THE UNIVERSITY of EDINBURGH

Edinburgh Research Explorer

Predicting outcomes in rheumatoid arthritis related interstitial lung disease

OPEN ACCESS

Predicting outcomes in rheumatoid arthritis related interstitial lung disease

Joseph Jacob[1,2], Nikhil Hirani[3], Coline H.M. van Moorsel[4,5], Srinivasan Rajagopalan[6], John T. Murchison[7], Hendrik W. van Es[8], Brian J. Bartholmai[6], Frouke T. van Beek[4], Marjolijn H.L. Struik[4], Gareth A. Stewart[9], Maria Kokosi[10], Ryoko Egashira[11], Anne Laure Brun[12], Gary Cross[13], Joseph Barnett[14], Anand Devaraj[14], George Margaritopoulos[10], Ronald Karwoski[15], Elisabetta Renzoni[10], Toby M. Maher[10,16] and Athol U. Wells[10]

ABSTRACT   The aim of this study was to compare radiology-based prediction models in rheumatoid arthritis-related interstitial lung disease (RAILD) to identify patients with a progressive fibrosis phenotype.

RAILD patients had computed tomography (CT) scans scored visually and using CALIPER and forced vital capacity (FVC) measurements. Outcomes were evaluated using three techniques, as follows. 1) Scleroderma system evaluating visual interstitial lung disease extent and FVC values; 2) Fleischner Society idiopathic pulmonary fibrosis (IPF) diagnostic guidelines applied to RAILD; and 3) CALIPER scores of vessel-related structures (VRS). Outcomes were compared to IPF patients.

On univariable Cox analysis, all three staging systems strongly predicted outcome (scleroderma system hazard ratio (HR) 3.78, $p=9\times10^{-5}$; Fleischner system HR 1.98, $p=2\times10^{-3}$; and 4.4% VRS threshold HR 3.10, $p=4\times10^{-4}$). When the scleroderma and Fleischner systems were combined, termed the progressive fibrotic system (C-statistic 0.71), they identified a patient subset (n=36) with a progressive fibrotic phenotype and similar 4-year survival to IPF. On multivariable analysis, with adjustment for patient age, sex and smoking status, when analysed alongside the progressive fibrotic system, the VRS threshold of 4.4% independently predicted outcome (model C-statistic 0.77).

The combination of two visual CT-based staging systems identified 23% of an RAILD cohort with an IPF-like progressive fibrotic phenotype. The addition of a computer-derived VRS threshold further improved outcome prediction and model fit, beyond that encompassed by RAILD measures of disease severity and extent.

Introduction

The development of lung fibrosis in rheumatoid arthritis is recognised in 2–8% of patients [1, 2], but is associated with a three-fold increased risk of mortality [2]. Given the limited number of patients with rheumatoid arthritis-related interstitial lung disease (RAILD) undergoing surgical lung biopsies, attention has focused on evaluating disease patterns on computed tomography (CT) imaging to predict patient outcomes. Several studies have analysed CT scans in RAILD patients using the idiopathic pulmonary fibrosis (IPF) CT diagnostic guidelines [3] and demonstrated the poor outcome in RAILD associated with a usual interstitial pneumonia (UIP) pattern [4–11]. However, despite two studies demonstrating similar outcomes between RAILD and IPF patients [4, 5], the median survival in RAILD patients with a CT UIP pattern has varied between 3.2 and 10.2 years [4–6, 8, 10] in different reports.

A potential limitation when extrapolating IPF patterns on CT to RAILD is that the distribution of disease in RAILD may not be basal-predominant, as honeycomb cysts and reticulation may be concentrated peripherally in the middle or upper zones of the lungs [12]. Accordingly, an IPF-like definite UIP pattern with basal predominance on CT may only capture a proportion of RAILD patients with true honeycombing. Yet, the importance of identifying RAILD patients in whom disease behaviour is relentlessly progressive, or IPF-like, is increasingly relevant. Mechanistic links between RAILD and IPF are increasingly recognised [13] and there is emerging preclinical evidence of the potential role in disease modification from antifibrotics in RAILD [14]. Furthermore, as highlighted in a recent perspective [15], there is growing recognition within the interstitial lung disease (ILD) community that a predominant focus on IPF may have curtailed the identification of rapidly progressive fibrotic phenotypes in non-idiopathic conditions.

Therefore, our primary study aim was to identify, on CT imaging in RAILD populations, patients with a progressive fibrotic phenotype using two staging systems: a system derived in scleroderma evaluating disease extent [16], and a modification of the Fleischner Society IPF diagnostic guidelines [17]. To avoid inherent biases associated with the evaluation of patients presenting to a single tertiary centre, we analysed independent RAILD cohorts presenting to tertiary referral centres in two countries of the United Kingdom. As a secondary aim, we compared survival in RAILD to survival in patients with IPF, and again to maintain robustness of our conclusions evaluated an multicentred international IPF population. Our final aim was to examine whether prognostication using computer analysis of CT imaging in RAILD was independently predictive of mortality when evaluated against the two RAILD CT staging systems.

Methods

Clinical data

Retrospective analyses of ILD databases identified all new consecutive patients presenting to two teritary ILD centres, the Royal Brompton Hospital (London, UK) from January 2007 to July 2014, and the Edinburgh Royal Infirmary (Edinburgh, UK) from January 2005 to December 2015. A diagnosis of rheumatoid arthritis was made according to the American College of Rheumatology/European League Against Rheumatism criteria [18] by specialist rheumatologists at both institutions. The presence of a fibrosing lung disease was initially diagnosed following review by a multidisciplinary team comprising pulmonologists, radiologists and, when biopsy samples were available, histopathologists. ILD presence was confirmed during the detailed scoring of the CT scans. Patients were defined as never-smokers following evaluation of clinical notes if they had a total lifetime tobacco exposure of <100 cigarettes [19].

