ERJ Express. Published on November 28, 2018 as doi: 10.1183/13993003.00869-2018

EUROPEAN RESPIRATORY *journal*

FLAGSHIP SCIENTIFIC JOURNAL OF ERS

Early View

Original article

Predicting outcomes in rheumatoid arthritis related interstitial lung disease

Joseph Jacob, Nikhil Hirani, Coline H. M. van Moorsel, Srinivasan Rajagopalan, John T. Murchison, Hendrik W. van Es, Brian J. Bartholmai, Frouke T. van Beek, Marjolijn H. L. Struik, Gareth A. Stewart, Maria Kokosi, Ryoko Egashira, Anne Laure Brun, Gary Cross, Joseph Barnett, Anand Devaraj, George Margaritopoulos, Ronald Karwoski, Elisabetta Renzoni, Toby M. Maher, Athol U. Wells

Please cite this article as: Jacob J, Hirani N, van Moorsel CHM, *et al*. Predicting outcomes in rheumatoid arthritis related interstitial lung disease. *Eur Respir J* 2018; in press (https://doi.org/10.1183/13993003.00869-2018).

Title: Predicting outcomes in rheumatoid arthritis related interstitial lung disease

Authors: 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], Athol U. Wells[10].

Affiliations:

[1]Department 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, Scotland, UK

[4]St Antonius ILD Center of Excellence, Department of Pulmonology, St. Antonius Hospital, Nieuwegein, The Netherlands

[5]Division of Heart and Lungs, University Medical Center Utrecht, The Netherlands.

[6]Division of Radiology, Mayo Clinic Rochester, Minnesota, USA

[7]Department of Radiology, Edinburgh Royal Infirmary, Edinburgh, UK

[8]Department of Radiology, St. Antonius Hospital, Nieuwegein, The Netherlands

[9]Edinburgh Lung Fibrosis Clinic, Edinburgh Royal Infirmary, Scotland, UK

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

[11]Department of Radiology, Faculty of Medicine, Saga University, Saga City, Japan.

[12]Imaging department, Hôpital Cochin, Paris-Descartes University, France

[13]Department of Radiology, Royal Free Hospital NHS Foundation Trust, London, UK.

[14]Department of Radiology, Royal Brompton Hospital, London, UK.

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

[16]Fibrosis Research Group, National Heart and Lung Institute, Imperial College, London

*Corresponding author: Joseph Jacob, Department of Respiratory Medicine, Rayne Institute, University College London, London, UK, WC1E 6JJ. Email: joseph.jacob@nhs.net

Authors contributions:

JJ, FTvB, CHMM, MHLS, HWvE, NH, GC, JB, GAS, MK, RE, ALB, AN, TMM, AD, GM, ER, AUW were involved in either the acquisition, or analysis or interpretation of data for the study.

JJ and AUW 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. None of the material has been published or is under consideration elsewhere, including the Internet.

Ethics committee approval

Approval for this study of clinically indicated CT and pulmonary function data was obtained from Liverpool Research Ethics Committee (Reference: 14/NW/0028) and

the Institutional Ethics Committee of the Royal Brompton Hospital, Mayo Clinic Rochester and St. Antonius Hospital, Nieuwegein. Informed patient consent was not required.

Declaration of Interests

JJ reports personal fees from Boehringer Ingelheim outside the current work. JJ was also supported by a Clinical Research Career Development Fellowship from the Wellcome Trust (209553/Z/17/Z).

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.

Prof Maher has, via his institution, received industry-academic funding from GlaxoSmithKline R&D, UCB and Novartis and has received consultancy or speakers fees from Apellis, Astra Zeneca, Bayer, Biogen Idec, Boehringer Ingelheim, Cipla, GlaxoSmithKline R&D, Lanthio, InterMune, ProMetic, Roche, Sanofi-Aventis, Takeda and UCB outside the current work.

Dr. Renzoni reports personal fees from Roche, Boehringer Ingelheim and Takeda, outside the submitted work.

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

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

Nik Hirani via his institution has received industry-academic funding from Boehringer Ingelheim, Intermune, Roche, Galecto and UCB.

The work 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.

Acknowledgements

Dr Jacob was supported by a Wellcome Trust Clinical Research Career Development Fellowship 209553/Z/17/Z.

Abstract: Aims: To compare radiology-based prediction models in rheumatoid arthritis-related interstitial lung disease (RA-ILD) to identify patients with a progressive fibrosis phenotype.

Methods: RAILD patients had CTs scored visually and by CALIPER and forced vital capacity (FVC) measurements. Outcomes were evaluated using three techniques: 1.Scleroderma system evaluating visual ILD extent and FVC values; 2.Fleischer Society IPF diagnostic guidelines applied to RAILD; 3.CALIPER scores of vessel-related structures (VRS). Outcomes were compared to IPF patients.

Results: On univariable Cox analysis, all three staging systems strongly predicted outcome: Scleroderma System:HR=3.78, $p=9x10^{-5}$; Fleischner System:HR=1.98, $p=2x10^{-3}$; 4.4% VRS threshold:HR=3.10, $p=4x10^{-4}$. 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, gender and smoking status, when analysed alongside the Progressive Fibrotic System, the VRS threshold of 4.4% independently predicted outcome (Model C-statistic=0.77).

Conclusions: 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.

ABBREVIATIONS

CALIPER	Computer-Aided Lung Informatics for Pathology Evaluation and Rating
CI	Confidence interval
CPI	Composite physiologic index
CT	Computed tomography
VRS	Vessel-related structure
DLco	Diffusing capacity for carbon monoxide
FEV1	Forced expiratory volume in one second
FVC	Forced vital capacity
HR	Hazard ratio
ILD	Interstitial lung disease
IPF	Idiopathic pulmonary fibrosis
PFT	Pulmonary function test
RA	Rheumatoid arthritis
TxBx	Traction bronchiectasis

TAKE HOME MESSAGE

Combining the Scleroderma and Fleischner Staging Systems in RAILD identifies patients with an IPF-like progressive fibrotic phenotype. CALIPER VRS scores further improve model fit, beyond that encompassed by RAILD measures of disease severity and extent.

INTRODUCTION

The development of lung fibrosis in rheumatoid arthritis is recognized in between 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 CT imaging to predict patient outcomes. Several studies have analysed CTs 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 3.2-10.2 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. And yet, the importance of identifying RAILD patients in whom disease behavior is relentlessly progressive, or IPF-like, is increasingly relevant. Mechanistic links between RAILD and IPF are increasingly

recognised(13) and there is emerging pre-clinical evidence of the potential role in disease modification from anti-fibrotics in RAILD(14). Furthermore, as highlighted in a recent perspective(15), there is growing recognition within the ILD community that a predominant focus on IPF may have curtailed the identification of rapidly progressive fibrotic phenotypes in non-idiopathic conditions.

Our primary study aim was therefore 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, and a modification of the Fleischner Society IPF diagnostic guidelines(16). 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 Kindgom. 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 multicentered 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 from January 2007 to July 2014, and to Edinburgh Royal Infirmary 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(17) 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 less than 100 cigarettes(18).

The RAILD population was compared to 284 IPF patients presenting to the Royal Brompton Hospital, London (n=179) and St Antonius Hospital, Utrecht, Netherlands (n=105), who were evaluated with multidisciplinary team diagnosis using consensus guidelines(3). The Royal Brompton Hospital IPF patients comprised consecutive new referrals presenting between July 2011 to December 2014 who had received a baseline volumetric non-contrast CT scan. The IPF patients presenting to the St Antonius Hospital comprised newly referred patients that had undergone a baseline volumetric non-contrast CT scan between 2004 and 2015.

