIP	usual interstitial pneumonia

ABSTRACT

Background: Unclassifiable-interstitial lung disease (uILD) represents a heterogeneous collection of pathologies encompassing those fibrosing lung diseases which do not fulfill current diagnostic criteria. We evaluated baseline and longitudinal functional and CT (visual and quantitative computer [CALIPER] analysis) variables to identify outcome predictors in uILD.

Methods: Consecutive patients with uILD on multidisciplinary review (n=95) had baseline functional (FVC, DLco, CPI [composite physiologic index]) and CT features (visual evaluation: CT pattern, fibrosis extent, honeycombing presence, traction bronchiectasis severity, pulmonary artery (PA) diameter; CALIPER evaluation: fibrosis extent, pulmonary vessel volume (PVV)) examined in univariate and multivariate Cox regression models. Change in functional and CT variables were examined in a patient subset (n=37), to identify indicators of outcome.

Results: On univariate analysis, CPI was the most powerful functional predictor of mortality ($p<0.0001$). Visual traction bronchiectasis ($p<0.0001$), PA diameter ($p<0.0001$) and honeycombing presence ($p=0.0001$) and CALIPER PVV ($p=0.0003$) were the strongest CT outcome predictors.

On multivariate analysis of baseline indices, traction bronchiectasis ($p=0.003$), PA diameter ($p=0.003$) and CPI ($p=0.0001$) independently predicted mortality. Colinearity with functional indices precluded the evaluation of CALIPER PVV in multivariate models.

On evaluation of longitudinal variables, increasing CALIPER fibrosis extent was the strongest outcome predictor, and remained so following adjustment for baseline disease severity, and when FVC declines were marginal.

Conclusions: In uILD patients, CPI, traction bronchiectasis severity and PA diameter independently predicted outcome at baseline. Increasing fibrosis extent measured by CALIPER was the most powerful index of outcome regardless of baseline disease severity and strongly predicted outcome in patients with marginal FVC declines.

INTRODUCTION

Multi-disciplinary teams (MDT) meetings are widely viewed as the preferred means of characterizing interstitial lung diseases[1, 2] and has improved agreement of the diagnosis of ILD between centres[3]. However, it is increasingly recognized that following MDT discussion, 10-25% of cases of interstitial lung disease cannot be classified and are labeled unclassifiable-ILD (uILD)[4-10].

Concerns with regard to the increased mortality associated with surgical lung biopsy has reduced the number of such procedures being performed, with the result that many patients are today evaluated using radiological and clinical data only in an MDT discussion[4]. Nevertheless, patients with uILD are ideally monitored in an MDT setting in case emerging clinical data allow a diagnostic label to be applied[6].

Our study aimed to characterize those baseline and longitudinal variables that best predict mortality in patients with uILD using a combination of clinical variables, functional indices and visual and computer-identified CT features.

METHODS

Clinical Data

The study population consisted of all consecutive new patients presenting over a four and a half year period (January 2007 to July 2011) who on re-examination in a multi-disciplinary team (MDT) setting were assigned a diagnosis of uILD. A diagnosis of uILD was reached if, following MDT discussion, a single diagnosis (in accordance with standard criteria where available[2, 11-13]) could not be reached with over 50% diagnostic certainty. Patients who had undergone a non-contrast, supine, volumetric thin section CT scans were selected and exclusions are shown in the flowchart in the Supplementary Appendix (Figure 1).

Standardized CT scanning protocols, CALIPER protocols and echocardiographic and pulmonary function test (PFT) protocols were utilized [14, 15]. Echocardiographic and PFT data was considered if performed within 3 months of the CT scan. Functional indices analysed in the current study included: forced expiratory volume in one second (FEV1), forced vital capacity (FVC), diffusing capacity for carbon monoxide (DLco), carbon monoxide transfer coefficient (Kco), total lung capacity (TLC), and the composite physiological index (CPI). The composite physiologic index is a score that considers varied weightings of FEV1, FVC and DLco (CPI = 91 - (0.65 x % predicted DLco) - (0.53 x % predicted FVC) + (0.34 x % predicted FEV1)) and was derived to account for the morphologic extent of pulmonary fibrosis on CT, correcting for the confounding effects of emphysema[16].

Approval for this study of clinically indicated CT and pulmonary function data was obtained from the Institutional Ethics Committee of the _____ and _____ and informed patient consent was not required.

Visual and CALIPER CT analysis

In accordance with a landmark study of uILD[4], classification of CT patterns into definite, possible and inconsistent UIP patterns according to accepted diagnostic criteria for idiopathic pulmonary fibrosis (IPF)[2] was performed by JJ (5 years thoracic imaging experience). Visual CT scoring was performed independently by two of four radiologists (AN and SLW or ALB and RE) with 5 to 9 years thoracic imaging experience, blinded to all clinical information. The same scorer pairs evaluated the same patient's serial CT (performed 6-30 months after the initial CT) when available, and the scorers were blinded to the time points of the serial CTs. Consenus formulation for visual CT scores is outlined in the Supplementary Appendix.