The RAILD population was compared to 284 IPF patients presenting to the Royal Brompton Hospital (n=179) and St Antonius Hospital (Utrecht, the Netherlands) (n=105), who were evaluated with multidisciplinary team diagnosis using consensus guidelines [3]. The Royal Brompton Hospital IPF

Affiliations: [1]Dept of Respiratory Medicine, University College London, London, UK. [2]Centre for Medical Image Computing, University College London, London, UK. [3]MRC Centre for Inflammation Research, Edinburgh Royal Infirmary, Edinburgh, UK. [4]St Antonius ILD Center of Excellence, Dept of Pulmonology, St Antonius Hospital, Nieuwegein, The Netherlands. [5]Division of Heart and Lungs, University Medical Center Utrecht, Utrecht, The Netherlands. [6]Division of Radiology, Mayo Clinic Rochester, Rochester, MN, USA. [7]Dept of Radiology, Edinburgh Royal Infirmary, Edinburgh, UK. [8]Dept of Radiology, St Antonius Hospital, Nieuwegein, The Netherlands. [9]Edinburgh Lung Fibrosis Clinic, Edinburgh Royal Infirmary, Edinburgh, UK. [10]Interstitial Lung Disease Unit, Royal Brompton Hospital, Royal Brompton and Harefield NHS Foundation Trust, London, UK. [11]Dept of Radiology, Faculty of Medicine, Saga University, Saga City, Japan. [12]Imaging Dept, Hôpital Cochin, Paris-Descartes University, Paris, France. [13]Dept of Radiology, Royal Free Hospital NHS Foundation Trust, London, UK. [14]Dept of Radiology, Royal Brompton Hospital, London, UK. [15]Dept of Physiology and Biomedical Engineering, Mayo Clinic Rochester, Rochester, MN, USA. [16]Fibrosis Research Group, National Heart and Lung Institute, Imperial College, London, UK.

Correspondence: Joseph Jacob, Dept of Respiratory Medicine, Rayne Institute, University College London, London, WC1E 6JJ, UK. E-mail: joseph.jacob@nhs.net

patients comprised consecutive new referrals presenting between July 2011 and December 2014 who had received a baseline volumetric noncontrast CT scan. The IPF patients presenting to St Antonius Hospital comprised newly referred patients who had undergone a baseline volumetric noncontrast CT scan between 2004 and 2015.

Approval for this study of clinically indicated CT and pulmonary function data was obtained from the institutional ethics committees of the Royal Brompton Hospital, St Antonius Hospital and Edinburgh Royal Infirmary, and informed patient consent was not required.

Visual CT evaluation

CT protocols are described in the online supplementary material. Each CT scan was evaluated independently by two radiologists (GC and JB) with 3 and 4 years imaging experience, respectively, blinded to all clinical information. CT patterns quantified on a lobar basis to the nearest 5% included ground glass opacities, reticular pattern and honeycombing (summed as ILD extent) [20]. Extent of emphysema and severity of traction bronchiectasis were quantified on a lobar basis, as previously described [21]. The most disparate 5% of visual scores (equating to two standard deviations) and any disagreement in presence/absence of honeycombing or emphysema were adjudicated by a third scorer (JJ) with 10 years imaging experience, blinded to all clinical information.

In addition, all CT scans were classified into one of five groups using a modification of the Fleischner Society IPF diagnostic guidelines [17] to allow applicability to a RAILD population (figure 1). The first modification involved eschewing a mosaic attenuation pattern from the list of inconsistent features prohibiting a UIP diagnosis. The second modification involved separating disease distribution into that typical of IPF (peripheral, basal or subpleural predominant) and disease distributions not typical of IPF (middle or upper zone predominant). The CT scans were classified accordingly into the following five groups. Group 1: definite UIP pattern in an IPF distribution; group 2: definite UIP pattern not in an IPF distribution; group 3: probable UIP in an IPF distribution; group 4: probable UIP pattern not in an IPF distribution; group 5: features inconsistent with UIP (excluding disease distribution and a mosaic attenuation pattern) (figure 1). The CT scans were scored by a radiologist (JJ) and a pulmonologist specialising in ILD (GM). Any discrepancies in scores were compared to a read by a third scorer (JB), and
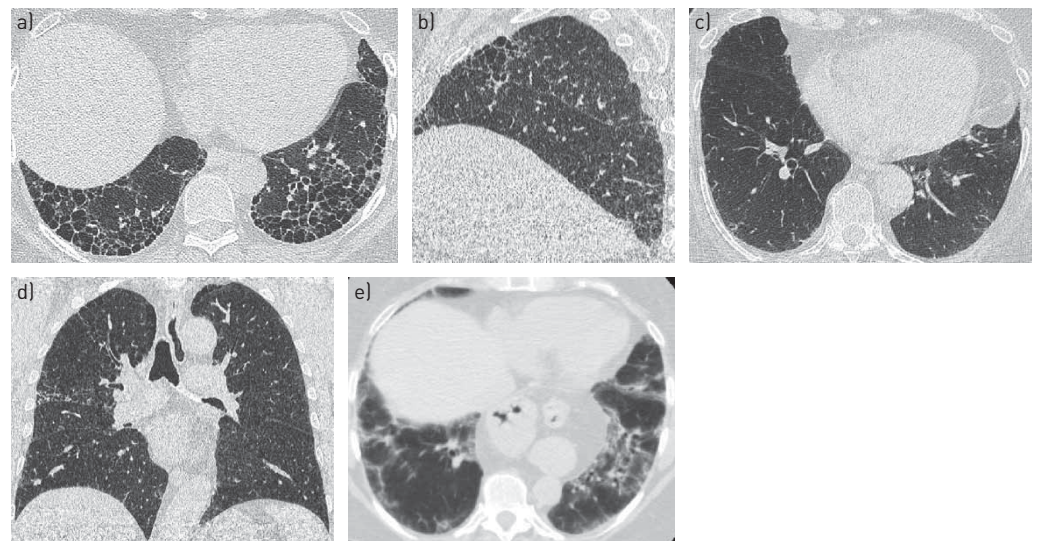