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

Visual CT Evaluation

CT protocols are described in the Supplementary Appendix. Each CT scan was evaluated independently by two radiologists (GC, 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) (19). Emphysema extent and traction bronchiectasis severity were quantified on a lobar basis as previously described(20). The most disparate 5% of visual scores (equating to two standard deviations) and any disagreement in presence/absence of honeycombing or emphysema was adjudicated by a third scorer (JJ) with 10 years imaging experience blinded to all clinical information.

All CTs were also classified into one of five groups using a modification of the Fleischner Society IPF diagnostic guidelines(16) to allow applicability to an 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, subpleural predominant) and disease distributions not typical of IPF (middle or upper zone predominant). The CTs were accordingly classified 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 CTs were scored by a radiologist (JJ) and a pulmonologist specializing in ILD (GM). Any discrepancies in scores were compared to a read by a third scorer (JB), and the majority view taken. In 17/157(11%) 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 by CALIPER as previously described(20). 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 Supplementary Appendix. Pulmonary function tests examined in the current study included forced expiratory volume in one second (FEV1), forced vital capacity (FVC), diffusing capacity for carbon monoxide (DLco) and the composite physiologic index (CPI).

Statistical analysis

Data are given as means with standard deviations, 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 as the weighted Kappa 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 $p<0{\cdot}05$.

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)(21), which separates patients into a good and bad-outcome group, 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 ATS/ERS/JRS/ALAT consensus diagnosis of a UIP pattern on CT (3) modified by the Fleischner Society (Fleischner System)(16). When using the Fleischner System, patients were categorized into one of five groups as previously described (Figure 1). The Fleischner System was also 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 vessel-related structures (VRS) of 4.4% of the lung, produced by automated computer-derived CT analytic 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 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 gender, smoking status (never versus ever) and baseline percent-predicted FVC unless stated. Model fit was evaluated using Harrells 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 with SPSS (IBM SPSS Statistics for Macintosh, Version 20·0. (IBM. Corp., Armonk, NY, USA)) and RStudio Team (2015). RStudio: Integrated Development for R. RStudio, Inc., Boston, MA.

RESULTS

Baseline data

The final RA 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 (Supplementary Figure 1). 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 Supplementary Table 1, 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% vs Brompton cohort FVC=76.0%; p<0.0001) and visual and CALIPER CT scores (Edinburgh cohort visual ILD extent=16.7% vs Brompton cohort visual ILD extent =31.3%; p<0.0001).

The IPF population comprised 179 patients presenting to the Royal Brompton Hospital, London and 105 presenting to the St Antonius Hospital, Utrecht, Netherlands. Interobserver variation scores are shown in Supplementary Table 2 and baseline differences between the RAILD and IPF populations are shown in Supplementary Table 3. The combined IPF population were older (median age=69 vs 65 in RAILD; p=0.0004) and more likely to be male (45% vs 79% in RAILD; p<0.0001) than RAILD patients. IPF patients had more

severe/extensive disease as measured using functional indices (FVC=74.2% vs RAILD FVC=84.4%; p<0.0001) and visual and CALIPER CT scores (IPF visual ILD extent=29.1% vs 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 5-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 a similar 3-year and 6-year survival. The weighted kappa for the scoring of the 5-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 similar to patients with IPF and worse than patients with a probable UIP pattern (Figure 3b). On univariable Cox regression analysis, the 3-group Fleischner System strongly predicted outcome: HR=1.98, 95%CI=1.38-2.85, $p=2x10^{-3}$, C-statistic=0.67, and the 3-point scale was used for analysis in the remainder of the

manuscript. On multivariable analysis, no difference in outcome was identified between RAILD patients with a definite UIP pattern (regardless of distribution) and IPF patients.

On univariable analysis, the Scleroderma System strongly predicted outcome: HR=3.78, 95%CI=2.10-6.81, $p=9x10^{-5}$, 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 (categorized 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]. On univariable Cox mortality analysis, the three-point Progressive Fibrotic System strongly predicted outcome: HR=2.46, 95%CI=1.73-3.50, $p=1x10^{-5}$, 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, $p=4x10^{-4}$, C-statistic=0.66 (Figure 3e). RAILD patients with a VRS ≥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 gender 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)[Supplementary Table 4]. 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 have identified 36/157 (23%) of the RAILD population that demonstrate a progressive fibrotic phenotype, with a 4-year survival that is indistinguishable from patients with IPF. We have also 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 gender and an increased DLco 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 pre-requisite 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, 38/157 (24%) 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, 55/157 (35%) 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(16). 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 Kappas 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 also 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.

Almost a quarter of our RAILD study population, had extensive and severe (UIP pattern) disease on CT, that 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 behavior as set out by an ATS/ERS expert group for use in unclassifiable ILD(23) 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 also evaluated the utility of automated software on predicting outcome in RAILD. Specifically, we evaluated the vessels-related structures 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(22). In patients with a VRS ≥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 (51/157 [32%] 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. Whilst 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 3 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 standardization of algorithms will also improve the likelihood of identifying subtle changes in pattern extents across serial CT examinations. The RAILD patients in the two study centres also had differing disease severities, with more advanced disease seen in the Royal Brompton Hospital population. However, we believe 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, Ulrichsen SP, Løkke A, Bendstrup E, et al. A population-based cohort study of rheumatoid arthritis-associated interstitial lung disease: comorbidity and mortality. Ann Rheum Dis. 2017;76(10):1700-6.
2. Bongartz T, Nannini C, Medina-Velasquez YF, Achenbach SJ, Crowson CS, Ryu JH, et al. Incidence and Mortality of Interstitial Lung Disease in Rheumatoid Arthritis: A Population Based Study. Arthritis Rheum. 2010;62(6):1583-91.
3. Raghu G, Collard HR, Egan JJ, Martinez FJ, Behr J, Brown KK, 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, Webb WR, Ryu JH, Van Uden JH, et al. Usual interstitial pneumonia in rheumatoid arthritis-associated interstitial lung disease. Eur Respir J. 2010;35(6):1322-8.
5. Solomon JJ, Ryu JH, Tazelaar HD, Myers JL, Tuder R, Cool CD, et al. Fibrosing interstitial pneumonia predicts survival in patients with rheumatoid arthritis-associated interstitial lung disease (RA-ILD). Respir Med. 2013;107(8):1247-52.
6. Solomon JJ, Chung JH, Cosgrove GP, Demoruelle MK, Fernandez-Perez ER, Fischer A, et al. Predictors of mortality in rheumatoid arthritis-associated interstitial lung disease. Eur Respir J. 2016;47(2):588.
7. Lee HK, Kim DS, Yoo B, Seo JB, Rho JY, Colby TV, et al. Histopathologic pattern and clinical features of rheumatoid arthritis-associated interstitial lung disease. Chest. 2005;127:2019-27.
8. Yunt ZX, Chung JH, Hobbs S, Fernandez-Perez ER, Olson AL, Huie TJ, et al. High resolution computed tomography pattern of usual interstitial pneumonia in rheumatoid arthritis-associated interstitial lung disease: Relationship to survival. Respir Med. 2017;126:100-4.
9. Kelly CA, Saravanan V, Nisar M, Arthanari S, Woodhead FA, Price-Forbes AN, et al. Rheumatoid arthritis-related interstitial lung disease: associations, prognostic factors and physiological and radiological characteristics—a large multicentre UK study. Rheumatology (Oxford). 2014;53:1676-82.
10. Tsuchiya Y, Takayanagi N, Sugiura H, Miyahara Y, Tokunaga D, Kawabata Y, et al. Lung diseases directly associated with rheumatoid arthritis and their relationship to outcome. Eur Respir J. 2011;37(6):1411-7.

