TITLE: **Serial CT analysis in idiopathic pulmonary fibrosis: comparison of visual features that determine patient outcome**

Joseph Jacob1,2, Leon Aksman2, Alex J. Procter3, Bahareh Gholipour3, Gary Cross4, Joseph Barnett5, Christopher Brereton6, Mark G. Jones6, Coline H.M. van Moorsel7,8,Hendrik W. van Es9, Frouke T. van Beek7,Marcel Veltkamp7, Sujal Desai5,Eoin P. Judge9, Teresa Burd10, Maria Kokosi11, Recep Savas12, Selen Bayraktaroglu12, Nesrin Mogulkoc13, Andre Altmann2, Athol U. Wells11

1Department of Respiratory Medicine, University College London

2Centre for Medical Image Computing, University College London

3Department of Radiology, University College Hospital, London

4Department of Radiology, Royal Free Hospital, London

5Department of Radiology, Royal Brompton Hospital, London

6National Institute for Health Research Biomedical Research Centre and Clinical and Experimental Sciences, University of Southampton, Southampton, United Kingdom

7St Antonius ILD Center of Excellence, Department of Pulmonology, St. Antonius Hospital, Nieuwegein, The Netherlands

8Division of Heart and Lungs, University Medical Center Utrecht, The Netherlands

9Department of Respiratory Medicine, Aintree University Hospital, Liverpool

10Department of Radiology, St Georges Hospital, London

11Department of Respiratory Medicine, Royal Brompton Hospital, London

12Department of Radiology, Ege University Hospital, Ismir, Turkey

13Department of Respiratory Medicine, Ege University Hospital, Ismir, Turkey

Corresponding Author: Joseph Jacob [j.jacob@ucl.ac.uk](mailto:j.jacob@ucl.ac.uk) Tel:07511033666

**Authors contributions**

JJ, LA, AJP, BG, GC, JB, CB, MGJ, FTvB, CHMM, MV, TB, HWvE, SD, EPJ, MK, RS, SB, NM, AA, AUW were involved in either the acquisition, or analysis or interpretation of data for the study.

JJ, LA, AA and AUW were also involved in the conception and design of 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 all aspects of the work and 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 Southampton Hospital, Ege University Hospital, Ismir, Turkey and St. Antonius Hospital, Nieuwegein, Netherlands. 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).

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

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

Work by CHMM, HWE, FTB, MV was supported by ZonMW TopZorg Care grant 842002001.

**Acknowledgements**

Dr Jacob was supported by a Wellcome Trust Clinical Research Career Development Fellowship209553/Z/17/Z.

Dr Altmann holds an MRC eMedLab Medical Bioinformatics Career Development Fellowship. This work was supported by the Medical Research Council (grant number MR/L016311/1).

Dr Brereton and Dr Jones were supported by the National Institute for Health Research Biomedical Research Centre at the University of Southampton.

**Word Count**: 2925

**ABSTRACT**

**Aims:** Idiopathic pulmonary fibrosis (IPF) patients receiving antifibrotic medication and patients with non-IPF fibrosing lung disease often demonstrate rates of annualised forced vital capacity (FVC) decline within the range of measurement variation (5.0-9.9%). We examined whether change in visual CT variables could help confirm whether marginal FVC declines represented genuine clinical deterioration rather than measurement noise.

**Methods:** In two IPF cohorts (test: n=103; validation: n=108), separate pairs of radiologists scored paired volumetric CTs (acquired between 6-24 months from baseline). Change in interstitial lung disease (ILD), honeycombing, reticulation, ground glass opacity extents and traction bronchiectasis severity was evaluated using a five-point scale, with mortality prediction analysed using univariable and multivariable Cox regression analyses. Both IPF populations were then combined to determine whether change in CT variables could predict mortality in patients with marginal FVC declines.

**Results:** On univariate analysis, change in all CT variables except ground glass opacity predicted mortality in both cohorts. On multivariate analysis adjusted for patient age, gender, antifibrotic use and baseline disease severity (diffusing capacity for carbon monoxide), change in traction bronchiectasis severity predicted mortality independent of FVC decline in both cohorts. Change in traction bronchiectasis severity was associated with good inter-observer agreement amongst both scorer pairs, independently predicted mortality in patients with marginal FVC declines and identified more patients with genuine deterioration than change in honeycombing extent.

**Conclusions:** Change in traction bronchiectasis severity is a sensitive measure of disease progression that could be used to adjudicate marginal FVC declines.

* **What is the key question?**

In IPF patients with marginal annualised FVC declines (5.0-9.9%), can visual evaluation of serial CT scans distinguish genuine clinical deterioration from measurement inaccuracy?

* **What is the bottom line?**

Change in traction bronchiectasis was shown to independently predict mortality in IPF patients with marginal FVC declines. As well as being a sensitive measure of disease progression, it was associated with good interobserver agreement in two separate scorer pairs.