FIGURE 1 Computed tomography (CT) images demonstrating the five groups of the modified Fleischner staging system. a) In a 55-year-old male ex-smoker, honeycombing is visible in a peripheral, basal subpleural distribution typical of a classical usual interstitial pneumonia (UIP) pattern on the axial CT image (group 1); b) a 58-year-old male ex-smoker has honeycomb cysts lying anteriorly above the horizontal fissure on a sagittal CT image with no basal subpleural honeycomb cysts in the lower lobe periphery (group 2); c) in a 57-year-old male never-smoker subtle traction bronchiectasis is visible in the peripheral basal subpleural region of the lower lobes on the axial CT image (group 3); d) a 61-year-old male ex-smoker demonstrates right mid-zone predominant traction bronchiectasis with sparing of the lung bases (group 4) on a coronal CT image; e) a 70-year-old female ex-smoker has evidence of pulmonary fibrosis in a bronchocentric distribution in the left lower lobe, with perilobular arcades of consolidation reminiscent of a fibrosing organising pneumonia pattern in the right lower lobe on the axial CT image; the bronchocentricity to the fibrosis and the organising pneumonia pattern made the CT inconsistent for a UIP pattern (group 5).

the majority view taken. In 17 (11%) out of 157 cases where there was disagreement between all three scorers, adjudication was performed by the original scorers (JJ and GM).

CALIPER CT evaluation

All CTs were evaluated using CALIPER, as previously described [21]. Patterns scored volumetrically by CALIPER included ground glass opacities, reticular pattern and honeycombing (summed as ILD extent), emphysema, normal lung and vessel-related structures (VRS). All parenchymal pattern volumes were expressed as a percentage of the lung, after correcting for total lung volume calculated by CALIPER.

Pulmonary function tests

Pulmonary function test protocols are described in the online supplementary material. Pulmonary function tests examined in the current study included forced expiratory volume in 1 s (FEV1), forced vital capacity (FVC), diffusing capacity of the lung for carbon monoxide (D_{LCO}) and the composite physiological index.

Statistical analysis

Data are presented as mean±SD, or n (%), where appropriate. Interobserver variation for visual scores was assessed using the single-determination standard deviation for continuous variables and as the weighted κ-statistic for ordinal categorical variables. Mean differences between groups were evaluated using a two-sample t-test for parametric continuous variables, and the Mann–Whitney U-test for medians. Differences between two categorical variables were evaluated using the Chi-squared test. Statistical significance was evaluated at a value of p<0.05.

The primary study aim was to identify a progressive fibrotic phenotype in RAILD, and to this end, univariable Cox proportional hazards analyses were used to evaluate mortality prediction in our RAILD population using two staging systems previously validated in other diseases. The first system was a composite staging system developed in scleroderma (scleroderma system) [16], which separates patients into good- and bad-outcome groups, based on CT ILD extent (bad outcome ⩾20% of total lung volume). For cases that might be considered indeterminate (15–25% ILD, representing 30% of the current study population), adjudication was made on the basis of FVC measurements (bad outcome ⩽70% predicted).

The second system was the American Thoracic Society (ATS)/European Respiratory Society (ERS)/ Japanese Respiratory Society (JRS)/Latin American Thoracic Association (ALAT) consensus diagnosis of a UIP pattern on CT [3] modified by the Fleischner Society (Fleischner system) [17]. When using the Fleischner system, patients were categorised into one of five groups, as previously described (figure 1). In addition, the Fleischner system was analysed as a three-group model by combining both definite and probable UIP groups (disregarding disease distribution as a discriminator).

As a separate analysis, we evaluated a threshold of pulmonary VRS of 4.4% of the lung, produced by automated computer-derived CT analytical software (figure 2). The 4.4% threshold, derived by CALIPER, has been shown to separate good- and bad-outcome IPF groups (VRS threshold).

The three systems (scleroderma, Fleischner and VRS) were evaluated separately in univariable Cox proportional hazards analysis and together in a multivariable Cox proportional hazards analysis to identify the contribution of each system to mortality prediction. Multivariable models were adjusted for patient age (years), male sex, smoking status (never versus ever) and baseline percent-predicted FVC, unless stated. Model fit was evaluated using Harrell's C-statistic. Survival estimation was performed via the Kaplan–Meier method. 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 using SPSS (IBM SPSS Statistics for Macintosh, version 20.0; IBM, Armonk, NY, USA) and R Studio Team (2015; RStudio: Integrated Development for R, RStudio, Boston, MA, USA).