11. Yang JA, Lee JS, Park JK, Lee EB, Song YW, Lee EY. Clinical characteristics associated with occurrence and poor prognosis of interstitial lung disease in rheumatoid arthritis Korean J Intern Med. 2017;0(0):0-.
12. Rajasekaran BA, Shovlin D, Lord P, Kelly CA. Interstitial lung disease in patients with rheumatoid arthritis: a comparison with cryptogenic fibrosing alveolitis. Rheumatology (Oxford). 2001;40(9):1022-5.
13. Paulin F, Doyle TJ, Fletcher EA, Ascherman DP, Rosas IO. Rheumatoid Arthritis-associated Interstitial Lung Disease and Idiopathic Pulmonary Fibrosis: Shared mechanistic and phenotypic traits suggest overlapping disease mechanisms. Revista de investigacion clinica; organo del Hospital de Enfermedades de la Nutricion. 2015;67(5):280-6.
14. Redente EF, Aguilar MA, Black BP, Edelman B, Bahadur A, Humphries SM, et al. Nintedanib reduces pulmonary fibrosis in a model of rheumatoid arthritis-associated interstitial lung disease. American Journal of Physiology-Lung Cellular and Molecular Physiology.0(0):null.
15. Wells AU, Brown KK, Flaherty KR, Kolb M, Thannickal VJ. What's in a name? That which we call IPF, by any other name would act the same. Eur Respir J. 2018;In press.
16. Lynch DA, Sverzellati N, Travis WD, Brown KK, Colby TV, Galvin JR, et al. Diagnostic criteria for idiopathic pulmonary fibrosis: a Fleischner Society White Paper. The Lancet Respiratory Medicine. 2018;6(2):138-53.
17. Aletaha D, Neogi T, Silman AJ, Funovits J, Felson DT, Bingham CO, et al. 2010 Rheumatoid arthritis classification criteria: an American College of Rheumatology/European League Against Rheumatism collaborative initiative. Ann Rheum Dis. 2010;69(9):1580-8.
18. Control CfD. Cigarette smoking among adults—United States. Morbidity and Mortality Weekly Report. 2002;51:642-5.
19. Jacob J, Bartholmai B, Rajagopalan S, Kokosi M, Maher T, Nair A, et al. Functional and prognostic effects when emphysema complicates idiopathic pulmonary fibrosis. Eur Respir J. 2017;50.
20. Jacob J, Bartholmai B, Rajagopalan S, Kokosi M, Nair A, Karwoski R, 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.
21. Goh NS, Desai SR, Veeraraghavan S, Hansell DM, Copley SJ, Maher TM, et al. Interstitial lung disease in systemic sclerosis: a simple staging system. AmJRespirCrit Care Med. 2008;177(11):1248-54.
22. Jacob J, Bartholmai BJ, Rajagopalan S, Moorsel CHMv, Es HWv, Beek FTv, et al. Predicting outcome in idiopathic pulmonary fibrosis using automated CT analysis. AJRCCM. 2018:doi: 10.1164/rccm.201711-2174OC.
23. Travis WD, Costabel U, Hansell DM, King TE, Jr., Lynch DA, Nicholson AG, 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.
24. Doyle TJ, Lee JS, Dellaripa PF, Lederer JA, Matteson EL, Fischer A, et al. A Roadmap to Promote Clinical and Translational Research in Rheumatoid Arthritis-Associated Interstitial Lung Disease. Chest. 2014;145(3):454-63.

Variable (n = 157 unless stated) Units are percentage unless stated	Value
Median Age (years)	65
Male/female	71/86
Survival (alive/dead)	103/54
Never smokers/ever/current-smokers (n=150)	45/83/22
Honeycombing presence (Y/N)	59/98
Emphysema presence (Y/N)	107/50
FEV1 % predicted (n=150)	79.5 ± 21.8
FVC % predicted (n=150)	84.4 ± 24.0
FEV1/FVC % predicted (n=150)	95.6 ± 13.7
DLco % predicted (n=131)	47.3 ± 17.7
CPI (n=128)	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
TxBx severity	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

Table 1. Patient age, gender and mean and standard deviations of pulmonary function indices and visually and CALIPER scored CT parameters in patients with rheumatoid arthritis-related interstitial lung disease. Data represent mean values with standard deviations for continuous variables and medians for categorical variables. FEV1=forced expiratory volume in one second, FVC=forced vital capacity, DLco=diffusing capacity for carbon monoxide, CPI=composite physiological index, ILD=interstitial lung disease.

	Number of patients	Hazard ratio	95.0% Confidence Interval		P Value
			Lower	Upper	
Age	157	1.03	1.01	1.06	0.02
Gender	157	1.79	1.04	3.06	0.03
Smoking (never vs ever)	150	1.56	0.82	2.96	0.18
VISUAL score					
ILD extent	157	1.03	1.02	1.04	$1x10^{-6}$
Fibrosis extent	157	1.05	1.04	1.07	$<1x10^{-6}$
Ground glass opacity extent	157	0.99	0.96	1.03	0.75
Reticular pattern extent	157	1.04	1.02	1.06	$2x10^{-6}$
Honeycombing extent	157	1.06	1.04	1.09	$1x10^{-6}$
Total emphysema extent	157	1.01	0.99	1.03	0.28
TxBx severity	157	1.22	1.13	1.31	$<1x10^{-6}$
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	$<1x10^{-6}$
Ground glass opacity extent	157	1.01	1.00	1.03	0.09
Reticular pattern extent	157	1.12	1.07	1.18	$3x10^{-6}$
Honeycombing extent	157	1.17	1.08	1.27	$9x10^{-5}$
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	$<1x10^{-6}$
Lung Function Indices					
FEV_1	150	0.99	0.98	1.00	0.08
FVC	150	0.98	0.96	0.99	0.001
DLco	131	0.96	0.94	0.98	$4x10^{-5}$
CPI	128	1.05	1.03	1.07	$1x10^{-5}$

Table 2. Univariable Cox regression analysis demonstrating mortality according to patient age, gender, visual and CALIPER CT indices and pulmonary function tests in patients with rheumatoid arthritis-related interstitial lung disease. ILD=Interstitial lung disease, TxBx=traction bronchiectasis, FEV_1=forced expiratory volume in one second, 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.

Variable	Progressive fibrosis RAILD Cohort (n=36)	IPF Cohort (n=284)	Cohort differences
Median Age (range)	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 (36)	74.2 ± 19.5 (283)	0.03
Visual ILD extent	41.3 ± 15.0	29.1 ± 12.1	<0.0001
Visual TxBx severity (max 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

Table 3. Patient age, gender and mean and standard deviations of forced vital capacity (FVC) and visually and CALIPER scored CT parameters in rheumatoid arthritis-related interstitial lung disease (RAILD) patients designated as having a bad outcome on the scleroderma staging system and also having a definite usual interstitial pneumonia pattern on CT (Column 1), compared to a population of idiopathic pulmonary fibrosis (IPF) patients (Column 2). Data represent mean values with standard deviations. Comparisons were made with the Students t-test, unless indicated: ^=Mann Whitey U test, *=Chi-squared test. FVC=forced vital capacity, ILD=interstitial lung disease, TxBx=traction bronchiectasis.

