Title: Predicting outcomes in idiopathic pulmonary fibrosis using automated CT analysis

Authors: Joseph Jacob[1*], Brian J. Bartholmai[1], Srinivasan Rajagopalan[1], Coline H.M. van Moorsel[2,3], Hendrik W. van Es[4], Frouke T. van Beek[2], Marjolijn H.L. Struik[2,3], Maria Kokosi[5], Ryoko Egashira[6], Anne Laure Brun[7], Arjun Nair[8], Simon L.F. Walsh[9], Gary Cross[10], Joseph Barnett[10], Angelo de Lauretis[11], Eoin P. Judge[5], Sujal, Desai[12], Ronald Karwoski[13], Sebastien Ourselin[14], Elisabetta Renzoni[5], Toby M. Maher[5], Andre Altmann[14], Athol U. Wells[6].

Affiliations: [1]Division of Radiology, Mayo Clinic Rochester, Minnesota, USA.

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

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

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

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

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

[7]Department of Radiology, Whittington Hospital, London.

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

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

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

[11]Division of Pneumology, "Guido Salvini" Hospital, Garbagnate Milanese, Italy

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

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

14Translational Imaging Group, Centre for Medical Image Computing, University

College London, London, UK

*Corresponding author: Joseph Jacob, 10 Wolsey road, Northwood, Middlesex, UK

HA6 2HW. Email: joseph.jacob@nhs.net

At a Glance: Quantification of CT parenchymal patterns in idiopathic pulmonary

fibrosis using computer tools has been suggested as a method that can improve on

mortality prediction using visual CT scoring. As computer technology advances, it has

now become possible to study CT parenchymal features that have no visual

correlate. Our study demonstrates that computer-derived vessel-related structure

scores can outperform current gold-standard measures of outcome in idiopathic

pulmonary fibrosis such as forced vital capacity decline. Specifically, we demonstrate

that using thresholds of computer-derived vessel-related structure scores for cohort

enrichment can identify idiopathic pulmonary fibrosis patients that respond to

antifibrotic medication with reduced FVC decline and improved survival.

Importantly, the vessel-related structure thresholds would be able to reduce

idiopathic pulmonary fibrosis drug-trial population sizes by 26% thereby dramatically

reducing study costs.

Authors contributions:

JJ, AA, FTvB, CHMM, MHLS, HWvE, SO, EPJ, AL, GC, JB, MK, RE, ALB, AN, SLFW, TMM, 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

Dr. Jacob reports personal fees from Boehringer Ingleheim outside the current work.

Abstract: Aims: To determine whether computer-derived computed tomography measures, specifically measures of pulmonary vessel-related structures, can better predict functional decline and survival in idiopathic pulmonary fibrosis (IPF) and reduce requisite sample sizes in drug trial populations.

Methods: IPF patients undergoing volumetric non-contrast CT imaging at the Royal Brompton Hospital, London and St Antonius Hospital, Utrecht, were examined to identify pulmonary functional measures (including forced vital capacity), and visual and computer-derived (CALIPER) CT features predictive of mortality and forced vital capacity decline. The discovery cohort constituted 247 consecutive patients with validation of results in a separate cohort of 284 patients all fulfilling drug trial entry criteria.

Results: In discovery and validation cohorts, CALIPER-derived features, particularly vessel-related structure (VRS) scores were amongst the strongest predictors of survival and forced vital capacity decline. CALIPER results were accentuated in patients with less extensive disease, outperforming pulmonary function measures. When used as a cohort enrichment tool, a CALIPER VRS score >4.4% of the lung was able to reduce the requisite sample size of an IPF drug trial by 26%.

Conclusions: Our study has validated a new quantitative CT measure in IPF patients fulfilling drug trial entry criteria, the VRS scores, that outperformed current gold-standard measures of outcome. When used for cohort enrichment in an IPF drug-trial setting, VRS threshold scores can reduce a required IPF drug trial population size by 25%, thereby limiting prohibitive trial costs. Importantly VRS scores identify patients in whom antifibrotic medication prolongs life and reduces forced vital capacity decline.