Results

Baseline data

The final rheumatoid arthritis study population comprised 90 patients presenting to the Royal Brompton Hospital and 67 patients presenting to Edinburgh Royal Infirmary with a multidisciplinary team diagnosis of RAILD (online supplementary figure S1). Patients underwent a maximum of 6 years of follow-up, with one patient censored before the end of the follow-up period. Baseline differences between the populations at the two institutions are demonstrated in online supplementary table S1, and baseline measures for the combined population are shown in table 1. Edinburgh Royal Infirmary patients were slightly older (p=0.03) than Royal Brompton Hospital patients and had less severe/extensive disease as measured using functional indices (Edinburgh cohort FVC 97.0% versus Brompton cohort FVC 76.0%, p<0.0001) and
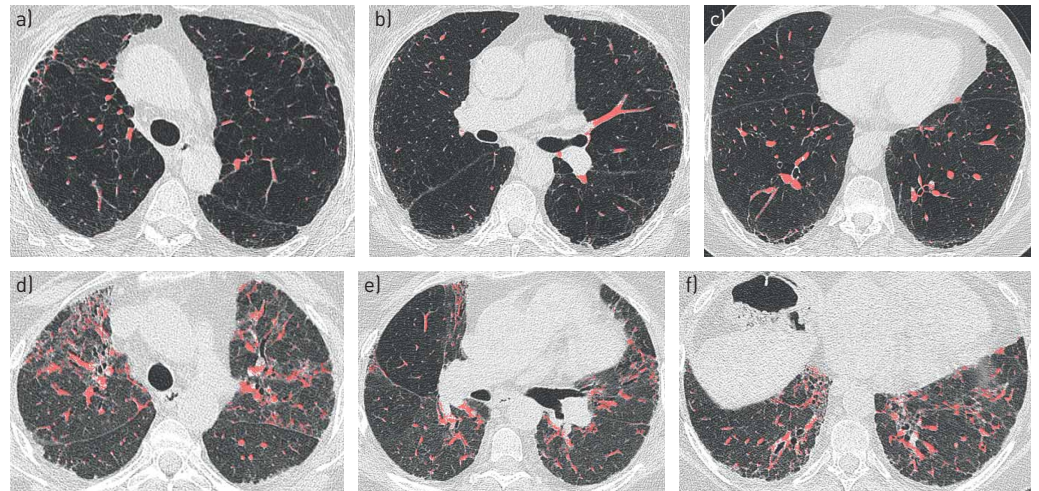
FIGURE 2 Axial computed tomography (CT) image colour maps demonstrating CALIPER-derived vessel-related structures (VRS; red). VRS represent pulmonary arteries and veins (excluding hilar vessels) and connected tubular structures, the latter primarily reflecting adjoining regions of fibrosis. a–c) axial sections in a 71-year-old female 30-pack-year ex-smoker with upper lobe emphysema and fibrosis visible in the lower lobes (VRS 2.1%); d–f) axial sections in a 62-year-old female never-smoker with upper lobe-predominant fibrosis (VRS 7.0%). Nonvascular region captures in the VRS signal are visible in the upper lobes (d) and adjacent to the right hemidiaphragm (f).

TABLE 1 Patient age, sex and pulmonary function indices and visually and CALIPER-scored computed tomography (CT) parameters in patients with rheumatoid arthritis-related interstitial lung disease (ILD)

Median age years	65
Male/female	71/86
Survival (alive/dead)	103/54
Never-smokers/ever-smokers/current smokers[#]	45/83/22
Honeycombing presence (yes/no)	59/98
Emphysema presence (yes/no)	107/50
FEV_1[#] % pred	79.5±21.8
FVC[#] % pred	84.4±24.0
FEV_1/FVC[#] % pred	95.6±13.7
D_{LCO}[¶] % pred	47.3±17.7
CPI[+] %	42.4±16.6
Visual CT scores %	
Total ILD extent	25.1±17.0
Fibrosis extent	21.7±14.2
Ground glass opacity	3.1±6.9
Reticular pattern	18.6±11.3
Honeycombing	3.2±7.5
Total emphysema	7.7±12.1
Traction bronchiectasis severity score (maximum score of 18)	7.7±3.6
CALIPER CT scores %	
Total ILD extent	15.6±16.0
Total fibrosis extent	5.6±5.2
Ground glass opacity	10.1±13.5
Reticular pattern	4.7±4.4
Honeycombing	0.9±2.1
Emphysema	2.5±6.8
Normal lung	77.8±17.1
Vessel-related structures	4.0±1.8

Data are presented as numbers of patients or mean±SD, unless otherwise stated. FEV_1: forced expiratory volume in 1 s; FVC: forced vital capacity; D_{LCO}: diffusing capacity of the lung for carbon monoxide; CPI: composite physiological index. [#]: n=150; [¶]: n=131; [+]: n=128.

visual and CALIPER CT scores (Edinburgh cohort visual ILD extent 16.7% *versus* Brompton cohort visual ILD extent 31.3%, p<0.0001).

The IPF population comprised 179 patients presenting to the Royal Brompton Hospital and 105 presenting to St Antonius Hospital. Interobserver variation scores are shown in online supplementary table S2 and baseline differences between the RAILD and IPF populations are shown in online supplementary table S3. The combined IPF population was older (median age 69 years *versus* 65 years in RAILD, p=0.0004) and more likely to be male (79% *versus* 45% in RAILD, p<0.0001) than RAILD patients. IPF patients had more severe/extensive disease as measured using functional indices (FVC 74.2% *versus* RAILD FVC 84.4%, p<0.0001) and visual and CALIPER CT scores (IPF visual ILD extent 29.1% *versus* RAILD visual ILD extent 25.1%, p=0.01).

Univariable outcome prediction in RAILD
On univariable analysis, visual (fibrosis and honeycombing extents, traction bronchiectasis severity) and CALIPER (fibrosis and honeycombing extents, VRS) measures of interstitial damage powerfully predicted outcome. Functional indices were less powerful in predicting outcome at baseline. A positive smoking history did not predict outcome (table 2).

Outcome prediction using staging systems
When the five-group Fleischner system was examined in RAILD, patients with a definite UIP pattern, regardless of whether the distribution was IPF-like or not, demonstrated a similar outcome (figure 3a). Similarly, patients with a probable UIP pattern, regardless of distribution demonstrated similar 3-year and 6-year survival. The weighted κ for the scoring of the five-group Fleischner system was 0.65.