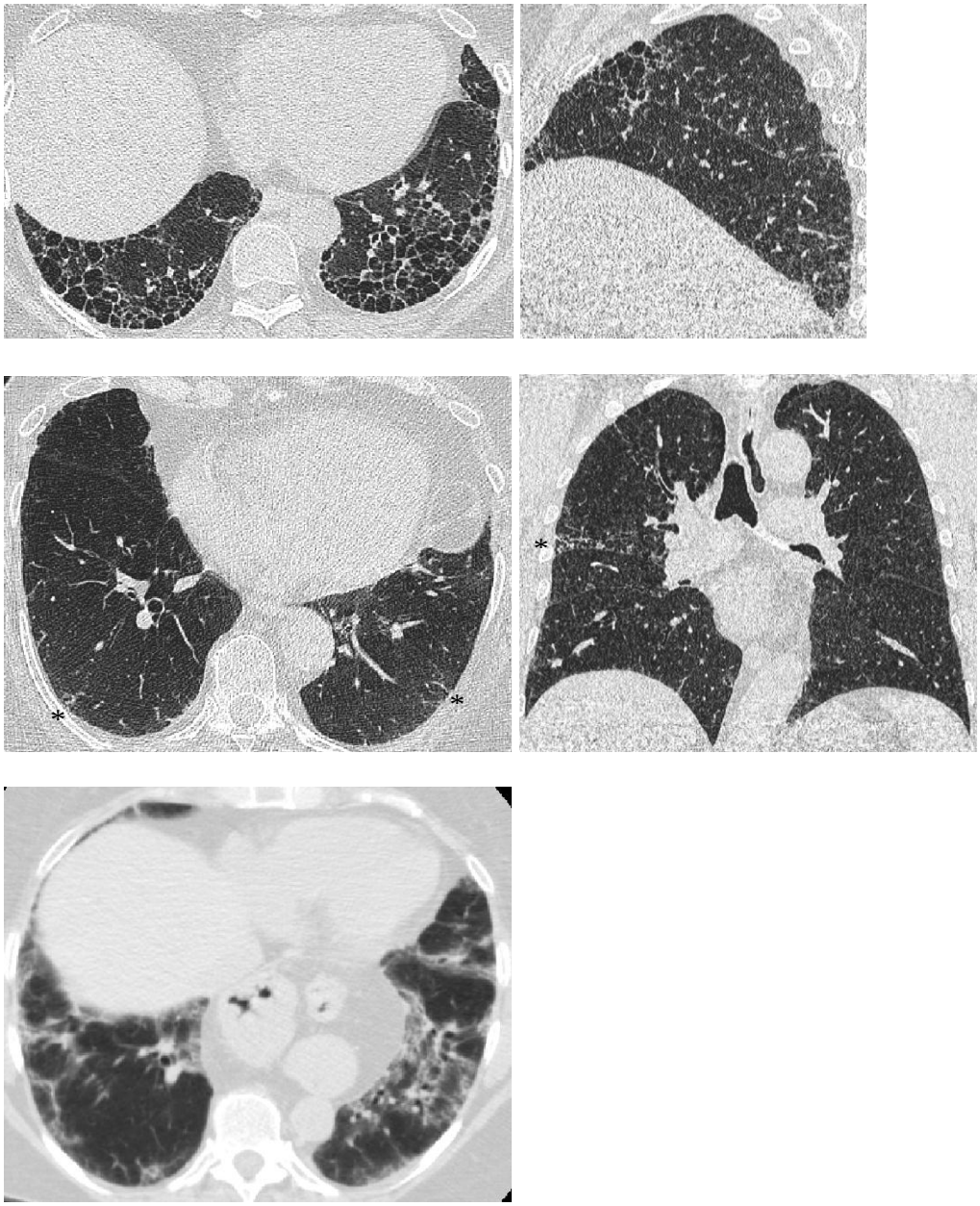


Figure 1. CT images demonstrating the five groups of the modified Fleischer Staging System. In a 55-year-old male ex-smoker (Figure 1a), honeycombing is visible in a peripheral, basal subpleural distribution typical of a classical usual interstitial pneumonia pattern on the axial CT image (Group 1). A 58-year-old male ex-smoker has honeycomb cysts lying anteriorly above the horizontal fissure on a sagittal CT image (Figure 1b) with no basal subpleural honeycomb cysts in the lower lobe periphery (Group 2). In a 57-year-old male never-smoker (Figure 1c) subtle traction bronchiectasis (asterix) is visible in the peripheral basal

subpleural region of the lower lobes on the axial CT image (Group 3). A 61-year-old male ex-smoker demonstrates right mid-zone predominant traction bronchiectasis (asterix) with sparing of the lung bases (Group 4) on a coronal CT image (Figure 1d). 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 organizing pneumonia pattern in the right lower lobe on the axial CT image (Figure 1e). The bronchocentricity to the fibrosis and the organizing pneumonia pattern made the CT inconsistent for a usual interstitial pneumonia pattern (Group 5).

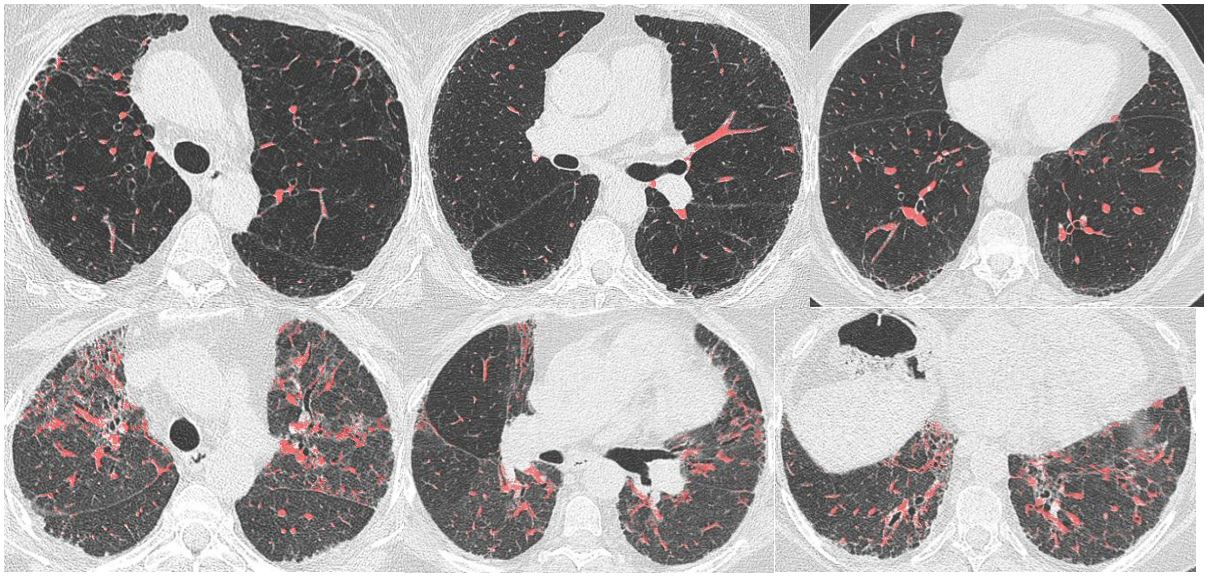


Figure 2. Axial CT image colour maps demonstrating CALIPER-derived vessel-related structures labelled in red. Vessel-related structures represent pulmonary arteries and veins (excluding hilar vessels) and connected tubular structures, the latter primarily reflecting adjoining regions of fibrosis. Figure 2a-c represents 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%). Figure 2d-f represents axial sections in a 62-year-old female never-smoker with upper lobe predominant fibrosis (VRS=7.0%). Non-vascular regions captures in the VRS signal are visible in the upper lobes (d) and adjacent to the right hemidiaphragm (f).