CT patterns scored visually as a percentage of a lobe included: ground glass opacity, reticular pattern, honeycombing and emphysema. The lobar scores for a single parenchymal pattern were summed and divided by 6 (the left upper lobe and lingula were considered as separate lobes) to generate a total lung percentage for each parenchymal pattern Traction bronchiectasis severity[17] was evaluated using a categorical four-point lobar scale (none=0, mild=1, moderate=2, severe=3) that reflected both the degree of airway dilatation within areas of fibrosis and the

number of lobar segments containing airways dilated secondary to fibrosis. Lobar scores were summed for six lobes resulting in a 19-point scale.

Indices reflecting pulmonary hypertension included the main pulmonary artery diameter and the main pulmonary artery:ascending aorta ratio. Both were assessed by one scorer using electronic caliper diameter measurements of the ascending aorta and pulmonary artery diameters at the level of the pulmonary artery bifurcation[18]. Details regarding scoring methodology and consensus formulation have been previously described.[14]

CALIPER represents a computer tool developed by the Biomedical Imaging Resource at the Mayo Clinic Rochester. CALIPER evaluates structural, textural and histogram characteristics of 15x15x15 voxel volume units across contiguous volumetric CT datasets, and classifies each voxel into one of eight patterns: normal lung, ground glass opacity, reticular pattern, honeycombing, three grades of low parenchymal attenuation and the pulmonary vessel volume (PVV). The PVV represents the total volume of pulmonary arteries and veins excluding vessels at the lung hilum[19]. Regarding the three grades of low parenchymal attenuation, Grade 2 and 3 low attenuation areas correspond to emphysema, whilst grade 1 low attenuation predominantly corresponds to normal lung[14]. The total lung volume of each of the eight CALIPER parenchymal patterns was divided by the total lung volume calculated by CALIPER to derive a total lung percentage for each parenchymal pattern.

For both visual and CALIPER scores, total fibrosis extent represented the sum of reticular pattern and honeycombing, whilst total ILD extent additionally summed ground glass opacities. For serial CT analyses, absolute change in CT patterns (Timepoint 2 minus Timepoint 1) expressed as a percentage of the lungs was calculated.

Calculation of change in CT and pulmonary function variables

Change in all CT variables and change in the composite physiologic index (CPI) were expressed as absolute change per year. Change in FVC and DLco was expressed as absolute and relative change per year. Kco and TLC change were only expressed as relative change per year.

Statistical analysis

Data are given as medians or means with standard deviations depending on distributions, or numbers of patients with percentages where appropriate. Interobserver variation for visual scores was assessed using the single determination standard deviation for continuous variables and the kappa statistic for categorical variables. Differences between groups were evaluated using a Chi-squared test for categorical variables or the T-test for normally distributed continuous variables and the Mann Whitney-U test for non-normally distributed variables. Statistical significance was evaluated at a value of $p<0.05$. Univariate and multivariate linear and logistic regression analyses were used to investigate relationships between variables. Univariate and multivariate Cox regression analyses were used to investigate variables predictive of mortality. Kaplan Meier survival curves were

compared using the Log rank test. Assumptions of linearity and proportional hazards were tested by visual inspection of Martingale residuals and scaled Schoenfeld residuals and were satisfied. Statistical analyses were performed with STATA (version 12, StatCorp, College Station, TX, USA).

RESULTS

Population demographics

107/1072 (10%) consecutive patients presenting to the _____ ILD unit when re-examined were regarded as uILD. 12 patients without departmental imaging were excluded and 95 patients were examined in the study (Supplementary Figure 1). Demographic data, mean visual CT scores, CALIPER scores, echocardiographic and pulmonary function test data are provided in Table 1. Change in CT and pulmonary function variables in 37 patients with serial imaging (median follow up time=1.27 years) and functional data is provided in Supplementary Table 1. Interobserver variation values for baseline and longitudinal visual scores are shown in Supplementary Tables 2a+b respectively.

The median patient age was 65 years and most patients were female and never smokers. At the end of the study follow up period, 45/95 (47%) patients were alive. One patient received a lung transplant and was censored at the time of transplantation. Only 21/95 [22%] patients underwent a surgical lung biopsy, with the majority demonstrating overlapping disease patterns on histology.

6/95 (6%) patients demonstrated a definite UIP pattern on CT, 25/95 (26%) patients had a possible UIP pattern, whilst 64/95 (67%) CTs were inconsistent with a UIP pattern. Criteria by which a CT pattern inconsistent with UIP was assigned are shown in Supplementary Table 3. For the 6 patients with a definite UIP pattern on CT, in three cases there was discrepancy between CT findings and histopathology, and in

three cases there was discrepancy between CT findings and clinical features (Supplementary Table 4).

Baseline results

On univariate analysis of CALIPER variables, the PVV (p=0.0003; C-Index=0.69) and CALIPER fibrosis extent (p=0.001) were the strongest predictors of mortality. Univariate analysis of visual CT variables identified traction bronchiectasis (p<0.0001) the PA diameter (p<0.0001) and the presence of honeycombing (n=52)[p=0.0001] as the strongest predictors of mortality. All PFTs except Kco were strongly predictive of outcome.

PVV was strongly associated with CALIPER ILD and fibrosis extent (R^2=0.73 and 0.47 respectively). PVV also showed strong linkages with other measures of baseline interstitial disease severity (DLco R^2=0.50; CPI R^2=0.61), and therefore PVV was not examined in multivariate models alongside functional indices.