ABBREVIATIONS

ILD	Interstitial lung disease
CT	Computed tomography
PFT	Pulmonary function test
FEV1	Forced expiratory volume in one second
FVC	Forced vital capacity
DLco	Diffusing capacity for carbon monoxide
Kco	Carbon monoxide transfer coefficient
CPI	Composite physiologic index
GAP	Gender, age physiology
CALIPER	Computer-Aided Lung Informatics for Pathology Evaluation and Rating
HR	Hazard ratio
CI	Confidence interval
TxBx	Traction bronchiectasis
IPF	Idiopathic pulmonary fibrosis
CILD	CALIPER interstitial lung disease extent
CFibrosis	CALIPER fibrosis extent
CVRS	CALIPER vessel-related structure
CHC	CALIPER honeycombing extent
UZ	Upper zone
MZ	Middle zone
LZ	Lower zone
UZVRS	Upper zone vessel-related structure

MZVRS   Middle zone vessel-related structure

LZVRS   Lower zone vessel-related structure

VILD   Visual interstitial lung disease extent

VFibrosis   Visual fibrosis extent

VHC   Visual honeycombing extent

INTRODUCTION

Idiopathic pulmonary fibrosis (IPF) is a progressive fibrosing lung disease of increasing prevalence(1-5) that results in a patient's median survival being curtailed to only 3-5 years from diagnosis(1, 6-11). In a drug trial setting, a decline in forced vital capacity (FVC) has been used as the primary measure of outcome in IPF(12-17), and has been correlated with survival in several IPF cohorts(13, 15, 18-21). FVC testing is associated with a degree of measurement variation(22), and evidence from several clinical studies(15, 18-20, 23) confirms that a 10% FVC decline threshold is a surrogate marker of mortality, and in the absence of an alternative explanation, an indicator of genuine disease progression(11).

Given the inexorable decline associated with late-stage IPF, characterisation of indicators suggesting a poor prognosis at baseline in IPF is essential to guide optimal management and allow timely referral for lung transplantation(24). In IPF drug trial settings, cohort enrichment strategies aim to create more homogenous study populations and identify patients that are likely to experience increased clinical events(25) thereby reducing sample sizes and lessening the prohibitive costs associated with clinical trials(26).

Given the rapid recent advances in computer technology, the focus of imaging based disease biomarkers in IPF has primarily fallen on computer analysis of CT imaging which is without the constraints of interobserver variation associated with visual CT scoring. New computer algorithms utilising three-dimensional volumetric CT datasets, can quantify parenchymal pattern extents(27, 28) and have been shown to

better predict survival in various fibrosing lung diseases at baseline, than visual CT scores(29, 30). Computer analysis also has the potential to uncover CT features hitherto under-recognised by visual CT analysis that predict mortality in IPF. The quantitation of pulmonary vessels (arteries and veins) and vessel-related structures (perivascular fibrosis)[Figure 1], which cannot be achieved by the human eye, have been shown to associate strongly with survival in a series of IPF patients(29).

However, the utility of quantitation of vessel-related structures (VRS) in predicting survival in IPF has not been examined across multicentred patient cohorts, and in patients with less extensive disease. It also remains unclear whether computer analysis of CT imaging can identify patients at risk of progressive disease and in turn act as a cohort enrichment tool in IPF. Our study therefore evaluated mortality prediction using pulmonary function tests (PFTs), composite indices and visual and computer-based CT scoring in a discovery cohort of IPF patients of varying disease severity. Prediction of decline in FVC using baseline measures, as well as a combined endpoint of a 10% relative FVC decline or death within 12 months was also investigated. Statistically significant variables from these analyses were then examined in a separate IPF validation cohort. We also examined whether specific thresholds of computer-derived CT variables, particularly VRS measures, could be used to enrich drug trial eligible IPF populations.

Materials and Methods

Study Design

All research subjects were diagnosed with IPF by a multidisciplinary team and had received a non-contrast volumetric CT scan as part of their clinical care. In this observational study, all consecutive IPF patients presenting to the Royal Brompton Hospital between January 2007 and June 2011 were included in the discovery cohort. All IPF patients presenting between July 2011 to December 2014 to the Royal Brompton Hospital and IPF patients presenting to the St Antonius Hospital were amalgamated to form the validation dataset.

Pulmonary function test, echocardiography and CT protocols are outlined in the Supplementary appendix. CALIPER CT analysis has been previously described(29, 31), and is outlined in the Supplementary appendix.