When disease distribution was ignored, and the Fleischner system was recoded into a three-group score (definite, probable and inconsistent UIP patterns), outcome in patients with a definite UIP pattern was

TABLE 2 Univariable Cox regression analysis demonstrating mortality according to patient age, sex, visual and CALIPER computed tomography (CT) indices and pulmonary function tests in patients with rheumatoid arthritis-related interstitial lung disease (ILD)

	Patients n	Hazard ratio (95% CI)	p-value
Age	157	1.03 (1.01–1.06)	0.02
Sex	157	1.79 (1.04–3.06)	0.03
Smoking (never *versus* ever)	150	1.56 (0.82–2.96)	0.18
Visual score			
ILD extent	157	1.03 (1.02–1.04)	1×10^{-6}
Fibrosis extent[#]	157	1.05 (1.04–1.07)	$<1\times10^{-6}$
Ground glass opacity extent	157	0.99 (0.96–1.03)	0.75
Reticular pattern extent	157	1.04 (1.02–1.06)	2×10^{-6}
Honeycombing extent	157	1.06 (1.04–1.09)	1×10^{-6}
Total emphysema extent	157	1.01 (0.99–1.03)	0.28
Traction bronchiectasis severity	157	1.22 (1.13–1.31)	$<1\times10^{-6}$
Honeycombing presence	157	2.45 (1.43–4.19)	0.001
CT pattern (5-point scale)	157	0.73 (0.61–0.87)	0.0004
CALIPER score			
Total ILD extent	157	1.02 (1.01–1.03)	0.002
Total fibrosis extent[#]	157	1.12 (1.08–1.17)	$<1\times10^{-6}$
Ground glass opacity extent	157	1.01 (1.00–1.03)	0.09
Reticular pattern extent	157	1.12 (1.07–1.18)	3×10^{-6}
Honeycombing extent	157	1.17 (1.08–1.27)	9×10^{-5}
Emphysema extent	157	1.00 (0.96–1.05)	0.96
Normal lung extent	157	0.98 (0.97–0.99)	0.0005
Vessel-related structures	157	1.47 (1.28–1.69)	$<1\times10^{-6}$
Lung function indices			
FEV_1	150	0.99 (0.98–1.00)	0.08
FVC	150	0.98 (0.96–0.99)	0.001
D_{LCO}	131	0.96 (0.94–0.98)	4×10^{-5}
CPI	128	1.05 (1.03–1.07)	1×10^{-5}

ILD extent additionally summed ground glass opacities. FEV_1: forced expiratory volume in 1 s; FVC: forced vital capacity; D_{LCO}: diffusing capacity of the lung for carbon monoxide; CPI: composite physiological index. [#]: sum of reticular pattern and honeycombing.
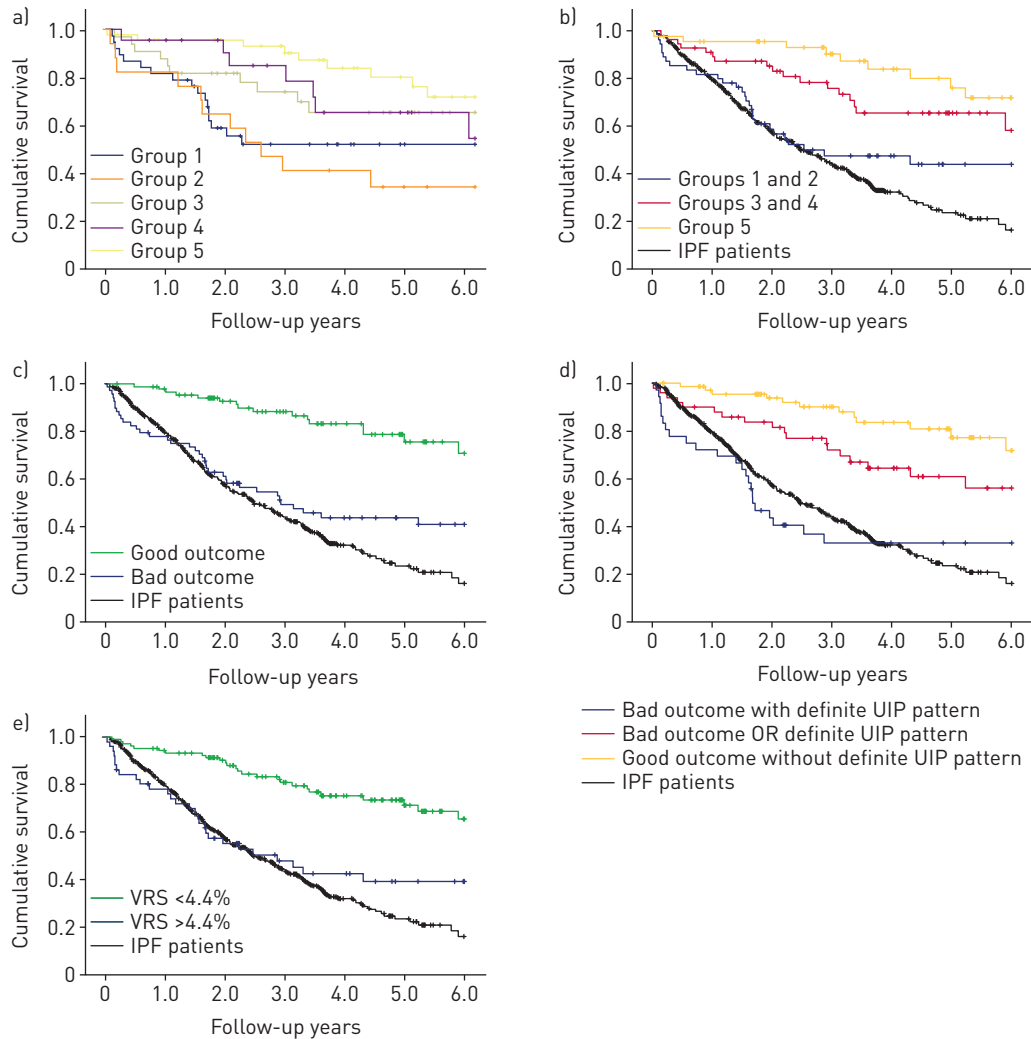
FIGURE 3 a) Kaplan–Meier curves demonstrating mortality for patients with rheumatoid arthritis-related interstitial lung disease (RAILD) subdivided according to the pattern of fibrosis on computed tomography (CT), based on the Fleischner Society idiopathic pulmonary fibrosis (IPF) diagnostic guidelines. Group 1 (honeycombing occurring in an IPF-like distribution): 53% 3-year and 6-year survival, n=38; group 2 (honeycombing occurring in a non IPF-like distribution): 40% 3-year and 35% 6-year survival, n=17; group 3 (fibrosis without honeycombing, occurring in an IPF-like distribution): 75% 3-year and 69% 6-year survival, n=33; group 4 (fibrosis without honeycombing, occurring in a non IPF-like distribution): 85% 3-year and 68% 6-year survival, n=23; group 5 (patients with CT features inconsistent with a universal interstitial pneumonia (UIP) diagnosis; distribution of disease and mosaic attenuation were not considered to be inconsistent features for the purposes of this RAILD study): 92% 3-year and 75% 6-year survival, n=46. b) Combined patients from groups 1 and 2 in figure 1a: 48% 3-year and 45% 6-year survival, n=55 and patients in groups 3 and 4 in figure 1a: 76% 3-year and 58% 6-year survival, n=56 while the group of patients with inconsistent CT features in figure 1a remained unchanged. All three groups were compared to a population of IPF patients: 42% 3-year and 18% 6-year survival, n=284. c) Kaplan–Meier curves demonstrating mortality for patients with RAILD subdivided according to the scleroderma system into good-outcome and bad-outcome groups, and compared with an IPF cohort. Patients with >25% interstitial lung disease (ILD) extent on CT were given a score of 1 (bad outcome), while patients with <15% ILD extent on CT were given a score of 0 (good outcome); patients with 15–25% ILD extent on CT were adjudicated on the basis of forced vital capacity (FVC), with an FVC >70% predicted accorded a score of 0, and FVC <70% pred was accorded a score of 1. Good-outcome scleroderma system 88% 3-year and 70% 6-year survival, n=88; bad-outcome scleroderma system 50% 3-year and 40% 6-year survival, n=68. IPF patients: 42% 3-year and 18% 6-year survival, n=284. d) Kaplan–Meier curves demonstrating mortality for patients with RAILD categorised using the scleroderma system and a definite usual interstitial pneumonia (UIP) pattern on CT (disease distribution and mosaicism were not considered as being inconsistent features of a definite UIP pattern). Patients with a definite UIP pattern on CT had a score of 1 added to their scleroderma system score. Scleroderma system bad outcome with a definite UIP pattern: 35% 3-year and 35% 6-year survival, n=36; either scleroderma system bad outcome or definite UIP pattern: 78% 3-year and 58% 6-year survival, n=50; scleroderma system good outcome without definite UIP: 90% 3-year and 78% 6-year survival, n=70. IPF patients: 42% 3-year and 18% 6-year survival, n=284. e) Kaplan–Meier curves demonstrating mortality for patients with RAILD categorised using a CALIPER vessel-related structure (VRS) score of 4.4%. VRS score >4.4%: 45% 3-year and 40% 6-year survival, n=51; VRS score <4.4% 82% 3-year and 70% 6-year survival, n=106. IPF patients: 42% 3-year and 18% 6-year survival, n=284.