In patients undergoing transthoracic echocardiography (n=82/95, 86%), no difference in baseline variables (age, gender, DLco, CPI [composite physiologic index], CALIPER ILD extent, visual ILD extent) was present when compared to patients that did not undergo transthoracic echocardiography (n=13/95; 14%). In the 53/95 patients with a measured RVSP, only weak correlations were identified with visually-scored PA diameter (R^2=0.28, p<0.0001), CALIPER PVV (R^2=0.01, p=0.52), DLco (R^2=0.08, p=0.05), and Kco (R^2=0.08, p=0.04).

Multivariate mortality analyses

On multivariate Cox regression analysis of CALIPER variables alone, PVV remained the strongest independent predictor of mortality, and improved in strength after correcting for patient age and gender (HR=1.23, CI 1.11-1.37, p<0.0001).

When visually derived CT indices were evaluated in a multivariate model, traction bronchiectasis (HR=1.17, CI 1.05-1.31, p=0.005) the presence of honeycombing (HR=2.18, CI 1.02-4.64, p=0.04) and PA diameter (HR=1.10, CI 1.05-1.16, p<0.0001) were independently predictive of mortality. When PFTs were examined, CPI was strongest single predictor of mortality.

When combined visual CT indices and the CPI were analysed, the resulting three independent predictors of mortality were: CPI (HR=1.06, CI 1.03-1.09, p=0.0002), traction bronchiectasis (HR=1.16, CI 1.05-1.28, p=0.003) and PA diameter (HR=1.08, CI 1.03-1.14, p=0.003). When DLco was substituted for CPI as a severity variable, the results were essentially unchanged: DLco (HR=0.95, CI 0.93-0.98, p=0.0005), traction bronchiectasis (HR=1.18, CI 1.07-1.30, p=0.001) and PA diameter (HR=1.08, CI 1.03-1.14, p=0.003). The three CT patterns (definite, possible and inconsistent with UIP) did not have any independent prognostic value. In all multivariate models, results were maintained when adjusting for patient age and gender. Results were also maintained in a separate subanalysis excluding five patients that demonstrated no traction bronchiectasis on CT.

Serial Analyses

When measures of longitudinal change were examined using univariate analyses in a subset of 37 patients with follow up data, the strongest predictors of outcome were relative FVC change, change in PVV and change in fibrosis extent measured visually and by CALIPER (Table 3). Following adjustment for baseline disease severity using CPI, change in CALIPER fibrosis extent (p=0.001, HR=1.16, CI 1.06-1.28) better predicted outcome than either change in visual fibrosis extent (p=0.01, HR=1.08, CI 1.02-1.14) or relative FVC change (p=0.003, HR=0.94, CI 0.90-0.98). The results were maintained when DLco was used to adjust for baseline disease severity. When examined alongside variables shown previously to predict outcome in uILD, CALIPER fibrosis change was a stronger predictor of outcome than baseline visual fibrosis extent and severity thresholds of DLco<35 % predicted and visual fibrosis extent >20%.

When group differences in baseline measures between patients with and without follow up imaging were examined to identify potential selection biases, the only significant difference between groups was the PA diameter (p=0.01)[Supplementary Table 5]. Kaplan-Meier survival curves did not significantly differ between groups with and without follow up imaging (Log rank test=0.17). Results were again maintained when two patients without traction bronchiectasis were excluded.

Predictors of disease progression

In 29/34 patients with <10% predicted relative FVC decline, the variables strongly predictive of outcome on univariate analysis remained unchanged (Table 4). After

adjustment for baseline disease severity using DLco, change in CALIPER fibrosis extent (p=0.0004, HR=1.19, CI 1.08-1.32) better predicted outcome than change in visual fibrosis extent (p=0.01, HR=1.07, CI 1.02-1.13), PVV change (p=0.02, HR=1.98, CI 1.09-3.58), relative FVC change (p=0.004, HR=0.94, CI 0.90-0.98) and relative DLco change (p=0.03, HR=0.96, CI 0.92-1.00). PVV change was independently predictive of outcome when examined alongside visual and CALIPER fibrosis extent change in separate models.

DISCUSSION

Our study has identified several new findings in patients with uILD. Firstly, at baseline, the CPI was the pulmonary function measure that best predicted mortality. Secondly, the strongest CT predictors of outcome were the severity of traction bronchiectasis and the pulmonary artery diameter. However when adjusting for baseline disease severity, change in CALIPER fibrosis extent was a stronger predictor of outcome than baseline or other longitudinal variables. Importantly, in patients with only marginal FVC change, change in CALIPER fibrosis extent was the stronger predictor of outcome.

Our findings regarding the prevalence of uILD in our tertiary centre, and the proportion of patients undergoing an surgical lung biopsy are both in line with previous reports[4-6, 20, 21]. UILD is often associated with the absence of histological data[4], with only 22% of patients in our cohort undergoing a diagnostic surgical lung biopsy. When histopathological assessment was possible, an overlap of patterns was seen in most cases, which when combined with a non-specific CT pattern and indeterminate clinical features led to no single diagnosis being favoured.