CT pattern evaluation

CT variables scored visually and by CALIPER included honeycombing, reticular pattern and ground glass opacity extents. Fibrosis extent represented the sum of reticular and honeycombing extents. Interstitial lung disease (ILD)extent additionally summed ground-glass opacification. Visual scores also quantified traction bronchiectasis extent and severity. CALIPER also quantified pulmonary VRS (details in Supplementary appendix). Volumes for all CALIPER parenchymal features were converted into a percentage using the total lung volume measured by CALIPER. The pulmonary VRS score was subdivided according to cross-sectional area and zonal location of the structures (described below). The VRS subdivisions were expressed as a percentage of the three CALIPER-derived zonal volumes.

Statistical Analysis

We compared the association of the five PFT scores (FVC, DLco, CPI, GAP score and GAP index), five visual CT scores (ILD extent, fibrosis extent, honeycombing extent, traction bronchiectasis severity and extent), four CALIPER CT parenchymal scores (ILD extent, fibrosis extent, honeycombing and CALIPER total VRS) and 18 detailed VRS variables (the five different cross-sectional areas [<5mm^2, 5-10 mm^2, 10-15 mm^2, 15-20 mm^2, >20 mm^2] occurring in either the upper, middle or lower zones of the lungs, and the three total zonal VRS scores) resulting in 32 predictors in total. We considered different measures of clinical outcome, i.e., death or longitudinal FVC trajectories. Each predictor variable was tested alone while correcting for confounders. Adjustment was made for covariates to demonstrate the additional benefit of CALIPER variables when compared to more routinely acquired clinical and functional variables.

In our primary analyses, to compare the predictor variables, the -log10 p-values for each measure from the discovery and validation sets were plotted on the x- and y-axes, respectively, labelling the axes using the p-value scale, and including both horizontal and vertical lines marking the Li and Ji corrected cutoff for statistical significance(32). The Bonferroni method is a conservative method for multiple testing correction: the significance threshold (alpha=0.05) is divided by the number of conducted tests (usually equivalent to the number of tested variables). Bonferroni, however, assumes that all tests are statistically independent. In practice however, this assumption is often violated because of strong correlations between tested variables. This includes the analysis presented here (e.g., CAL VRS and UZ VRS

are strongly correlated $R^2=0.77$). We therefore used the method by Li and Ji in order to estimate the effective number of independent tests (Meff), which is derived from the eigenvalues of the correlation matrix of all tested variables. We then used Meff in order to adjust the significance threshold: alpha=0.05/Meff. Essentially the method by Li and Ji allows calculation of the number of effective independent tests, and then a Bonferroni correction is applied to that number instead of being applied indiscriminately to all variables.

Longitudinal analysis

A subset of subjects in both cohorts was followed longitudinally with repeated pulmonary function tests. Percent change from baseline FVC (raw) was modelled using linear mixed effects with random intercept and random slope. More precisely, each follow-up FVC value was divided by the baseline FVC value and multiplied by 100 (FVC%). FVC% was the target variable and confounders were baseline raw FVC, sex, age at CT image, smoking status (ever vs. never) and time since baseline FVC as well as interactions between time and sex, age, and smoking status. As an additional confounder, we used the predictor variable and then measured the effect of the predictor-by-time interaction as a means on how the predictor affects the longitudinal percentage decline in FVC. The lmer function from the R package lme4(33) was used for the analysis. We present the –log10 p-values of the predictor-by-time interaction for discovery and validation cohort.

In addition to the linear mixed effects analysis, we calculated, for each subject the presence or absence of a 10% decline in FVC at twelve months. More precisely, for

each subject we estimated a 10% FVC loss within 12 months based on best linear unbiased predictions (BLUPs) from the longitudinal mixed model (a minimum of four months of follow-up data was required). The presence of a 10% FVC decline at twelve months was defined based on the fitted trajectory, i.e., the predicted value at twelve months was 10 percent lower than the baseline FVC measurement.

We combined the endpoint of a 10% FVC decline at twelve months with a second endpoint of death within twelve months and using logistic regression, considered both endpoints as the outcome variable. One logistic regression was fitted per predictor and the models were corrected for sex, age at CT imaging and smoking status (ever vs. never). The glm function in R was used for this analysis. As above we present the –log10 p-values of the predictor for discovery and validation cohorts.

Survival analysis

We performed a survival analysis using right-censored Cox proportional hazards models. Time to event (death) or censoring was measured from CT imaging. The models were corrected for sex, age at CT imaging and smoking status (ever vs. never). The analysis was conducted using the coxph function of the survival package(34) in R(35). As above we present the –log10 p-values of the predictor for discovery and validation cohorts.