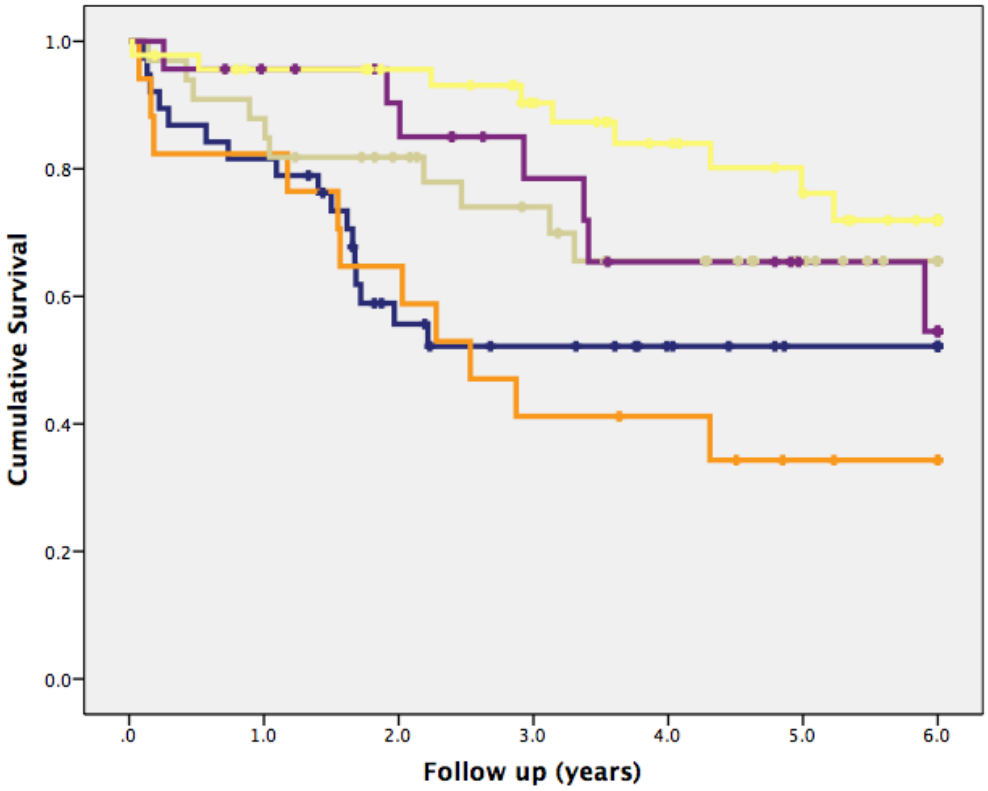


Figure 3a. Kaplan Meier curves demonstrating mortality for patients with rheumatoid arthritis-related interstitial lung disease (RAILD) subdivided according to the pattern of fibrosis on CT, based on the Fleischner Society idiopathic pulmonary fibrosis (IPF) diagnostic guidelines (a). Group 1 (honeycombing occurring in an IPF-like distribution): blue, 53% 3-year and 6-year survival, n=38; group 2 (honeycombing occurring in a non IPF-like distribution): orange, 40% 3-year and 35% 6-year survival, n=17; group 3 (fibrosis without honeycombing, occurring in an IPF-like distribution): beige, 75% 3-year and 69% 6-year survival, n=33; group 4 (fibrosis without honeycombing, occurring in a non IPF-like distribution): violet, 85% 3-year and 68% 6-year survival, n=23; group 5 (patients with CT features inconsistent with a UIP diagnosis – nb distribution of disease and mosaic attenuation were not considered to be inconsistent features for the purposes of this RAILD study): yellow, 92% 3-year and 75% 6-year survival, n=46.

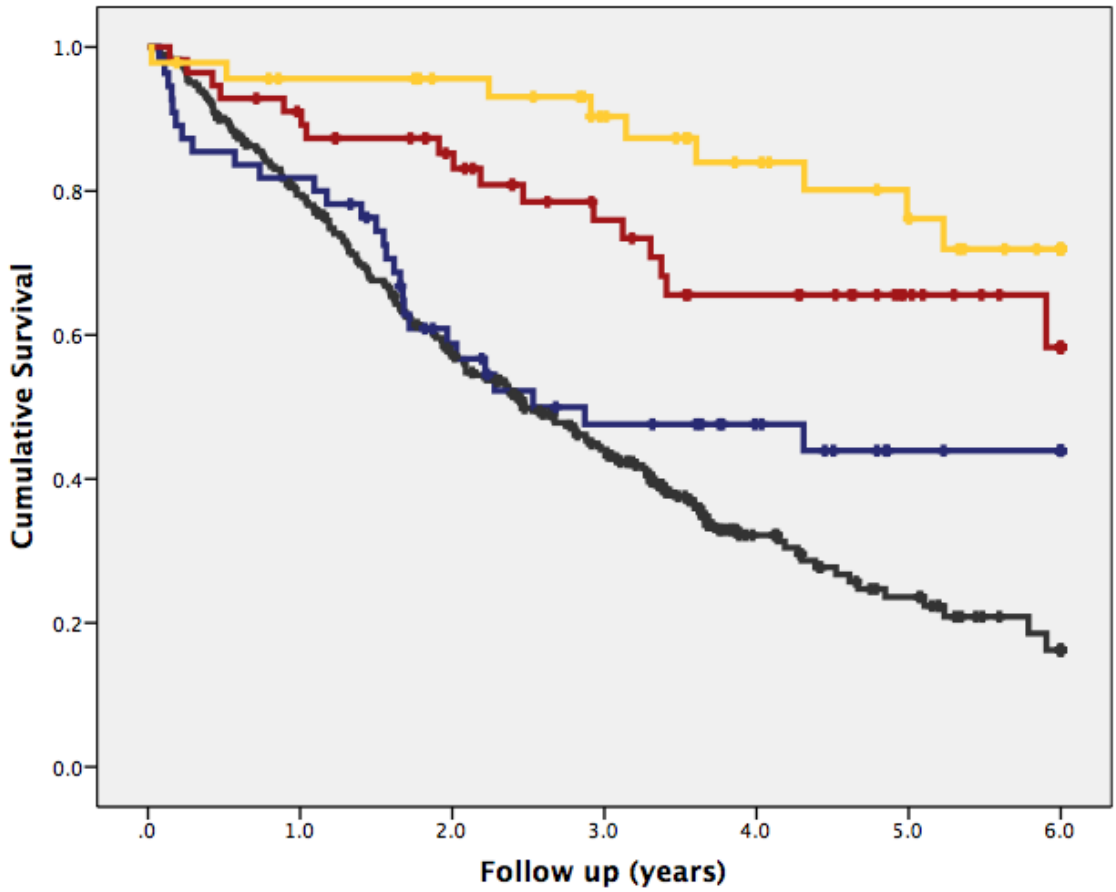


Figure 3b. Combined patients from groups 1 and 2 in Figure 1a: blue, 48% 3-year and 45% 6-year survival, n=55 and patients in groups 3 and 4 in Figure 1a: red, 76% 3-year and 58% 6-year survival, n=56 whilst the group of patients with inconsistent CT features in Figure 1a (yellow) remained unchanged. All three groups were compared to a population of IPF patients: black, 42% 3-year and 18% 6-year survival, n=284.