Amongst baseline CALIPER variables, PVV was the strongest determinant of mortality, reproducing findings identified in other fibrosing lung diseases[15, 22]. PVV was also found to be a more malignant prognostic determinant in patients with increased fibrosis on CT. Regarding the visual CT variables that were independently predictive of mortality, traction bronchiectasis severity[23-26] and the pulmonary artery diameter[27, 28] have both been previously shown to strongly predict outcome across

various fibrosing lung diseases, but were not examined in previous studies investigating uILD[4, 29].

In a previous study detailing a population of patients with uILD, Ryerson et al[4] classified patients according to current consensus IPF-CT diagnostic guidelines[30]. In our study, the weak link between a definite UIP pattern and outcome (p=0.04) was similar to the weak links for combined definite and possible UIP patterns (p=0.01) identified by Ryerson et al[4]. Furthermore in both studies, relationships between CT pattern and outcome were weak compared to CT fibrosis extent. Though the relative proportions of patients in each of the three CT pattern groups (definite, possible and inconsistent with UIP) varied between studies, taken together, these findings suggest that in patients with uILD, utilization of the IPF-UIP CT diagnostic criteria may not be useful in indicating patient outcome.

Change in fibrosis extent scored visually and by CALIPER were equivalent at predicting outcome on univariate analysis, but after adjusting for baseline disease severity, change in CALIPER fibrosis extent was shown to be a stronger measure of outcome than any baseline variable. The cardinal description of uILD by Ryerson et al[4] identified baseline CT fibrosis as being independently associated with outcome. The very nature of uILD presupposes a heterogeneous population in which patients with inflammatory etiologies are grouped with established fibrotic conditions. Consequently, with variations in disease trajectories across the population, baseline disease severity measures may be proportionally less important than robust

measures of longitudinal change such as worsening lung fibrosis or relative FVC change that reflect disease behavior.

The uILD patient subgroup with longitudinal FVC measurements showed limited functional decline (2.5% predicted per year) in contrast to IPF populations in which an annual decline of 10% predicted FVC is common in patients not receiving therapy[31]. When FVC declines are marginal, the measurement variation associated with FVC[32] can reduce its strength as an outcome measure. Yet we have demonstrated, albeit in what is essentially pilot data, that in patients with marginal declines (<10% predicted relative FVC decline), CALIPER measures of change in fibrosis extent strongly and independently predicted outcome after adjustment for baseline disease severity. Our findings suggest that quantitative CT analysis has the potential to be part of a composite staging system to identify true decline when FVC change is marginal.

The interobserver variation associated with visual CT scoring was accentuated when variation in scores across two time points were summed to identify disease change. A similar effect may be seen with longitudinal DLco assessment where measurement inaccuracy would be compounded when DLco is evaluated at two time points. In our study, all DLco measurements were performed in a single rigorously calibrated laboratory, and yet DLco change had limited associations with outcome.

A limitation of our study is that only a subset of patients had follow up imaging in our study, however the similarities in baseline functional indices between patients

with and without follow up imaging, suggests that an unintentional selection bias towards patients with less severe disease did not explain the prognostic strength of CALIPER fibrosis extent change. Our findings suggest that larger studies using CALIPER are warranted to characterize the robustness of change in quantitative CT markers as measures of disease progression and outcome.

In conclusion, our study of unclassifiable-interstitial lung disease has identified that the CPI and visual increased traction bronchiectasis severity and pulmonary artery diameter are the strongest independent predictors of outcome at baseline. However, in a subset of patients with longitudinal data and similar baseline measures, change in fibrosis extent measured by CALIPER was independent of, and more powerful than baseline measures of disease severity in patients with uILD. Importantly, when FVC decline was marginal, CALIPER fibrosis extent change was able to independently predict patient outcome.

Variable (n = 95 unless stated) Units are percentage unless stated	Value
Median Age (years)	65
Male/female	45/50
Survival (alive/dead)	55/50
Never smokers/ex/current smokers	54/36/5
FEV1 % predicted (n=90)	72.0 ± 23.7
FVC % predicted (n=90)	72.1 ± 23.8
DLco % predicted (n=90)	41.3 ± 14.4
Kco % predicted (n=90)	71.6 ± 18.4
TLC% predicted (n=87)	69.2 ± 17.2
CPI (n=87)	52.0 ± 13.3
Echocardiography RVSP (mmHg) (n=53)	39.7 ± 16.6
CALIPER ILD extent	26.5 ± 23.8
CALIPER Fibrosis extent	6.7 ± 6.3
CALIPER Ground glass opacity	19.9 ± 20.0
CALIPER Reticular pattern	6.2 ± 6.0
CALIPER Honeycombing	0.4 ± 1.0
CALIPER Grade 1+2 decreased attenuation	0.4 ± 1.0
CALIPER Pulmonary vessel volume	4.8 ± 2.2
CALIPER Normal lung	68.2 ± 25.6
Visual ILD extent	44.9 ± 23.1
Visual fibrosis extent	21.4 ± 15.9
Visual Ground glass opacity	22.1 ± 20.8
Visual Reticular pattern	17.7 ± 12.6
Visual Honeycombing	3.8 ± 7.9
Visual Emphysema	3.3 ± 7.7
Visual TxBx severity (max score 18)	4.6 ± 3.1
Main pulmonary artery diameter (mm)	30.5 ± 5.9
Ascending aorta diameter (mm)	33.9 ± 3.9

Table 1. Patient age, gender and mean and standard deviations of pulmonary function indices, CALIPER and visually scored CT parameters and echocardiography data. Data represent mean values with standard deviations. CT = computed tomography, FEV1 = forced expiratory volume in one second, FVC = forced vital capacity, DLCO = diffusing capacity for carbon monoxide, Kco = carbon monoxide transfer coefficient, TLC = total lung capacity, CPI = composite physiological index, ILD = interstitial lung disease, RVSP = right ventricular systolic pressure TxBx = traction bronchiectasis.