We also calculated the C-index(36) in the discovery cohort based on 500 bootstrap replicates for the logistic regression (10% FVC decline/death within 12 months) and Cox mortality models. We further computed C-indices in the validation cohort by

13

applying models with regression coefficients for the variable of interest and confounders estimated in the discovery cohort. We also calculated the improvement in the model C-index when a powerful VRS measure was added to a model adjusted for confounders, and containing either FVC, DLco or CPI.

Drug trial power calculation

We conducted a power analysis to explore possible sample size reductions in clinical trials using CVRS as an enrichment parameter. The power analysis was based on a two-sample t-test with p-value threshold of 0.05 and assuming 90% power. Further, we assumed that the control group in an IPF drug trial would be receiving antifibrotic medication rather than a placebo and would on average decline by 120 mls/year in FVC. The rate of decline in FVC in the control arm is therefore an average of rates of FVC decline in the treated arms of anti-fibrotic trials[12,13].

Three different drug effects were investigated: 25%, 40% and 50% effect leading to annual average decline in FVC of 90 ml, 72 ml and 60 ml, respectively. In order to derive the effect size (Cohen's d), we computed the standard deviation in the cohort with DLco≥30. We used the stable estimator for standard deviation based on median absolute deviation (MAD), resulting in 230.27. Thus, in the unenriched design, the effect sizes (d) are 0.13, 0.21 and 0.26, respectively. For the enrichment design, we considered CVRS thresholds that retain 70%, 50% and 30% of subjects with DLco≥30% predicted: 3.7, 4.4, and 5.1. Again, we used the stable estimator to compute the standard deviation of annual FVC decline in these enriched subgroups: 228.1, 197.4 and 195.6. These values were used to derive the corresponding effect

size that would be seen when considering the potency of each drug and the corresponding sample sizes for a trial. In published anti-fibrotic studies where these data are available(16, 37), standard deviations of FVC decline graphed or derived from graphed standard errors, did not differ materially between treatment and non-treatment arms.

The C-index was calculated for the logistic regression and mortality models that contained the CVRS thresholds in both the discovery and validation cohorts where patients fulfilled drug trial entry criteria (DLco≥30% predicted). We finally also examined the effect on study outcome measures from antifibrotic use in drug trial eligible patients selected for cohort enrichment.

RESULTS

Baseline data

To evaluate the potential of computer-derived indices to aid cohort enrichment in a drug-trial setting, our cardinal analyses considered IPF patients fulfilling drug trial inclusion criteria(17). Accordingly, patients with a percent-predicted DLco between 30-90% were analysed in the discovery (n=163) and validation cohort (n=200). Two secondary analyses were also performed. Firstly, after combining the discovery and validation cohorts, we separately examined patients with a DLco≥30% predicted who were not exposed to antifibrotic medication (n=200), and those that had received antifibrotics (n=159), to characterise the effects of medication on CT measures predicting FVC decline and survival. Antifibrotic use was categorised on an intention to treat basis. Secondly, we examined patients with a DLco<30% predicted in both

discovery (n=84) and validation cohorts (n=84) to evaluate the performance of computer tools in predicting the various study outcome measures in patients with severe disease.

Patients with a DLco≥30% predicted in the validation cohort had a marginally higher mean DLco than patients in the discovery cohort with no statistically significant difference identified in other baseline variables (Supplementary Table 1). No statistically significant difference in survival curves was found between cohorts (Supplementary Figure 1a and b).

FVC decline

Across the entire discovery (n=181) and validation (n=207) cohorts, CALIPER variables (particularly upper zone VRS and CALIPER fibrosis extent) were the strongest predictors of FVC decline (Figure 2A). When analyses were limited to patients in the discovery (n=130) and validation (n=168) cohorts with a DLco≥30% predicted, upper and midzone VRS subdivisions and visual traction bronchiectasis extent were more powerful than functional indices and visual ILD and fibrosis extents in predicting FVC decline (Figure 2D).

10% FVC decline/death at 12 months

When all patients in the discovery (n=224) and validation (n=251) cohorts were evaluated, CALIPER VRS variables (particularly total VRS and upper and midzone VRS subdivisions) were the strongest predictors of the combined endpoint, outperforming all functional indices (Figure 2B). In patients with a DLco≥30%

16

predicted across the discovery (n=148) and validation (n=176) cohorts, UZ VRS

subdivisions, and visual traction bronchiectasis severity and extent were the most

powerful predictors of the combined endpoint, though no single variable satisfied

both Li and Ji cutoffs for statistical significance (Figure 2E).