* **Why read on?**

The use of antifibrotic medication has resulted in larger proportions of IPF patients having annualised rates of FVC decline that lie within the range of measurement noise. In non-IPF fibrosing lung diseases, patients with similar marginal FVC declines are not uncommon. Yet determining whether marginal declines reflect measurement variation or true clinical deterioration will be important for clinical management and for future non-inferiority drug trials. We demonstrate the utility of using a simple measure of change in traction bronchiectasis severity to help adjudicate marginal FVC declines and highlight its good reproducibility and sensitivity for identifying disease progression.

**ABBREVIATIONS**

CI Confidence interval

CPI Composite physiologic index

CT Computed tomography

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

TxBx Traction bronchiectasis

**INTRODUCTION**

The introduction of antifibrotic medication for the treatment of IPF has resulted in a reduction in the rate of forced vital capacity (FVC) decline(1, 2). Fewer patients will now undergo definitive declines of ≥10% of FVC and more patients will be seen with declines that lie within the range of measurement variation (between 5.0-9.9% annualised FVC decline). There are also trials underway examining the prognostic benefit of anti-fibrotic medication in non-IPF fibrosing lung diseases(3, 4) where patients are likely to undergo less dramatic declines than are seen in IPF. Knowing whether marginal declines in FVC values reflect measurement variability or genuine clinical deterioration is therefore going to be an increasingly challenging problem both in clinical practice and in future drug trials in fibrosing lung diseases.

CT analysis has been considered as a complimentary tool to FVC measurement whereby identification of worsening of disease on CT, could be used to confirm that a marginal FVC decline reflects clinical deterioration. To date however, the focus has been on quantitative CT analysis(5) which can be expensive and of limited availability. Our study chose to examine whether change in visual CT parameters could be used in the same way as quantitative tools in adjudicating marginal FVC declines. We examined CT parameters routinely evaluated by radiologists and examined change on a simple five-point scale. As well as global measures of CT pattern change, we examined whether lobar scores of change in interstitial lung disease (ILD) extent added more prognostic information than global ILD change measures.

**METHODS**

Study population

The discovery population consisted of patients diagnosed by a multidisciplinary team with IPF according to published guidelines (6), presenting to the Royal Brompton Hospital, London with longitudinal CT imaging performed between 2007 and 2014. The validation cohort comprised patients presenting to St Antonius Hospital, Utrecht (between 2004 and 2015), Ege Hospital Ismir, Turkey (between 2008 and 2015), and Southampton General Hospital, UK (between 2013 and 2015). Patients were included in the study if they had undergone two non-contrast, supine, volumetric thin section CT scans (maximum collimation of 2mm) within a 6-24-month time period. Approval for this study of clinically indicated CT and pulmonary function data at the Royal Brompton Hospital was obtained from the Liverpool Research Ethics Committee (Reference: 14/NW/0028), and the Institutional Ethics Committees of St Antonius Hospital, Ege University Hospital and Southampton General Hospital and informed patient consent was not required.

Patients were included in the study if they had undergone two non-contrast, supine, volumetric thin section CT scans within a 6-24-month time period. Pulmonary function tests (PFTs) evaluated included baseline forced expiratory volume in the first second, forced vital capacity (FVC), diffusion capacity for carbon monoxide (DLco), the composite physiologic index (CPI) and longitudinal FVC measurements.

Individual visual CT analysis

All baseline CTs were evaluated by a specialist thoracic radiologist (JJ) with 12 years imaging experience, and were classified according to the 2018 ATS/ERS/ALAT/JRS international consensus guidelines(7). Each CT in the discovery cohort was scored independently by two radiologists (GC, JB) with 3- and 4-years thoracic imaging experience respectively. Each CT in the validation cohort was scored independently by two radiologists (BG, AJP) with 3- and 5-years thoracic imaging experience respectively. Observers were blinded to all clinical information and the time points of the serial CTs. CT analysis involved interrogating images on dual monitor work-stations. CT patterns were classified according to the Fleischer Society glossary of terms(8) with the following modifications: areas of increased density lung with overlying reticulation or traction bronchiectasis were characterised as ground glass. Increased density lung with no overlying reticulation, representing pure ground glass opacity was not quantified as it was felt likely to represent inflammation rather than interstitial fibrosis.

Change in total ILD extent and total lung change in extents of ground glass opacity, reticular pattern and honeycombing and severity of traction bronchiectasis (Figure 1) were all scored on a categorical 5-point scale: 1=markedly improved, 2=slightly improved, 3=no change, 4=slightly worsened, 5= markedly worsened. CT pairs also had total ILD extent change scored on a lobar basis across six lobes (with the left middle lobe demarcated by the origin of the lingula bronchus). Consensus formulation for visual scores used the continuous learning method(9), whereby each CT was consensed by the scorers immediately after each CT read. CT change scores were adjusted to reflect the true timepoints of the CT pairs prior to statistical analysis. In a case where chronological randomisation meant that the second timepoint CT had been examined under the illusion that it was the first CT, and the first CT was thought to be the second timepoint CT, observer scores of disease improvement were changed to equivalent scores of disease worsening. For example a score of 1 or 2 was converted to a score of 5 or 4 respectively.