	Number of patients	Hazard ratio	P Value	95.0% Confidence Interval	
				Lower	Upper
CALIPER score					
Total ILD extent	95	1.02	0.002	1.01	1.03
Total fibrosis extent	95	1.06	0.001	1.02	1.10
Ground glass opacity	95	1.02	0.007	1.01	1.03
Reticular pattern	95	1.06	0.003	1.02	1.09
Honeycombing	95	1.27	0.006	1.07	1.52
Grade 2+3 DA	95	1.15	0.29	0.89	1.48
Normal lung	95	0.98	0.001	0.98	0.99
PVV	95	1.21	0.0003	1.09	1.35
VISUAL score					
ILD extent	95	1.02	0.005	1.01	1.03
Fibrosis extent	95	1.03	0.0002	1.02	1.05
Ground glass opacity	95	1.00	0.85	0.99	1.01
Reticular pattern	95	1.03	0.01	1.01	1.05
Honeycombing	95	1.04	0.0003	1.02	1.07
Total emphysema	95	1.01	0.44	0.98	1.05
TxBx severity	95	1.27	<0.0001	1.16	1.38
Main PA	95	1.11	<0.0001	1.05	1.16
PA:Ao ratio	95	1.59	0.003	1.16	2.14
Honeycombing presence	95	3.39	0.0001	1.81	6.35
Definite UIP pattern	95	2.72	0.04	1.07	6.90
Lung Function Indices					
FEV_1	90	0.97	0.0002	0.96	0.99
FVC	90	0.97	<0.0001	0.95	0.98
TLC	87	0.95	<0.0001	0.94	0.97
DLco	90	0.94	<0.0001	0.91	0.96
Kco	90	0.98	0.04	0.96	1.00
CPI	87	1.09	<0.0001	1.06	1.12
ECHOCARDIOGRAPHY					
RVSP	53	1.02	0.02	1.00	1.04

Table 2. Univariate Cox regression analysis demonstrating mortality according to CALIPER indices (top white), visually derived HRCT indices (light grey), pulmonary function tests (dark grey) and echocardiography (bottom white). ILD=Interstitial lung disease, DA=decreased attenuation, TxBx=traction bronchiectasis, PA=pulmonary artery, Ao=Aorta, PVV= pulmonary vessel volume, FEV_1=forced expiratory volume in one second, FVC=forced vital capacity, TLC=total lung capacity, DLco=diffusing capacity for carbon monoxide, Kco=Carbon monoxide transfer coefficient, CPI=composite physiologic index, RVSP=right ventricular systolic pressure. Fibrosis extent is sum of reticular pattern and honeycombing. ILD extent additionally summed ground glass opacities.

Change in variables (per year)	Number of patients	Hazard ratio	P Value	95.0% Confidence Interval	
				Lower	Upper
CALIPER score					
Total ILD change	37	1.02	0.22	0.99	1.04
Total fibrosis change	37	1.17	0.0002	1.08	1.27
Ground glass opacity change	37	1.00	0.84	0.98	1.03
Reticular pattern change	37	1.16	0.001	1.07	1.26
Honeycombing change	37	1.42	0.17	0.86	2.34
Grade 2+3 DA change	37	1.95	0.09	0.90	4.19
Normal lung change	37	0.98	0.15	0.96	1.01
PVV change	37	2.07	0.006	1.23	3.50
VISUAL score					
ILD extent change	37	1.05	0.02	1.01	1.09
Fibrosis extent change	37	1.09	<0.0001	1.04	1.14
Ground glass opacity change	37	1.00	0.86	0.97	1.04
Reticular pattern change	37	1.07	0.008	1.02	1.12
Honeycombing change	37	1.11	0.002	1.04	1.19
Total emphysema change	37	0.91	0.32	0.77	1.09
TxBx severity change	37	1.31	0.04	1.01	1.69
Main PA change	37	1.07	0.30	0.95	1.20
Lung Function Indices					
Absolute FVC change	34	0.92	0.003	0.88	0.97
Relative FVC change	34	0.93	0.0002	0.90	0.97
Absolute DLco change	34	0.92	0.03	0.85	0.99
Relative DLco change	34	0.96	0.01	0.92	0.99
Absolute CPI change	32	1.13	0.009	1.03	1.24

Table 3. Univariate Cox regression analysis demonstrating mortality according to annualized change in CALIPER indices (top white), visually derived HRCT indices (light grey) and pulmonary function tests (dark grey). ILD=Interstitial lung disease, DA=decreased attenuation, TxBx=traction bronchiectasis, PA=pulmonary artery, PVV=pulmonary vessel volume, FVC=forced vital capacity, DLco=diffusing capacity for carbon monoxide, CPI=composite physiologic index. Fibrosis extent is sum of reticular pattern and honeycombing. ILD extent additionally summed ground glass opacities.