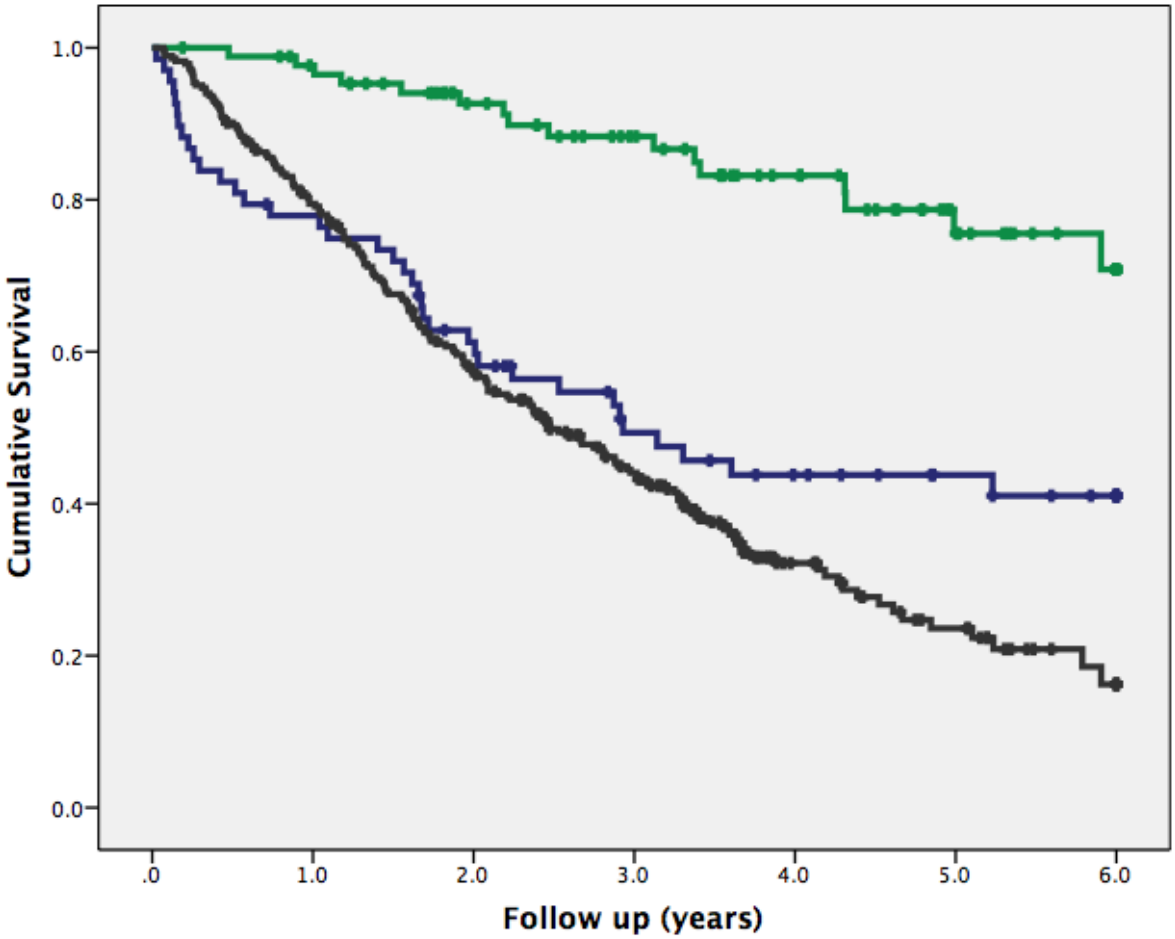


Figure 3c. Kaplan Meier curves demonstrating mortality for patients with rheumatoid arthritis-related interstitial lung disease (RAILD) subdivided according to the Scleroderma System into good-outcome (green) and bad-outcome (blue) groups, and compared with an idiopathic pulmonary fibrosis (IPF) cohort (black). Patients with >25% ILD extent on CT were given a score of 1 (blue), whilst patients with <15% ILD extent on CT were given a score of 0 (green). Patients with between 15-25% ILD extent on CT, were adjudicated on the basis of FVC, with an FVC >70% predicted accorded a score of 0, whilst an FVC <70% predicted was accorded a score of 1. Good-outcome Scleroderma System (green) 88% 3-year and 70% 6-year survival, n=88; bad-outcome Scleroderma System (blue) 50% 3-year and 40% 6-year survival, n=68. IPF patients: black, 42% 3-year and 18% 6-year survival, n=284.

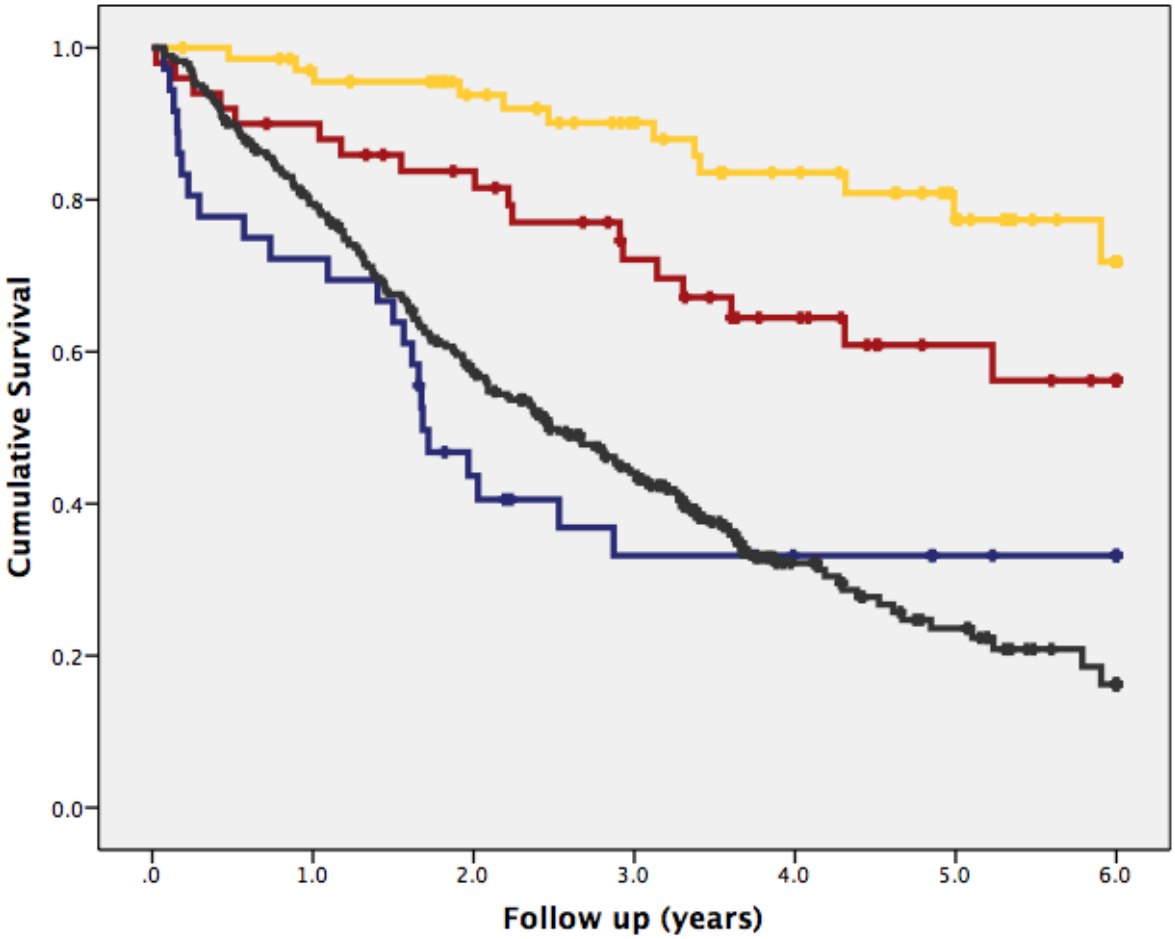


Figure 3d. Kaplan Meier curves demonstrating mortality for patients with rheumatoid arthritis-related interstitial lung disease (RAILD) categorized using the Scleroderma System and a definite usual interstitial pneumonia (UIP) pattern on CT (disease distribution and mosaicism were not considered as being 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: blue, 35% 3-year and 35% 6-year survival, n=36; either Scleroderma System bad-outcome OR definite UIP pattern: red, 78% 3-year and 58% 6-year survival, n=50; Scleroderma System good-outcome without definite UIP: yellow, 90% 3-year and 78% 6-year survival, n=70. IPF patients: black, 42% 3-year and 18% 6-year survival, n=284.