Statistical analysis

Data are given as medians or means with standard deviations, or numbers of patients with percentages where appropriate. Interobserver variation for visual scores was assessed using the quadratic-weighted Kappa statistic for categorical variables. Weighted kappa coefficients were categorized as follows: poor (0-0·20), fair (0·21-0·40), moderate (0·41-0·60), good (0·61-0·80) and excellent (0·81-1·00). The students t-test was used to measure mean differences between continuous variables, and the Chi-Squared test evaluated differences between categorical variables.

For both the discovery and validation cohorts the temporal trajectory of subjects’ FVC volumes was modelled using a linear mixed effects (LME) model with fixed effects of baseline age, sex (M/F), baseline antifibrotic use (never/ever) and study time, along with random intercepts and random slopes. We then estimated the change in FVC volume from baseline to follow-up CT measurement times by taking the difference in the LME models’ predicted FVC volume at these times.

In each cohort we performed a survival analysis for each of the eleven CT variables of interest using a separate multivariate Cox regression. Time was measured from the second CT and an event was either death or transplantation. We modelled survival as a function of the change in the CT variable with adjustment for the estimated change in FVC volume (estimated via the LME model), baseline age, sex, baseline DLco and baseline antifibrotic use. We adjusted the p-values of the CT variables’ hazard ratios for multiple comparisons by calculating the effective number of independent tests via the method of Li and Ji(10).

We also combined the discovery and validation cohorts to maximize the power of detecting whether CT derived change score predict mortality in three groups of subjects: those with an annualized FVC volume decline of at least 10% (definite FVC decline), those with decline between 5% and <10% (marginal FVC decline) and those with decline between -5% and <5% (FVC stability). We first estimated FVC volume trajectories for all 211 subjects using LME modelling with the same fixed and random effects as before. We estimated the change in FVC volume over the CT scanning period as well as the annualized FVC volume decline over this period. We then performed separate survival analyses for each CT variable for the 110 subjects with an annualized FVC volume decline of at least 10%, the 50 subjects with decline between 5% and <10% and the 36 subjects with decline between 0% and <5%, with multiple comparisons correction as before. Logistic regression analyses were used to identify relationships between a UIP pattern on baseline CT and annualized FVC decline thresholds. A p value threshold of ≤0.05 was considered significant.

**RESULTS**

Baseline Analyses

The discovery cohort comprised 103 IPF patients presenting to the Royal Brompton Hospital, London. The validation cohort comprised 108 patients: St Antonius Hospital, Utrecht (n=52), Ege Hospital Ismir, Turkey (n=46), and Southampton General Hospital, UK (n=10). 8 patients in the St Antonius cohort who underwent a lung transplant were censored at the date of transplantation. No patients were lost to follow up.

Patient age, gender and baseline lung function tests were similar between discovery and validation populations (Table 1). Significantly more patients in the validation cohort received antifibrotic medication than in the discovery cohort. A definite usual interstitial pneumonia pattern on CT was significantly more common in the discovery cohort than the validation cohort, which was also reflected in the increased mortality seen in the discovery cohort (Table 1). Patients in the validation cohort also had a slightly shorter interval between CTs.

Agreement between observers for change in CT variables was weakest for honeycombing but good/excellent for all other variables (Supplementary Table 2).

Longitudinal CT analyses

On univariable Cox regression analysis of the discovery (Table 2) and validation (Table 3) cohorts, all CT variables except ground glass opacity significantly predicted mortality. The same CT variables of change were the strongest predictors of mortality in both cohorts:

honeycombing extent, traction bronchiectasis severity, right middle lobe and right and left lower lobe ILD extents. When all the global lung CT variables (total ILD extent, ground glass opacity, reticulation, honeycombing and traction bronchiectasis) were examined together in a multivariable Cox regression model, change in ground glass opacity and honeycombing independently predicted mortality in the discovery cohort whilst change in traction bronchiectasis severity alone independently predicted mortality in the validation cohort.

Longitudinal CT and PFT models

In multivariable Cox regression models adjusted for patient age, gender, antifibrotic use and baseline DLco, the concordance indices were 0.64 in the discovery cohort and 0.67 in the validation cohort. When CT measures of change were separately added to this model (Table 4), the change scores that most powerfully predicted mortality were: traction bronchiectasis severity (discovery cohort: HR=2.15, p value=0.0005, CI=0.71; validation cohort: HR=2.33, p value=0.005, CI=0.71) and right middle lobe ILD extent (discovery cohort: HR=1.90, p value=0.003, CI=0.70; validation cohort: HR=2.12, p value=0.0005, CI=0.72). The results were maintained when baseline disease severity was evaluated using CPI (Supplementary Table 2) or FVC. In a final mortality model that additionally incorporated FVC decline, change in traction bronchiectasis severity predicted mortality independent of the degree of FVC decline in both cohorts, again with adjustment for patient age, gender, antifibrotic use and baseline DLco (Table 5).

Examination of FVC decline thresholds

Both cohorts were combined to evaluate patients with varying thresholds of annualised FVC decline (≥10%, n=110; 5.0-9.9%, n=50; -5.0-4.9%, n=47). Patients with ≥10% FVC decline had significantly more severe disease at baseline (assessed using baseline DLco and CPI) than patients with <10% annualised FVC decline (n=101). However, no significant difference in baseline disease severity (DLco, CPI) was seen between patients undergoing FVC declines of 5.0-9.9% or ≥10%. The presence of a definite UIP pattern (versus a probable UIP pattern) at baseline did not distinguish between patients undergoing an annualised FVC decline ≥10% or <10% and did not distinguish between patients undergoing an FVC decline ≥10% from patients undergoing an FVC decline between 5.0-9.9%.