Change in variables (per year)	Number of patients	Hazard ratio	P Value	95.0% Confidence Interval	
				Lower	Upper
CALIPER score					
Total ILD change	29	1.02	0.13	1.00	1.04
Total fibrosis change	29	1.16	0.001	1.07	1.26
Ground glass opacity change	29	1.01	0.37	0.98	1.04
Reticular pattern change	29	1.14	0.001	1.06	1.24
Honeycombing change	29	1.41	0.16	0.87	2.28
Grade 2+3 DA change	29	1.71	0.18	0.77	3.71
Normal lung change	29	0.98	0.10	0.96	1.00
PVV change	29	2.27	0.005	1.28	4.04
VISUAL score					
ILD extent change	29	1.04	0.05	1.01	1.09
Fibrosis extent change	29	1.08	0.0005	1.04	1.13
Ground glass opacity change	29	0.99	0.67	0.94	1.04
Reticular pattern change	29	1.06	0.03	1.01	1.11
Honeycombing change	29	1.10	0.009	1.02	1.18
Total emphysema change	29	4.10	0.02	1.30	12.93
TxBx severity change	29	1.19	0.24	0.89	1.58
Main PA change	29	1.02	0.72	0.91	1.15
Lung Function Indices					
Absolute FVC change	29	0.93	0.01	0.87	0.98
Relative FVC change	29	0.93	0.001	0.89	0.97
Absolute DLco change	27	0.91	0.04	0.83	0.99
Relative DLco change	27	0.95	0.009	0.91	0.99
Absolute CPI change	27	1.13	0.02	1.02	1.25

Table 4. Univariate Cox regression analysis in patients with a relative FVC decline <10% predicted. The table demonstrates mortality according to annualized change in CALIPER indices (top white), visually derived HRCT indices (light grey) and pulmonary function tests (dark grey). ILD=Interstitial lung disease, DA=decreased attenuation, TxBx=traction bronchiectasis, PA=pulmonary artery, PVV=pulmonary vessel volume, FVC=forced vital capacity, DLco=diffusing capacity for carbon monoxide, CPI=composite physiologic index. Fibrosis extent is sum of reticular pattern and honeycombing. ILD extent additionally summed ground glass opacities.