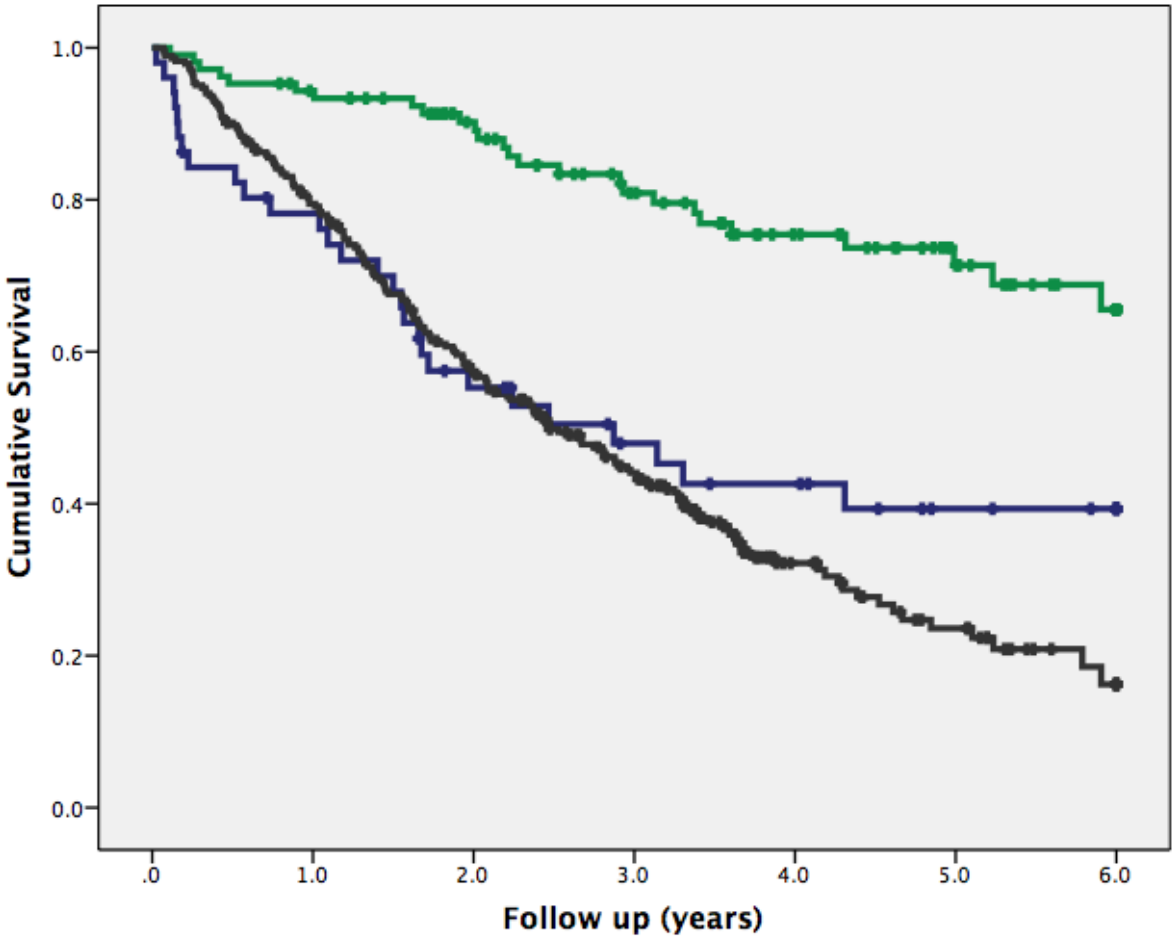


Figure 3e. Kaplan Meier curves demonstrating mortality for patients with rheumatoid arthritis-related interstitial lung disease (RAILD) categorized using a CALIPER vessel-related structure (VRS) score of 4.4%. VRS score >4.4% (blue) 45% 3-year and 40% 6-year survival, n=51; VRS score <4.4% (green) 82% 3-year and 70% 6-year survival, n=106. IPF patients: black, 42% 3-year and 18% 6-year survival, n=284.

Supplementary Appendix

Pulmonary Function Protocol

Pulmonary function tests (PFT) were analysed if performed within 3 months of the corresponding CT scan according to established protocols (25). Spirometry (Jaeger Master screen PFT, Carefusion Ltd., Warwick, UK and Houten, NL, Alpha Spirometer; Vitalograph, Buckingham, UK), plethysmographic lung volumes (Jaeger Master screen Body, Carefusion Ltd., Warwick, UK and Houten, NL), and diffusion capacity for carbon monoxide (Jaeger Master screen PFT, Carefusion Ltd., Warwick. UK and Houten, NL). Parameters assessed: forced expiratory volume in one second (FEV1), forced vital capacity (FVC) and single breath carbon monoxide diffusing capacity corrected for hemoglobin concentration (DLco). The composite physiologic index (CPI) was calculated using the formula: 91·0 - (0·65 x % predicted DLco) - (0·53 x % predicted FVC) + (0·34 x % predicted FEV1). (26)

CT Protocol

CT scans at the Royal Brompton Hospital were obtained using a 64-slice multiple detector CT scanner (Somatom Sensation 64; Siemens, Erlangen, Germany) or a 4-slice multiple-detector CT scanner (Siemens Volume Zoom; Siemens, Erlangen, Germany). All images were reconstructed using a high spatial frequency, B70 kernel (Siemens, Munich, Germany). CT scans at the St. Antonius Hospital, Nieuwegein, were obtained using a 64-slice multiple detector CT scanner (Phillips Brilliance 64; Cleveland, Ohio, USA) or a 256-slice multiple-detector iCT scanner (Phillips, Cleveland, Ohio, USA). All images were reconstructed using C, EC, L or YC kernels (Phillips, Cleveland, Ohio, USA). CT images at Edinburgh Royal Infirmary were acquired using a high-resolution CT (HRCT) scanner (Aquilion One [Toshiba Medical Systems, Toshiba Corp] or Somatom Volume Zoom [Siemens AG]). The following kernals

were used for the CT imaging reconstructions: Siemens (B20f, B60f, B60s or I70f\3), General Electric (Bone, Lung) and Toshiba (FC03, FC07, FC12 and FC51). The Siemens B80 algorithm was markedly edge-enhancing and resulted in misclassification of parenchymal features such as honeycombing by CALIPER. As a result, CT imaging reconstructed with the Siemens B80 algorithm (St Antonius n=6, Edinburgh Royal Infirmary n=3) were excluded from the study.

All patients were scanned from lung apices to bases, at full inspiration, using a peak voltage of 120 kVp with tube current modulation (range, 30 to 140 mA). Images of 0.5-2mm thickness were viewed at window settings optimized for the assessment of the lung parenchyma (width 1500 HU; level -500 HU).

Variable	Royal Brompton Cohort (n=90)	Edinburgh Royal Infirmary Cohort (n=67)	Cohort differences
Median Age (range)	64 (38-85)	67 (30-85)	0.03^
Male/female	43/47	28/39	0.46*
Survival (alive/dead)	48/42	55/12	0.0002*
Never smokers/ever-smokers (n=230)	27/57	18/48	0.52*
Honeycombing presence (N/Y)	48/42	50/17	0.006*
FEV1 % predicted	70.2 ± 19.5 (87)	91.5 ± 18.5 (67)	<0.0001
FVC % predicted	74.8 ± 22.9 (87)	97.0 ± 19.4 (67)	<0.0001
FEV1/FVC % predicted	96.3 ± 16.2 (83)	94.7 ± 9.8 (67)	0.50
DLco % predicted	40.9 ± 15.8 (84)	58.7 ± 15.1 (47)	<0.0001
CPI	48.6 ± 15.8 (81)	31.9 ± 12.1 (47)	<0.0001
Visual CT scores (%)			
Total ILD extent	31.3 ± 19.0	16.7 ± 8.5	<0.0001
Emphysema extent	7.9 ± 13.4	7.5 ± 10.2	0.86
Honeycombing extent	5.2 ± 9.3	0.4 ± 1.0	0.00004
TxBx severity (max score 18)	8.7 ± 4.0	6.4 ± 2.5	0.0001
CALIPER CT scores (%)			
Total ILD extent	18.0 ± 16.9	12.5 ± 14.1	0.03
Emphysema extent	2.2 ± 7.8	3.0 ± 5.3	0.45
Honeycombing extent	1.2 ± 2.7	0.4 ± 0.7	0.02
Vessel-related structures	4.5 ± 1.8	3.4 ± 1.5	0.0001

Supplementary Table 1. Patient age, gender and mean and standard deviations of pulmonary function indices and visually scored CT parameters in patients with rheumatoid arthritis-related interstitial lung disease subdivided according to study centre. Data represent mean values with standard deviations. Comparisons were made with the Students t-test, unless indicated: ^=Mann Whitey U test, *=Chi-squared test. FEV1=forced expiratory volume in one second, FVC=forced vital capacity, DLco=diffusing capacity for carbon monoxide, CPI=composite physiological index, ILD=interstitial lung disease, TxBx=traction bronchiectasis.