In the patient group with the largest FVC declines (≥10%), change in honeycombing extent independently predicted mortality (Table 6). No CT variables were linked to mortality prediction in patients with the smallest FVC declines (-5.0-4.9%) and in this group of patients, none had any change in honeycombing identified by either scorer pair.

In patients with marginal FVC declines of 5.0-9.9%, change in traction bronchiectasis severity and change in honeycombing extent independently predicted mortality in separate models (Table 6). 5/50 (10%) patients with an FVC decline between 5.0-9.9%, had honeycombing change identified on CT pairs, whilst 12/50 (24%) patients were identified as having a change in traction bronchiectasis severity.

**DISCUSSION**

Our study specifically set out to examine whether change in visual CT variables across serial CT studies could be used to adjudicate marginal FVC declines of between 5.0-9.9% predicted. Marginal FVC changes are challenging to interpret in routine clinical practice as they could reflect either measurement noise or genuine physiological deterioration. We demonstrate that change in traction bronchiectasis severity, scored using a simple five-point categorical scale predicts mortality independent of patient treatment with antifibrotics and baseline disease severity in separate IPF cohorts. Importantly, in patients with FVC declines between 5.0-9.9% traction bronchiectasis severity change was a more sensitive measure of disease progression than honeycombing extent change and could therefore be used to determine whether the functional decline is clinically meaningful.

Several previous baseline studies in patients with idiopathic pulmonary fibrosis (11, 12) and other idiopathic and non-idiopathic fibrosing lung diseases (13-18) have underlined the prognostic value of visual estimations of traction bronchiectasis severity using categorical scores. Yet no previous studies have examined whether change in visual traction bronchiectasis scores can predict mortality in patients with IPF. Prior baseline studies examining traction bronchiectasis have highlighted its improved interobserver agreement when compared to scores of parenchymal pattern extents such as honeycombing. Our study, though examining longitudinal change on CTs, has demonstrated good-to-excellent agreement between observer pairs for traction bronchiectasis change, with agreement better than that seen for honeycombing change.

Change in traction bronchiectasis severity was the only CT variable that demonstrated a prognostic signal in both patient cohorts as well as in patients with indeterminate FVC declines. As awareness around the prognostic value of longitudinal CT analysis grows by the year, IPF patients are likely to undergo more frequent CT imaging necessitating a better understanding of longitudinal CT measures of deterioration. Change in traction bronchiectasis severity predicted mortality independent of patient therapy, suggesting that it could represent an important measure of disease worsening in drug trials. 24% of patients with marginal FVC declines were identified as exhibiting change in traction bronchiectasis severity. A sensitive measure of disease worsening will have particular importance in non-inferiority IPF trials where standard of care with antifibrotic therapy will result in a high proportion of patients undergoing marginal FVC declines.,

Our findings also have relevance in other fibrosing lung diseases. With elucidation of the progressive fibrotic phenotype(19), it is now acknowledged that identifying disease worsening over time may be just as important as reaching a specific diagnosis. Patients enrolled in the INBUILD trial(3), examining non-IPF progressive fibrotic conditions, are likely to undergo FVC declines of smaller magnitudes than those typically seen in patients with IPF. The INBUILD trial protocol requires patients undergoing FVC declines between 5.0-9.9% to demonstrate symptomatic worsening or worsening on CT imaging(3). Our findings demonstrate that progression on CT imaging can act as a surrogate for mortality in patients whose FVC declines when taken alone may not confidently suggest progression. Importantly, the mortality linkages identified using our CT variables indicate that they reflect disease progression (involvement of new lung tissue) rather than disease maturation (evolving changes in lung tissue already damaged). We also demonstrate that rather than evaluating total ILD extent progression as has been traditionally used, change in traction bronchiectasis might be the best of the existing suite of CT variables to examine. Reassuringly, our results were maintained in a validation cohort where the multicentred nature of the CT imaging analysed is similar to the range of image acquisitions captured in drug trials. Our positive findings suggest that potentially noisy and heterogenous imaging data can still prove prognostically valuable.

A major focus of CT image analysis in recent years has involved using computer tools to quantify lung damage, both at baseline and longitudinally. Volumetric computer analysis has the potential to afford a greater degree of precision and sensitivity for detecting lung damage when compared to more crude semiquantitative visual CT scores. Yet several challenges may constrain the uptake of computer technologies. Analytic tools employed on longitudinal CT imaging will need to take into account the noise associated with variations in CT acquisition parameters at baseline and between longitudinal scan pairs, as well as variation in patient respiratory effort across CT timepoints. Volumetric quantitation of lung damage evolution, as performed by computer tools, can also underestimate increases in the extent of damage in a disease such as IPF where affected lung shrinks in volume, and spared lung hyper expands in compensation. Over time, though ILD extent increases, a proportion of the involved lung shrinks, resulting in underestimation of disease worsening over time. Lastly, computer tools are often proprietary, and may consider unique variables that cannot be quantified by the human eye, making comparisons between clinicians, hospitals or countries challenging.