similar to patients with IPF and worse than patients with a probable UIP pattern (figure 3b). On univariable Cox regression analysis, the three-group Fleischner system strongly predicted outcome (hazard ratio (HR) 1.98, 95% CI 1.38–2.85; p=2×10^{-3}; C-statistic 0.67); and the three-point scale is used for analysis in the remainder of the article. On multivariable analysis, no difference in outcome was identified between RAILD patients with a definite UIP pattern (regardless of distribution) and IPF patients.

TABLE 3 Patient age, sex and forced vital capacity (FVC) and visually and CALIPER-scored computed tomography (CT) parameters in rheumatoid arthritis-related interstitial lung disease (RAILD) patients designated as having a bad outcome on the scleroderma staging system and having a definite usual interstitial pneumonia pattern on CT, compared to a population of idiopathic pulmonary fibrosis (IPF) patients

	Progressive fibrosis RAILD cohort	IPF cohort	Cohort differences p-value
Subjects	36	284	
Age years	66 (42–82)	69 (37–92)	0.32[#]
Male/female	21/15	225/59	0.005[¶]
Survival (alive/dead)	13/23	95/189	0.75[¶]
FVC % predicted	66.4±19.3	74.2±19.5[+]	0.03
Visual ILD extent	41.3±15.0	29.1±12.1	<0.0001
Visual traction bronchiectasis severity (maximum score 18)	11.7±2.8	10.2±3.2	0.005
CALIPER ILD extent	23.7±15.0	23.5±16.3	0.94
CALIPER vessel-related structures	5.5±1.5	5.3±1.7	0.41

Data are presented as n, median (range) or mean±SD, unless otherwise stated. Comparisons were made using t-test, unless otherwise stated. ILD: interstitial lung disease. [#]: Mann–Whitney U-test; [¶]: Chi-squared test; [+]: n=283 (as one IPF patient did not have baseline FVC).

On univariable analysis, the scleroderma system strongly predicted outcome (HR 3.78, 95% CI 2.10–6.81, $p=9\times10^{-5}$; C-statistic 0.69) (figure 3c). When examined in a multivariable model, no difference in outcome was identified between IPF patients and poor-outcome RAILD patients (categorised using the scleroderma system). In all multivariable models examining the two staging systems, the results were maintained when visual or CALIPER ILD extent replaced FVC as a measure of baseline disease severity.

Combining real-world staging systems

In order to identify RAILD patients that demonstrated a progressive fibrotic phenotype, the patients with a definite UIP pattern on CT (irrespective of disease distribution) in the bad-outcome group of the scleroderma system were identified and compared to the IPF population (progressive fibrotic system) (figure 3d; table 3). On univariable Cox mortality analysis, the three-point progressive fibrotic system strongly predicted outcome (HR 2.46, 95% CI 1.73–3.50; $p=1\times10^{-5}$; C-statistic 0.71). Results were maintained on multivariable analysis when using FVC, visual or CALIPER ILD extents as measures of baseline disease severity.