Survival analyses

When all study patients were examined in the discovery (n=247) and validation

(n=284) cohorts, DLco and CPI were the most powerful predictors of survival. Total

VRS and UZ VRS also strongly predicted survival across both cohorts (Figure 2C). In

patients fulfilling drug trial entry criteria (DLco≥30% predicted)[n=363], functional

indices (CPI and DLco), total and UZ VRS and large VRS (>20mm) in the middle and

upper zones best predicted survival (Figure 2F).

Subanalysis related to antifibrotic administration

Analyses in patients who had never received antifibrotic medication but fulfilled

drug trial entry criteria (DLco≥30% predicted), in the discovery (n=128) and

validation cohorts (n=72) are demonstrated in Figure 2G-I, and Figure 3A-C. Upper

zone VRS variables were the strongest predictors of FVC decline (Figure 2G and 3A),

10% FVC decline/death at 12 months (Figure 2H and 3B) and survival (Figure 2I and

3C). Upper and midzone VRS variables outperformed functional indices for all three

study outcome measures. Results were maintained when patients were stratified

based on the median cohort DLco of 40.15% predicted - below median DLco: Figures

3B, E and H; above median DLco: Figures 3C, F and I.

In patients with a DLco≥30% predicted who had received antifibrotics (n=32 in the

discovery cohort, n=127 in the validation cohort), upper and midzone VRS variables

were the strongest predictors of FVC decline (Supplementary Figure 2D) and 10%

FVC decline/death at 12 months (Supplementary Figure 2E) and outperformed DLco

and CPI. UZ VRS subdivisions were similar to DLco and CPI in their ability to predict

survival (Supplementary Figure 2F).

C-Index analyses

The relative strength of adjusted models predicting the combined endpoint of 10%

FVC decline/death at 12 months were compared using the C-index for models

containing CALIPER VRS subdivision scores, functional indices and visual CT variables.

All models were adjusted for patient age, gender, smoking status and reconstruction

algorithm in patients with a DLco≥30% predicted who had never received

antifibrotics (n=175). Models were examined separately in the combined cohorts,

the discovery cohort and the validation cohort (Supplementary Table 2). Models in

the combined cohorts, containing CALIPER VRS subdivisions outperformed models

containing functional indices in predicting the combined endpoint.

When adjusted Cox mortality models were compared using the C-index, models

containing CALIPER VRS subdivisions demonstrated higher C-indices than models

containing FVC, DLco and CPI in the combined cohort (Supplementary Table 2). The

addition of an upper or midzone VRS subdivision score to a model containing a

functional index increased, to a modest degree, the model C-index in all cases

(Supplementary Table 3).

Patients with severe disease

When patients with severe disease (DLco<30% predicted) were evaluated (n=168), CPI and DLco were clearly shown to be the strongest predictors of survival (Supplementary Figure 3C). Results were maintained on subanalysis of patients who received antifibrotic medication (n=29)[Supplementary Figure 2C]. No variables strongly predicted FVC decline or a 10% FVC decline/death at 12 months in all study patients with a DLco<30% predicted (Supplementary Figure 3A and B) or when those receiving antifibrotic medication were subanalysed (n=27)[Supplementary Figure 2A and B].

When patients with a DLco<30% predicted, who had never received antifibrotics were subanalysed (n=139), lower zone VRS subdivisions were better predictors of FVC decline than DLco or CPI (Supplementary Figure 3D). Upper zone VRS subdivisions demonstrated stronger relationships with 10% FVC decline/death at 12 months and survival than DLco or CPI in patients with severe disease that were not exposed to antifibrotics (Supplementary Figure 3E and F).

Cohort enrichment using CALIPER VRS thresholds

The primary aim of or study was to see whether computer quantitation of CT imaging could have a role in cohort enrichment of IPF drug trials. In a final analysis, we aimed to calculate potential savings to an IPF drug trial that would result from cohort enrichment of a study population using various CALIPER total VRS thresholds (Table 1). Using IPF patients with a DLco≥30% predicted, we modelled sample size

savings for a drug with three potential effect sizes on FVC decline (25%, 40% and 50% reduction in FVC decline). At each drug effect size we looked at cohort enrichment using three CALIPER VRS thresholds (representing the total VRS score throughout the entire lungs) corresponding to 70% (CVRS=3.7% of the lung), 50% (CVRS =4.4% of the lung) and 30% (CVRS =5.1% of the lung) of the original IPF study population.