In contrast, longitudinal visual CT analysis is free, and to date has examined well-accepted CT variables that have been long defined by radiology glossaries(20). Whilst variation resulting from patient-related respiratory effort will still account for some measurement noise, careful selection of appropriate visual CT variables may negate problems with lung shrinkage that could underestimate disease progression. For example, traction bronchiectasis and honeycombing extent scores, constitute variables that represent larger proportions of damaged lung as disease worsens. Their linear relationship with disease progression may account for their sensitivity when compared to more traditionally examined global scores of worsening such as total ILD extent.

There were several limitations to the current study. As the study was retrospective, CT imaging was not performed at predefined regular intervals. Yet the mean interval between CTs was approximately 1 year in both cohorts which is within the bounds of follow up duration expected in IPF drug trials. Different scorer pairs evaluated the two cohorts and showed differences in agreement across several parenchymal patterns. Our a priori aim had been to examine real-world observer interpretations of CT images, using real-world clinical methods (side-by-side longitudinal CT examination) and not rely on adjudication by experienced sub-specialist radiologists. Accordingly, we believe our methods reflect a realistic interpretation of longitudinal CT images.

In conclusion our study has shown that change in traction bronchiectasis is a reliable and sensitive measure of disease progression in IPF. Change in traction bronchiectasis can be used to adjudicate indeterminate FVC declines, which are likely to be seen with increasing frequency in IPF patients receiving anti-fibrotics, and potentially in non-IPF patients with a progressive fibrotic phenotype.

 

 

 

Figure 1. Serial axial CT images in patients with idiopathic pulmonary fibrosis. In a 50-year-old male patient who did not receive antifibrotic medication and who demonstrated a >10% annualised FVC decline, images acquired 6-months apart (A) show change in traction bronchiectasis categorised as markedly worsened (Score=5) by scoring radiologists. In a 62-year-old male patient who received antifibrotic medication (B), with images were acquired 13-months apart, annualised FVC decline was between 5.0-9.9% and change in traction bronchiectasis was categorised as mildly worsened (Score=4). In a 77-year-old male who did not receive antifibrotic medication (C) and who had CTs acquired 15-months apart, change in traction bronchiectasis severity (Score=3) and annualised FVC decline (-5.0-4.9%) were both considered stable. Changes visible on the CT may reflect disease maturation rather than disease progression.

|  |  |  |  |
| --- | --- | --- | --- |
| **Variable** |  | **Discovery Cohort**  **(n=103)** | **Validation Cohort**  **(n=108)** |
| Median Age (years) |  | 67 (60-73) | 65 (58-71) |
| Male/female (numbers) |  | 85/18 | 80/28 |
| Alive/Dead (numbers) |  | 33/70 | 52/56\* |
| Mean CT time interval (years) |  | 1.1 ± 0.4 | 1.0 ± 0.4\* |
| Antifibrotic use (Y/N) |  | 33/70 | 74/34# |
| Definite UIP pattern on CT (Y/N) |  | 45/58 | 32/76\* |
| Baseline FVC (% predicted) |  | 74.3 ± 20.0 | 75.9 ± 19.6 |
| Baseline DLco (% predicted) |  | 41.3 ± 12.2 | 48.2 ± 14.4 |
| Baseline CPI (% predicted) |  | 50.6 ± 11.4 | 46.2 ± 12.3 |

Table 1. Patient age, gender and baseline values for various pulmonary function indices in the discovery and validation study cohorts of patients with idiopathic pulmonary fibrosis. Data represent mean values with standard deviations or medians with interquartile ranges unless otherwise indicated. FVC = forced vital capacity, DLco = diffusing capacity for carbon monoxide, CPI=composite physiologic index, UIP=usual interstitial pneumonia, \*p<0.05, #p<0.001.

|  |  |  |  |
| --- | --- | --- | --- |
| **Visual CT variable change** | **Hazard Ratio** | **95% Confidence Interval** | **P value** |
| Total ILD extent | 1.89 | 1.30-2.73 | 0.001 |
| Ground glass opacity extent | 1.91 | 1.30-2.79 | 0.001 |
| Reticular pattern extent | 1.54 | 1.07-2.22 | 0.02 |
| Honeycombing extent | 2.21 | 1.50-3.28 | 0.00007 |
| Traction Bronchiectasis severity | 2.06 | 1.40-3.06 | 0.0003 |
| Right upper lobe ILD extent | 1.69 | 1.20-2.40 | 0.003 |
| Left upper lobe ILD extent | 1.72 | 1.24-2.39 | 0.001 |
| Right middle lobe ILD extent | 2.02 | 1.40-2.92 | 0.0002 |
| Left middle lobe ILD extent | 1.60 | 1.14-2.23 | 0.006 |
| Right lower lobe ILD extent | 1.99 | 1.41-2.82 | 0.0001 |
| Left lower lobe ILD extent | 1.90 | 1.37-2.64 | 0.0001 |