Additional impact of VRS threshold

When the CALIPER VRS threshold of 4.4% was examined, outcome prediction was strong (HR 3.10, 95% CI 1.81–5.29; $p=4\times10^{-4}$; C-statistic 0.66) (figure 3e). RAILD patients with a VRS $\geqslant$4.4% had a similar outcome to IPF patients on multivariable analysis, with results maintained when visual or CALIPER ILD extent replaced FVC as a measure of baseline disease severity.

When evaluated in a multivariable model, adjusted for patient age, male sex and smoking status (never versus ever), the VRS threshold independently predicted outcome when separately examined with the scleroderma system (C-statistic 0.78), the Fleischner system (C-statistic 0.75) and the progressive fibrotic system (C-statistic 0.77) (online supplementary table S4). The results demonstrate the additional prognostic information provided by quantitative CT measures in fibrosing lung diseases beyond that captured by visual analysis of CT imaging.

Discussion

Our study has combined two simple staging systems, readily available to clinicians and previously validated in scleroderma and IPF. On applying these staging systems to RAILD patients, we identified 36 (23%) out of 157 of the RAILD population who demonstrate a progressive fibrotic phenotype, with a 4-year survival that is indistinguishable from patients with IPF. In addition, we have shown the additional prognostic value of automated CT analysis in patients with RAILD.

The landmark study of KIM et al. [4] was the first direct comparison of outcomes in RAILD with those of IPF patients. Our findings are in line with that study [4] as a definite UIP pattern, traction bronchiectasis and honeycombing were all associated with worsened survival and female sex and an increased D_{LCO} were linked to better outcome. Yet a definite UIP pattern in RAILD was only just significant as an independent predictor of outcome (p=0.03) in the study by KIM et al. [4], which may relate to the fact that a UIP

pattern in RAILD was evaluated on IPF terms. A basal distribution of disease was a prerequisite for a definite UIP pattern designation. However, as has been documented previously in RAILD, honeycombing on CT, indicative of histological UIP, can predominate in the middle and upper lobes in RAILD [12] or appear bronchocentric [7], causing classification inconsistencies when compared to IPF.

In the current study, when strict IPF diagnostic criteria were used, 38 (24%) out of 157 RAILD patients had a definite UIP pattern on CT. However, when a broader definition, ignoring disease distribution and the presence of a mosaic attenuation pattern (which can occur secondary to co-existing small airways disease in RAILD) was considered, 55 (35%) out of 157 RAILD patients were classified as having a definite UIP pattern. Patients with either UIP definition (strict or broad) demonstrated similar outcomes which matched to those of a control IPF population. Importantly, both definitions demonstrated distinct outcomes when compared to patients classified as probable UIP (using strict and broad definitions) according to the Fleischner Society IPF diagnostic guidelines [17]. By slightly modifying IPF diagnostic criteria to better suit patients with RAILD, we have more comprehensively captured patients demonstrating a bad outcome, and in so doing have confirmed the notion that it is the presence of honeycombing in RAILD, rather than its location, that best determines patient outcome. The clinical applicability of the modified Fleischner system is also evident in the acceptable κ between scorers, which are higher than those documented in the 2011 ATS/ERS/JRS/ALAT diagnostic guidelines [3].

The prognostic strength of the scleroderma staging system in our study reinforces the importance of disease extent as a predictor of outcome in RAILD, a finding highlighted by SOLOMON et al. [5], where fibrosis extent was an independent predictor of survival in RAILD. Indeed, once disease extent had been accounted for, the Fleischner system, though remaining independently predictive of outcome, only added slightly to the goodness of fit of the model. However, when bad-outcome scleroderma patients with a definite UIP pattern were subanalysed, they mimicked a progressive fibrotic phenotype, akin to IPF, more consistently (table 3).

Almost a quarter of our RAILD study population, had extensive and severe (UIP pattern) disease on CT, which behaved like IPF. As the recognition of similarities in disease mechanisms and phenotypes between subsets of patients with RAILD and IPF grows [13], RAILD patients demonstrating a progressive fibrotic phenotype could be considered to be future potential recipients of antifibrotic medication. In the remainder, monitoring of longitudinal disease behaviour as set out by an ATS/ERS expert group for use in unclassifiable ILD [22] might best inform identification of a possible progressive fibrotic phenotype.

In addition to the scleroderma staging system and the modification of the Fleischner Society diagnostic guidelines, we evaluated the utility of automated software on predicting outcome in RAILD. Specifically, we evaluated the VRS in the lungs (the volume of pulmonary arteries and veins, and surrounding fibrosis, excluding vessels at the lung hilum), which have been shown to predict outcome in IPF, at a threshold of 4.4% of the lung volume [23]. In patients with a VRS $\geqslant$4.4%, survival at 3 years was almost identical to IPF patients. As well as independently predicting outcome when analysed alongside the other staging systems, the VRS threshold improved the fit of both the scleroderma and Fleischner systems, suggesting that it provides additional information on outcome beyond that encompassed by simple measures of disease severity and extent. Almost a third of the RAILD population (51 (32%) out of 157 patients) were above the 4.4% threshold, indicating a degree of sensitivity to the VRS measure, which may have utility in analyses of large multicentred RAILD populations, as envisioned in a recent RAILD editorial [24], and where visual scoring of CTs would be impractical.

There were limitations to the current study. While the obvious benefits of automated analysis include its objectivity and reproducibility, extremely edge-enhancing CT algorithms can result in misclassification of patterns such as honeycombing. For this reason, in the current study, CTs from three patients with RAILD CTs reconstructed with a Siemens B60f algorithm were not analysed. To avoid the exclusion of such cases, the routine acquisition of computer-friendly reconstruction algorithms should be encouraged and will become increasingly important as quantitative CT analysis becomes more widespread. In addition, the standardisation of algorithms will improve the likelihood of identifying subtle changes in pattern extents across serial CT examinations. The RAILD patients in the two study centres had differing disease severities, with more advanced disease seen in the Royal Brompton Hospital population. However, we believe that this heterogeneity is a potential strength rather than a major limitation, as it better reflects a real-world cohort of RAILD patients, rather than a selection of advanced, complex RAILD patients referred to a single tertiary London centre. It was for this reason that we also chose to compare the RAILD patients to IPF patients originating from two tertiary centres, instead of a single, homogenous, potentially biased IPF cohort.