As shown in Table 1, restricting a clinical trial cohort to IPF patients with a CVRS of 4.4% of the lung or greater, it would be possible to reduce sample size by 26% to identify the same drug treatment effect size. Importantly, half of all IPF patients studied were included in the CVRS threshold of >4.4% of the lung. The model C-index for the CAL VRS threshold of 4.4% in the discovery cohort of patients with a DLco≥30% predicted, not on antifibrotics (n=128) was 0.67 for survival and 0.73 for the combined endpoint of 10% FVC decline/death in 12 months. The equivalent C-indices in the validation cohort (n=72, following exclusion of patients on antifibrotics and those with a DLco<30% predicted) were 0.56 for survival and 0.64 for the combined endpoint.

As a final analysis, to exclude the possibility that the CVRS threshold of 4.4% was identifying patients that were likely to progress regardless of any medical intervention, we examined the effect of antifibrotic medication in all study patients with a DLco≥30% predicted, with a VRS >4.4% (n=190). Our results demonstrated that antifibrotic use in this subpopulation increased life expectancy (Mortality analysis: OR=0.437, 95%CI 0.298 - 0.641; Z= -4.239, p=2.25x10-5; 10% FVC

20

decline/death in 12 months: OR= 0.246, 95%CI 0.122 - 0.498; Z= -3.89, p=9.73x10-5)

and reduced FVC decline (fitted between group difference in annual relative FVC

change: 3.36% ; 95%CI 0.39% - 6.5%; T=2.22, p=0.027). Accordingly, it would seem

that patients being selected using CALIPER VRS thresholds are patients that would

respond to antifibrotic medication and represent a good population with which to

enrich an IPF drug trial.

DISCUSSION

Our study has demonstrated, across two cohorts of patients with IPF, that computer-

derived CT variables, particularly the CALIPER vessel-related structures, predict two

separate disease end-points, survival and FVC decline at 12 months and demonstrate

an enhanced effect size in IPF patients with less extensive disease. Most importantly,

VRS scores can be utilized in cohort enrichment of an IPF population. Patients

selected using a CALIPER total VRS threshold of over 4.4% were shown to reduce an

IPF drug trial sample size by 26%, and were shown to have an increased life-

expectancy and reduced rate of FVC decline when receiving antifibrotic medication

compared to patients not receiving antifibrotics.

Existing computer algorithms that have approached the analysis of CT imaging in IPF

have done so by characterising and quantifying the standard library of CT

parenchymal patterns, recognised visually by radiologists for over 20 years(28, 38-

40). The advent of volumetric CT datasets however has allowed three-dimensional

structural information within the lungs to be analysed across several hundred CT

images. Furthermore, features such as the pulmonary vessel-related structures, that

have not been associated with prognostication outside the ambit of CALIPER studies, can now be quantified.

The VRS measure predominantly quantifies pulmonary arteries and veins but also captures connected tubular structures mainly representing adjoining regions of fibrosis. Our results regarding the utility of CALIPER-defined VRS quantitation as a prognostic tool, may represent the first example of a non-traditional CT parameter, derived by a computer and therefore having no inherent measurement variation constraints, that strongly predicts mortality in IPF. As computer tools evolve and supervised machine learning progresses, the scope for identifying novel prognostic features that have been overlooked using visual analysis can only increase.

Pulmonary function tests, specifically DLco and CPI were the strongest predictors of survival in IPF patients with severe disease (DLco<30% predicted). Without the benefit of comprehensive right heart catheter measurements however, the possibility that a major determinant of mortality in patients with extensive disease reflects pulmonary hypertension can only remain speculative. The prognostic value of functional tests however reduced in strength in patients with less extensive disease. Similarly, the GAP index did not strongly predict the likelihood of future FVC decline in patients with less extensive disease, confirming a previous report(41).

Whilst traction bronchiectasis has been recognised as a predictor of mortality in previous reports in IPF(42, 43), our results are the first proof of their utility in predicting FVC decline. The strength with which a relatively simple visual traction

bronchiectasis score predicted FVC decline suggests that the development of an automated measure of traction bronchiectasis would be a desirable tool for further prognostic analyses in IPF.