Table 2. Univariable Cox regression analysis in the discovery cohort (n=103) demonstrating mortality prediction determined by change in various visual CT variables. ILD=interstitial lung disease.

|  |  |  |  |
| --- | --- | --- | --- |
| **Visual CT variable change** | **Hazard Ratio** | **95% Confidence Interval** | **P value** |
| Total ILD extent | 1.60 | 1.18-2.17 | 0.002 |
| Ground glass opacity extent | 1.40 | 1.00-1.97 | 0.054 |
| Reticular pattern extent | 1.68 | 1.17-2.42 | 0.005 |
| Honeycombing extent | 1.87 | 1.27-2.76 | 0.002 |
| Traction Bronchiectasis severity | 2.61 | 1.60-4.28 | 0.0001 |
| Right upper lobe ILD extent | 1.52 | 1.03-2.24 | 0.04 |
| Left upper lobe ILD extent | 1.50 | 1.03-2.18 | 0.03 |
| Right middle lobe ILD extent | 1.98 | 1.37-2.84 | 0.0002 |
| Left middle lobe ILD extent | 1.49 | 1.03-2.16 | 0.04 |
| Right lower lobe ILD extent | 1.56 | 1.14-2.14 | 0.006 |
| Left lower lobe ILD extent | 1.58 | 1.15-2.16 | 0.004 |

Table 3. Univariable Cox regression analysis in the validation cohort (n=108) demonstrating mortality prediction determined by change in various visual CT variables. ILD=interstitial lung disease.

|  |  |  |
| --- | --- | --- |
| **Categorical change in**  **visual CT variables** | **Discovery Cohort**  **(HR, p value, CI)** | **Validation Cohort**  **(HR, p value, CI)** |
| Total ILD extent | 1.72, 0.02, 0.69 | 1. 67, 0.005, 0.72 |
| Ground glass opacity extent | 1.86, 0.005, 0.69 | 1.59, 0.03, 0.70 |
| Reticular pattern extent | 1.49, 0.15, 0.68 | 1.96, 0.001, 0.73 |
| Honeycombing extent | 2.03, 0.003, 0.69 | 1.84, 0.02, 0.71 |
| Traction Bronchiectasis severity | 2.15, 0.0005, 0.71 | 2.33, 0.005, 0.71 |
| Right upper lobe ILD extent | 1.58, 0.05, 0.67 | 1.65, 0.05, 0.70 |
| Left upper lobe ILD extent | 1.62, 0.02, 0.69 | 1.60, 0.05, 0.70 |
| Right middle lobe ILD extent | 1.90, 0.003, 0.70 | 2.12, 0.0005, 0.72 |
| Left middle lobe ILD extent | 1.54, 0.05, 0.68 | 1.56, 0.05, 0.70 |
| Right lower lobe ILD extent | 1.79, 0.003, 0.71 | 1.59, 0.01, 0.73 |
| Left lower lobe ILD extent | 1.86, 0.001, 0.72 | 1.59, 0.01, 0.72 |

Table 4. Multivariable Cox regression analyses models demonstrating mortality prediction determined by change in various visual CT variables in separate discovery (n=103) and validation (n=108) cohorts of patients with idiopathic pulmonary fibrosis. Each visual CT variable was analysed in a separate model adjusted for patient age, gender, baseline disease severity using the diffusion capacity for carbon monoxide and antifibrotic use (never/ever). HR=Hazard ratio, CI=concordance index, ILD=interstitial lung disease. P values shown are adjusted for multiple comparisons.

|  |  |  |
| --- | --- | --- |
| **Categorical change in**  **visual CT variables** | **Discovery Cohort**  **(HR, p value, CI)** | **Validation Cohort**  **(HR, p value, CI)** |
| Total ILD extent | 1.62, 0.09, 0.66 | 1.38, 0.33, 0.70 |
| Ground glass opacity extent | 1.78, 0.02, 0.65 | 1.36, 0.42, 0.70 |
| Reticular pattern extent | 1.37, 0.62, 0.64 | 1.65, 0.09, 0.71 |
| Honeycombing extent | 1.91, 0.045, 0.66 | 1.37, 1.00, 0.70 |
| Traction Bronchiectasis severity | 2.08, 0.003, 0.68 | 2.08, 0.02, 0.72 |
| Right upper lobe ILD extent | 1.48, 0.25, 0.64 | 1.18, 1.00, 0.69 |
| Left upper lobe ILD extent | 1.55, 0.12, 0.66 | 1.20, 1.00, 0.69 |
| Right middle lobe ILD extent | 1.81, 0.01, 0.66 | 1.68, 0.13, 0.71 |
| Left middle lobe ILD extent | 1.43, 0.37, 0.65 | 1.21, 1.00, 0.69 |
| Right lower lobe ILD extent | 1.72, 0.01, 0.68 | 1.29, 0.70, 0.70 |
| Left lower lobe ILD extent | 1.87, 0.007, 0.68 | 1.29, 0.69, 0.70 |