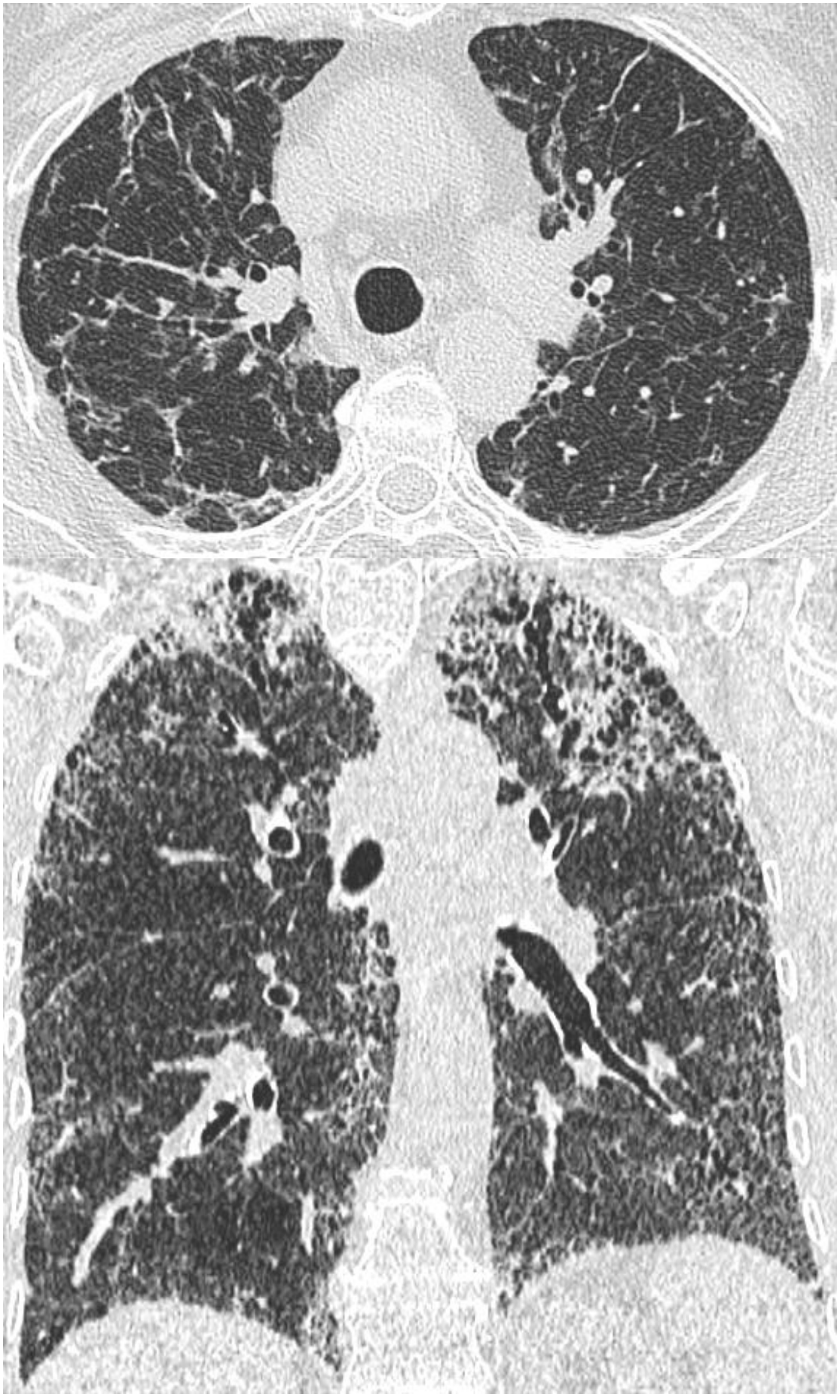

Figure 1. CT imaging for two patients who were found to be inconsistent for a UIP pattern on CT. Patient A (Axial image), a 77-year-old never smoker demonstrated peribronchovascular and perilobular consolidation in keeping with an organizing pneumonia pattern, but did not fulfill criteria for undifferentiated connective tissue disease. Patient B (Coronal image) a 79-year-old never smoker, demonstrated an upper lobe predominant pattern of fibrosis but had no associated CT or clinical features of hypersensitivity pneumonitis and was given a differential diagnosis of idiopathic pulmonary fibrosis/chronic hypersensitivity pneumonitis.

Declaration of Interests

Dr Jacob reports personal fees from Boehringer Ingelheim, outside the submitted work.

BJB, RK, SR report a grant from the Royal Brompton Hospital during the conduct of the study; another from Imbio, LLC, was outside the submitted work; and all have a patent: SYSTEMS AND METHODS FOR ANALYZING IN VIVO TISSUE VOLUMES USING MEDICAL IMAGING DATA licensed to Imbio, LLC.

Dr. Wells reports personal fees from Intermune, personal fees from Boehringer Ingelheim, personal fees from Gilead, personal fees from MSD, personal fees from Roche, personal fees from Bayer, personal fees from Chiesi, outside the submitted work.

Dr. Walsh reports personal fees from Boehringer Ingelheim, personal fees from Roche, outside the submitted work.

Dr. Nicholson reports personal fees from Boehringer Ingelheim, personal fees from Sanofi, personal fees from Roche, personal fees from Intermune, personal fees from MED IQA, outside the submitted work.

Dr. Hansell reports personal fees from AstraZeneca, grants and personal fees from Intermune, personal fees from Boehringer Ingelheim, personal fees from Sanofi, personal fees from Glaxo Smith Kline, personal fees from Roche, outside the submitted work. Dr Hansell is the recipient of a National Institute of Health Research Senior Investigator Award.

The manuscript was supported by the National Institute of Health Research Respiratory Disease Biomedical Research Unit at the Royal Brompton and Harefield NHS Foundation Trust and Imperial College London.

Funding source

There is no funding source for the current study. Joseph Jacob had full access to all the data in the study and had final responsibility for the decision to submit for publication.

REFERENCES

1. Flaherty KR, Andrei AC, King TE, Jr., et al. Idiopathic interstitial pneumonia: do community and academic physicians agree on diagnosis? *AmJRespirCrit Care Med*. 2007;**175**(10):1054-60.
2. Raghu G, Collard HR, Egan JJ, et al. An official ATS/ERS/JRS/ALAT statement: idiopathic pulmonary fibrosis: evidence-based guidelines for diagnosis and management. *Am J Respir Crit Care Med*. 2011;**183**(6):788-824.
3. Walsh SLF, Wells AU, Desai SR, et al. Multicentre evaluation of multidisciplinary team meeting agreement on diagnosis in diffuse parenchymal lung disease: a case-cohort study. *The Lancet Respiratory Medicine*. 2016;**4**(7):557-65.
4. Ryerson CJ, Urbania TH, Richeldi L, et al. Prevalence and prognosis of unclassifiable interstitial lung disease. *Eur Respir J*. 2013 September 1, 2013;**42**(3):750-7.
5. Hyldgaard C, Hilberg O, Muller A, et al. A cohort study of interstitial lung diseases in central Denmark. *Respir Med*. 2014;**108**(5):793-9.
6. Troy L, Glaspole I, Goh N, et al. Prevalence and prognosis of unclassifiable interstitial lung disease. *Eur Respir J*. 2014 May 1, 2014;**43**(5):1529-30.
7. Ooi A, Iyenger S, Ferguson J, et al. VATS Lung Biopsy in Suspected, Diffuse Interstitial Lung Disease Provides Diagnosis, and Alters Management Strategies. *Heart, Lung and Circulation*. 2005;**14**(2):90-2.
8. Sigurdsson MI, Isaksson HJ, Gudmundsson G, et al. Diagnostic Surgical Lung Biopsies for Suspected Interstitial Lung Diseases: A Retrospective Study. *The Annals of Thoracic Surgery*. 2009;**88**(1):227-32.
9. Zhang D, Liu Y. Surgical Lung Biopsies in 418 Patients with Suspected Interstitial Lung Disease in China. *Intern Med*. 2010;**49**(12):1097-102.
10. Casoni GL, Tomassetti S, Cavazza A, et al. Transbronchial Lung Cryobiopsy in the Diagnosis of Fibrotic Interstitial Lung Diseases. *PLoS One*. 2014;**9**(2):e86716.
11. 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**(6):733-48.
12. Lacasse Y, Selman M, Costabel U, et al. Clinical diagnosis of hypersensitivity pneumonitis. *Am J Respir Crit Care Med*. 2003;**168**(8):952-8.
13. Mosca M, Neri R, Bencivelli W, et al. Undifferentiated connective tissue disease: analysis of 83 patients with a minimum followup of 5 years. *The Journal of Rheumatology*. 2002;**29**(11):2345-9.
14. Jacob J, Bartholmai B, Rajagopalan S, et al. Automated quantitative CT versus visual CT scoring in idiopathic pulmonary fibrosis: validation against pulmonary function. *J Thorac Imaging*. 2016;**31**:304-11.
15. Jacob J, Bartholmai B, Rajagopalan S, et al. Mortality prediction in idiopathic pulmonary fibrosis: evaluation of automated computer tomographic analysis with conventional severity measures. *Eur Respir J*. 2016;**doi: 10.1183/13993003.01011-2016**.
16. Wells AU, Desai SR, Rubens MB, et al. Idiopathic pulmonary fibrosis: a composite physiologic index derived from disease extent observed by computed tomography. *Am J Respir Crit Care Med*. 2003;**167**:962-9.