In conclusion, we have demonstrated that by combing two staging systems, we are able to identify RAILD patients with a CT UIP pattern, demonstrating a progressive fibrotic phenotype. Representing 23% of our

RAILD population, this cohort had a 4-year survival indistinguishable from patients with IPF. We have also shown the additional prognostic strength of computer analysis of CT imaging which provides information on patient outcome in RAILD, beyond that described by visually scored CT pattern extents.

References

1   Hyldgaard C, Hilberg O, Pedersen AB, et al. A population-based cohort study of rheumatoid arthritis-associated interstitial lung disease: comorbidity and mortality. Ann Rheum Dis 2017; 76: 1700–1706.

2   Bongartz T, Nannini C, Medina-Velasquez YF, et al. Incidence and mortality of interstitial lung disease in rheumatoid arthritis: a population-based study. Arthritis Rheum 2010; 62: 1583–1591.

3   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: 788–824.

4   Kim EJ, Elicker BM, Maldonado F, et al. Usual interstitial pneumonia in rheumatoid arthritis-associated interstitial lung disease. Eur Respir J 2010; 35: 1322–1328.

5   Solomon JJ, Ryu JH, Tazelaar HD, et al. Fibrosing interstitial pneumonia predicts survival in patients with rheumatoid arthritis-associated interstitial lung disease (RA-ILD). Respir Med 2013; 107: 1247–1252.

6   Solomon JJ, Chung JH, Cosgrove GP, et al. Predictors of mortality in rheumatoid arthritis-associated interstitial lung disease. Eur Respir J 2016; 47: 588–596.

7   Lee HK, Kim DS, Yoo B, et al. Histopathologic pattern and clinical features of rheumatoid arthritis-associated interstitial lung disease. Chest 2005; 127: 2019–2027.

8   Yunt ZX, Chung JH, Hobbs S, et al. High resolution computed tomography pattern of usual interstitial pneumonia in rheumatoid arthritis-associated interstitial lung disease: relationship to survival. Respir Med 2017; 126: 100–104.

9   Kelly CA, Saravanan V, Nisar M, et al. Rheumatoid arthritis-related interstitial lung disease: associations, prognostic factors and physiological and radiological characteristics – a large multicentre UK study. Rheumatology 2014; 53: 1676–1682.

10   Tsuchiya Y, Takayanagi N, Sugiura H, et al. Lung diseases directly associated with rheumatoid arthritis and their relationship to outcome. Eur Respir J 2011; 37: 1411–1417.

11   Yang JA, Lee JS, Park JK, et al. Clinical characteristics associated with occurrence and poor prognosis of interstitial lung disease in rheumatoid arthritis. Korean J Intern Med 2017; in press [https://doi.org/10.3904/kjim.2016.349].

12   Rajasekaran BA, Shovlin D, Lord P, et al. Interstitial lung disease in patients with rheumatoid arthritis: a comparison with cryptogenic fibrosing alveolitis. Rheumatology 2001; 40: 1022–1025.

13   Paulin F, Doyle TJ, Fletcher EA, *et al.* Rheumatoid arthritis-associated interstitial lung disease and idiopathic pulmonary fibrosis: shared mechanistic and phenotypic traits suggest overlapping disease mechanisms. *Rev Invest Clin* 2015; 67: 280–286.

14   Redente EF, Aguilar MA, Black BP, *et al.* Nintedanib reduces pulmonary fibrosis in a model of rheumatoid arthritis-associated interstitial lung disease. *Am J Physiol Lung Cell Mol Physiol* 2018; 314: L998–L1009.

15   Wells AU, Brown KK, Flaherty KR, *et al.* What's in a name? That which we call IPF, by any other name would act the same. *Eur Respir J* 2018; 51: 1800692.

16   Goh NS, Desai SR, Veeraraghavan S, *et al.* Interstitial lung disease in systemic sclerosis: a simple staging system. *Am J Respir Crit Care Med* 2008; 177: 1248–1254.

17   Lynch DA, Sverzellati N, Travis WD, *et al.* Diagnostic criteria for idiopathic pulmonary fibrosis: a Fleischner Society white paper. *Lancet Respir Med* 2018; 6: 138–153.

18   Aletaha D, Neogi T, Silman AJ, *et al.* 2010 Rheumatoid arthritis classification criteria: an American College of Rheumatology/European League Against Rheumatism collaborative initiative. *Ann Rheum Dis* 2010; 69: 1580–1588.

19   Centers for Disease Control and Prevention. Cigarette smoking among adults – United States, 2000. *MMMWR Morb Mortal Wkly Rep* 2002; 51: 642–645.

20   Jacob J, Bartholmai B, Rajagopalan S, *et al.* Functional and prognostic effects when emphysema complicates idiopathic pulmonary fibrosis. *Eur Respir J* 2017; 50: 1700379.

21   Jacob J, Bartholmai B, Rajagopalan S, *et al.* Mortality prediction in idiopathic pulmonary fibrosis: evaluation of automated computer-based CT analysis with conventional severity measures. *Eur Respir J* 2017; 49: 1601011.

22   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: 733–748.

23   Jacob J, Bartholmai BJ, Rajagopalan S, *et al.* Predicting outcome in idiopathic pulmonary fibrosis using automated CT analysis. *Am J Respir Crit Care Med* 2018; 198: 767–776.

24   Doyle TJ, Lee JS, Dellaripa PF, *et al.* A roadmap to promote clinical and translational research in rheumatoid arthritis-associated interstitial lung disease. *Chest* 2014; 145: 454–463.