Table 5. Multivariable Cox regression analyses models demonstrating mortality prediction determined by change in various visual CT variables in separate discovery (n=103) and validation (n=108) cohorts of patients with idiopathic pulmonary fibrosis. Each visual CT variable was analysed in a separate model adjusted for patient age, gender, baseline disease severity using the diffusion capacity for carbon monoxide, antifibrotic use (never/ever) and forced vital capacity decline calculated using mixed effects models. HR=Hazard ratio, CI=concordance index, ILD=interstitial lung disease. P values shown are adjusted for multiple comparisons.

|  |  |  |
| --- | --- | --- |
| **Categorical change in**  **visual CT variables** | **FVC 5.0-9.9% decline**  **(HR, p value, CI)** | **FVC≥10% decline**  **(HR, p value, CI)** |
| Total ILD extent | 2.30, 0.20, 0.58 | 1.42, 0.18, 0.55 |
| Ground glass opacity extent | 1.89, 0.74, 0.51 | 1.28, 0.57, 0.54 |
| Reticular pattern extent | 2.27, 0.27, 0.57 | 1.10, 1.00, 0.52 |
| Honeycombing extent | 3.89, 0.02, 0.58 | 1.88, 0.005, 0.58 |
| Traction Bronchiectasis severity | 3.20, 0.02, 0.59 | 1.29, 0.88, 0.52 |
| Right upper lobe ILD extent | 2.03, 0.33, 0.53 | 0.96, 1.00, 0.52 |
| Left upper lobe ILD extent | 3.05, 0.02, 0.60 | 1.02, 1.00, 0.52 |
| Right middle lobe ILD extent | 2.58, 0.10, 0.57 | 1.31, 0.60, 0.54 |
| Left middle lobe ILD extent | 2.31, 0.11, 0.57 | 1.03, 1.00, 0.52 |
| Right lower lobe ILD extent | 1.85, 0.71, 0.51 | 1.29, 0.46, 0.54 |
| Left lower lobe ILD extent | 1.44, 1.00, 0.50 | 1.41, 0.11, 0.55 |

Table 6. Various visual CT variables were examined in separate multivariable Cox regression models adjusted for patient age, gender, baseline disease severity (using diffusion capacity for carbon monoxide) and antifibrotic use (never/ever). Analyses were performed in patients with an annualised forced vital capacity decline of 5.0-9.9% (n=50) and ≥10% (n=110). HR=Hazard ratio, CI=concordance index, ILD=interstitial lung disease. P values shown are adjusted for multiple comparisons.

Supplementary Appendices

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Visual CT Variable** |  | **Discovery Cohort** | **Validation Cohort** | |
| Interstitial lung disease change |  | 0.87 | | 0.87 |
| Ground glass opacity change |  | 0.79 | | 0.65 |
| Reticular pattern change |  | 0.83 | | 0.80 |
| Honeycombing change |  | 0.64 | | 0.70 |
| Traction bronchiectasis change |  | 0.81 | | 0.66 |
| Right upper lobe ILD extent change |  | 0.87 | | 0.83 |
| Left upper lobe ILD extent change |  | 0.83 | | 0.77 |
| Right middle lobe ILD extent change |  | 0.87 | | 0.71 |
| Left middle lobe ILD extent change |  | 0.83 | | 0.75 |
| Right lower lobe ILD extent change |  | 0.83 | | 0.77 |
| Left lower lobe ILD extent change |  | 0.83 | | 0.80 |

Supplementary Table 1. Weighted Kappa measurements indicating variation in visual scores of disease change on CT between the pairs of radiologist scorers. Separate pairs of scorers evaluated CTs of idiopathic pulmonary fibrosis patients in the discovery and validation cohorts. ILD =interstitial lung disease.

|  |  |  |
| --- | --- | --- |
| **Categorical change in visual CT variable** | **Discovery Cohort**  **(HR, p value, CI)** | **Validation Cohort**  **(HR, p value, CI)** |
| Total ILD extent | 1.69, 0.005, 0.70 | 1. 67, 0.0009, 0.72 |
| Ground glass opacity extent | 1.88, 0.001, 0.69 | 1.50, 0.01, 0.70 |
| Reticular pattern extent | 1.48, 0.03, 0.69 | 2.04, 0.0002, 0.73 |
| Honeycombing extent | 1.97, 0.0009, 0.70 | 1.91, 0.002, 0.70 |
| Traction Bronchiectasis severity | 2.08, 0.0002, 0.73 | 2.40, 0.0009, 0.70 |
| Right upper lobe ILD extent | 1.53, 0.02, 0.68 | 1.67, 0.01, 0.69 |
| Left upper lobe ILD extent | 1.55, 0.009, 0.70 | 1.63, 0.008, 0.70 |
| Right middle lobe ILD extent | 1.83, 0.001, 0.71 | 2.10, 0.0001, 0.72 |
| Left middle lobe ILD extent | 1.47, 0.03, 0.69 | 1.56, 0.02, 0.70 |
| Right lower lobe ILD extent | 1.70, 0.002, 0.72 | 1.60, 0.002, 0.72 |
| Left lower lobe ILD extent | 1.89, 0.0001, 0.73 | 1.61, 0.002, 0.71 |

Supplementary Table 2. Multivariable Cox regression analyses models demonstrating mortality prediction determined by change in various visual CT variables in separate discovery (n=103) and validation (n=108) cohorts of patients with idiopathic pulmonary fibrosis. Each visual CT variable was analysed in a separate model adjusted for patient age, gender, baseline disease severity using the composite physiologic index and antifibrotic use (never/ever). HR=Hazard ratio, CI=concordance index, ILD=interstitial lung disease.