17. Hansell DM, Bankier AA, MacMahon H, et al. Fleischner Society: glossary of terms for thoracic imaging. *Radiology*. 2008;**246**(3):697-722.
18. Ng CS, Wells AU, Padley SP. A CT sign of chronic pulmonary arterial hypertension: the ratio of main pulmonary artery to aortic diameter. *J Thorac Imaging*. 1999;**14**(4):270-8.
19. Maldonado F, Moua T, Rajagopalan S, et al. Automated quantification of radiological patterns predicts survival in idiopathic pulmonary fibrosis. *Eur Respir J*. 2014;**43**(1):204-12.
20. Morell F, Reyes L, Doménech G, et al. Diagnósticos y procedimientos en 500 pacientes consecutivos con sospecha clínica de enfermedad pulmonar intersticial. *Arch Bronconeumol*. 2008;**44**(4):185-91.
21. Xaubet A, Ancochea J, Morell F, et al. Report on the incidence of interstitial lung diseases in Spain. *Sarcoidosis Vasc Diffuse Lung Dis*. 2004;**21**(1):64-70.
22. Jacob J, Bartholmai BJ, Rajagopalan S, et al. Evaluation of computer-based computer tomography stratification against outcome models in connective tissue disease-related interstitial lung disease: a patient outcome study. *BMC Med*. 2016;**14**:190.
23. Edey AJ, Devaraj AA, Barker RP, et al. Fibrotic idiopathic interstitial pneumonias: HRCT findings that predict mortality. *Eur Radiol*. 2011;**21**(8):1586-93.
24. Sumikawa H, Johkoh T, Colby TV, et al. Computed tomography findings in pathological usual interstitial pneumonia: relationship to survival. *Am J Respir Crit Care Med*. 2008;**177**(4):433-9.
25. Walsh SL, Sverzellati N, Devaraj A, et al. Chronic hypersensitivity pneumonitis: high resolution computed tomography patterns and pulmonary function indices as prognostic determinants. *Eur Radiol*. 2012;**22**(8):1672-9.
26. Walsh SL, Sverzellati N, Devaraj A, et al. Connective tissue disease related fibrotic lung disease: high resolution computed tomographic and pulmonary function indices as prognostic determinants. *Thorax*. 2014;**69**(3):216-22.
27. Walsh SLF, Wells AU, Sverzellati N, et al. An integrated clinicoradiological staging system for pulmonary sarcoidosis: a case-cohort study. *The Lancet Respiratory Medicine*. 2014;**2**(2):123-30.
28. Iwasawa T, Kato S, Ogura T, et al. Low-Normal Lung Volume Correlates With Pulmonary Hypertension in Fibrotic Idiopathic Interstitial Pneumonia: Computer-Aided 3D Quantitative Analysis of Chest CT. *American Journal of Roentgenology*. 2014;**203**(2):W166-W73.
29. Ryerson CJ, Vittinghoff E, Ley B, et al. Predicting survival across chronic interstitial lung disease: The ild-gap model. *CHEST Journal*. 2014;**145**(4):723-8.
30. Raghu G, Collard HR, Egan JJ. An official ATS/ERS/JRS/ALAT statement: idiopathic pulmonary fibrosis—evidence-based guidelines for diagnosis and management. *Am J Respir Crit Care Med*. 2011;**183**:788-824.
31. Noble PW, Albera C, Bradford WZ, et al. Pirfenidone in patients with idiopathic pulmonary fibrosis (CAPACITY): two randomised trials. *Lancet*. 2011;**377**(9779):1760-9.
32. Pellegrino R, Viegi G, Brusasco V, et al. Interpretative strategies for lung function tests. *Eur Respir J*. 2005;**26**(5):948-68.