Visual CT Variable	Single determination standard deviation
CT Interstitial lung disease extent	3.95
CT Ground glass opacity	3.49
CT Reticular pattern	4.01
CT Honeycombing	2.08
CT Total emphysema	3.47
CT Traction bronchiectasis	1.74

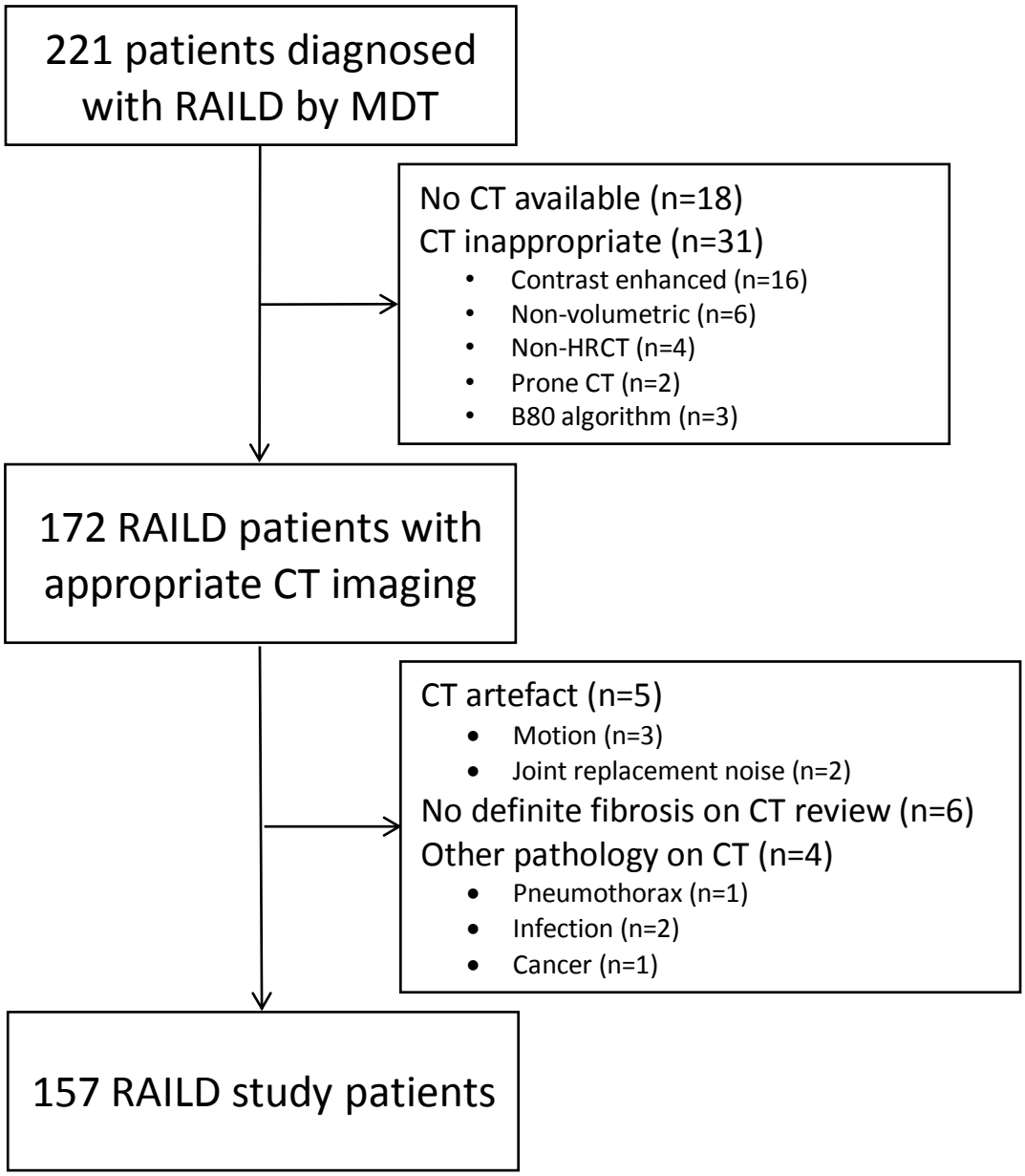
Supplementary Table 2. Variation in visual scores of CT parenchymal patterns between the two scorers for the rheumatoid arthritis-related interstitial lung disease patients, calculated using the single determination standard deviation. CT = computed tomography.

Variable	RAILD Cohort (n=157)	IPF Cohort (n=284)	Cohort differences
Median Age (range)	65 (30-85)	69 (37-92)	0.004^
Male/female	71/86	225/59	<0.0001*
Survival (alive/dead)	103/54	95/189	<0.0001
FVC % predicted	84.4 ± 24.1 (154)	74.2 ± 19.5 (283)	<0.0001
Visual ILD extent	25.1 ± 17.0	29.1 ± 12.1	0.01
Visual TxBx severity (max score 18)	7.7 ± 3.6	10.2 ± 3.2	<0.0001
CALIPER ILD extent	15.7 ± 16.0	23.5 ± 16.3	<0.0001
CALIPER Vessel-related structures	4.0 ± 1.8	5.3 ± 1.7	<0.0001

Supplementary Table 3. Patient age, gender and mean and standard deviations of forced vital capacity (FVC) and visually and CALIPER scored CT parameters in patients with rheumatoid arthritis-related interstitial lung disease (RAILD, Column 1), and idiopathic pulmonary fibrosis (IPF, Column 2). Data represent mean values with standard deviations. Comparisons were made with the Students t-test, unless indicated: ^=Mann Whitey U test, *=Chi-squared test. FVC=forced vital capacity, ILD=interstitial lung disease, TxBx=traction bronchiectasis.

Variable	Hazard ratio	95.0% Confidence Interval		P Value
		Lower	Upper	
Age	1.05	1.02	1.08	0.001
Male Gender	1.09	0.61	1.95	0.78
Smoking (never vs ever)	2.97	1.45	6.09	0.003
Scleroderma System	3.80	1.90	7.63	0.0002
4.4% VRS Threshold	2.43	1.28	4.60	0.007
Age	1.03	1.00	1.06	0.06
Male Gender	1.29	0.72	2.28	0.39
Smoking (never vs ever)	2.05	1.02	4.10	0.04
Fleischner System	1.65	1.14	2.40	0.008
4.4% VRS Threshold	3.93	2.21	7.02	<0.0001
Age	1.04	1.01	1.07	0.008
Male Gender	1.08	0.60	1.94	0.80
Smoking (never vs ever)	2.69	1.32	5.51	0.007
Progressive Fibrotic System	2.19	1.45	3.31	0.0002
4.4% VRS Threshold	2.57	1.36	4.87	0.004

Supplementary Table 4. Multivariable Cox regression analysis demonstrating mortality in patients with rheumatoid arthritis-related interstitial lung disease. Three staging systems (Scleroderma System, Fleischner System and Progressive Fibrotic System) were compared to the CALIPER vessel-related threshold (VRS) of 4.4% of the lung, after adjustment for patient age, gender and smoking status (never versus ever).



Supplementary Figure 1. CONSORT diagram illustrating the selection of rheumatoid arthritis-related interstitial lung disease (RAILD) patients for the study. CT = computed tomography.