**REFERENCES**

1. Richeldi L, du Bois RM, Raghu G, Azuma A, Brown KK, Costabel U, et al. Efficacy and Safety of Nintedanib in Idiopathic Pulmonary Fibrosis. N Engl J Med. 2014;370(22):2071-82.

2. King TE, Bradford WZ, Castro-Bernardini S, Fagan EA, Glaspole I, Glassberg MK, et al. A Phase 3 Trial of Pirfenidone in Patients with Idiopathic Pulmonary Fibrosis. N Engl J Med. 2014;370(22):2083-92.

3. Flaherty KR, Brown KK, Wells AU, Clerisme-Beaty E, Collard HR, Cottin V, et al. Design of the PF-ILD trial: a double-blind, randomised, placebo-controlled phase III trial of nintedanib in patients with progressive fibrosing interstitial lung disease. BMJ Open Respiratory Research. 2017;4(1):e000212.

4. Maher TM, Corte TJ, Fischer A, Kreuter M, Lederer DJ, Molina-Molina M, et al. Pirfenidone in patients with unclassifiable progressive fibrosing interstitial lung disease: design of a double-blind, randomised, placebo-controlled phase II trial. BMJ Open Respiratory Research. 2018;5(1):e000289.

5. Jacob J, Bartholmai BJ, van Moorsel CHM, Rajagopalan S, Devaraj A, van Es HW, et al. Longitudinal prediction of outcome in idiopathic pulmonary fibrosis using automated CT analysis. Eur Respir J. 2019:1802341.

6. 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(6):788-824.

7. Raghu G, Remy-Jardin M, Myers JL, Richeldi L, Ryerson CJ, Lederer DJ, et al. Diagnosis of Idiopathic Pulmonary Fibrosis. An Official ATS/ERS/JRS/ALAT Clinical Practice Guideline. Am J Respir Crit Care Med. 2018;198(5):e44-e68.

8. Hansell DM, Bankier AA, MacMahon H, McLoud TC, Müller NL, Remy J. Fleischner Society: glossary of terms for thoracic imaging. Radiology. 2008;246:697.

9. Sverzellati N, Devaraj A, Desai SR, Quigley M, Wells AU, Hansell DM. Method for Minimizing Observer Variation for the Quantitation of High-Resolution Computed Tomographic Signs of Lung Disease. J Comput Assist Tomogr. 2011;35(5):596-601.

10. Li J, Ji L. Adjusting multiple testing in multilocus analyses using the eigenvalues of a correlation matrix. Heredity (Edinb). 2005;95(3):221-7.

11. Sumikawa H, Johkoh T, Colby TV, Ichikado K, Suga M, Taniguchi H, et al. Computed tomography findings in pathological usual interstitial pneumonia: relationship to survival. Am J Respir Crit Care Med. 2008;177(4):433-9.

12. 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.

13. Walsh SL, Sverzellati N, Devaraj A. Chronic hypersensitivity pneumonitis: high resolution computed tomography patterns and pulmonary function indices as prognostic determinants. Eur Radiol. 2012;22.

14. Walsh SL, Sverzellati N, Devaraj A, Keir GJ, Wells AU, Hansell DM. Connective tissue disease related fibrotic lung disease: high resolution computed tomographic and pulmonary function indices as prognostic determinants. Thorax. 2014;69(3):216-22.

15. Jacob J, Bartholmai BJ, Egashira R, Brun AL, Rajagopalan S, Karwoski R, et al. Chronic hypersensitivity pneumonitis: identification of key prognostic determinants using automated CT analysis. BMC Pulm Med. 2017;17(1):81.

16. Jacob J, Bartholmai BJ, Rajagopalan S, Egashira R, Brun AL, Kokosi M, et al. Unclassifiable-interstitial lung disease: Outcome prediction using CT and functional indices. Respir Med. 2017;130:43-51.

17. Jacob J, Bartholmai BJ, Rajagopalan S, Brun AL, Egashira R, Karwoski R, et al. Evaluation of computer-based computer tomography stratification against outcome models in connective tissue disease-related interstitial lung disease: a patient outcome study. BMC Med. 2016;14:190.

18. Jacob J, Hirani N, van Moorsel CHM, Rajagopalan S, Murchison JT, van Es HW, et al. Predicting outcomes in rheumatoid arthritis related interstitial lung disease. Eur Respir J. 2018:1800869.

19. Wells AU, Brown KK, Flaherty KR, Kolb M, Thannickal Victor J. What's in a name? That which we call IPF, by any other name would act the same. Eur Respir J. 2018;51(5):1800692.

20. Hansell DM, Bankier AA, MacMahon H, McLoud TC, Müller NL, Remy J. Fleischner Society: glossary of terms for thoracic imaging. Radiology. 2008;246(3):697-722.