VRS in the upper zones were amongst the strongest predictors of mortality in most analyses whilst no powerful mortality signal was obtained from VRS in the lower lung zones, regardless of the severity of underlying fibrosis. The results argue for the selective evaluation of VRS in discrete lung subunits.

The potential utility of CALIPER as a cohort enrichment tool could help reduce the prohibitive cost of modern drug trials in IPF. With patients in the control arms of new trials receiving anti-fibrotics and therefore demonstrating a slower decline in FVC when compared to the earlier placebo-controlled Nintedanib(16) and Pirfednidone(17) trials, the recognition of a drug treatment effect from a novel agent will require larger sample sizes in a trial and potentially longer patient follow up. However, the use of CALIPER total VRS scores may allow selective recruitment of IPF patients that are likely to demonstrate more rapid FVC decline. Accordingly, savings in trial size and follow up and therefore cost are likely.

One of the main limitations of computer analysis of CT imaging relates to the requirement of non-contrast volumetric imaging acquired using appropriate reconstruction algorithms. In the current study, most of the CT scans were performed using edge-enhancing algorithms that can result in the misclassification of honeycombing, reticulation or even ground glass opacities. It was to avoid such

misclassification, that the most severely edge enhanced algorithm (Siemens B80) was not evaluated in our study.

For institution of CALIPER CT analysis as a cohort enrichment tool for a drug trial, several preparatory steps would be necessary at the various clinical trial centres. Firstly, scrupulous attention would need to be paid to the quality of the CT scan acquired. An optimal full inspiratory breath at total lung capacity is essential and could be aided with a form of spirometric gating. The desired scanner algorithms used to reconstruct the CT would be clearly specified to avoid severely edge-enhancing algorithms.

Once acquired, the anonymized volumetric, non-contrast digital imaging and communications in medicine (DICOM) axial CT files would be electronically sent by a CT technician to a central processing hub, for example at the Biomedical Imaging Resource, Mayo Clinic, Rochester, Minnesota, USA. Some pre-processing of CT data might be necessary prior to computer analysis to reduce image noise, and though this step can be inadvertently overlooked, its importance, to ensure optimal parenchymal characterisation, cannot be overstated. Once ready for analysis, CALIPER processing of an entire CT examination takes less than one minute, though is accompanied by careful quality control of segmentation accuracy. The anonymized data-files can then be electronically sent to a trial coordinator the same working day. In our study, CALIPER VRS quantitation remained predictive of mortality across a spectrum of CT algorithms in the validation cohort suggesting that CALIPER VRS measures may have a role in future multicentred studies.

There were limitations to the current study. The measurement of VRS by CALIPER predominantly included pulmonary arteries and veins, but in patients with extensive fibrosis, there is invariably a degree of capture of reticular densities and peribronchial fibrosis. It therefore remains unclear just how much the associated perivascular fibrosis may contribute to overall variable strength. Importantly, the VRS scores strongly predicted survival when measured in the upper lung zones, where fibrosis is least extensive in IPF, and retained utility across independent patient populations and a range of reconstruction algorithms suggesting a robustness to the measure. Lastly, whilst the time intervals of FVC measurement in our study were not standardised in line with current drug trial protocols, our study has the advantage of a longer follow up than is usually possible in a drug trial setting.

In conclusion, our study has shown that in IPF, computer analysis of CT imaging, in particular, quantitation of pulmonary vessel-related structures can strongly predict survival and likelihood of FVC decline with effects enhanced over functional indices in patients with less extensive disease. Importantly, CALIPER VRS scores can selectively identify IPF patients who will reach drug trial endpoints and respond to antifibrotic medication. CALIPER VRS scores may therefore have a major role in drug trial cohort enrichment reducing the prohibitive costs of current IPF trials.

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 (44). Spirometry (Jaeger Master screen PFT, Carefusion Ltd., Warwick, UK and Houten, NL), 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). (45) The gender, age, physiology (GAP) score and GAP index were also calculated for both cohorts by summing threshold ranges of patient age, gender, FVC and DLco(46).

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).

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 0.5-1mm thickness were viewed at window settings optimized for the assessment of the lung parenchyma (width 1500 HU; level -500 HU).

Visual CT scoring

Each CT was independently evaluated by two radiologists (AN and SLW of 5 and 7 years thoracic imaging experience or GC and JB of 3 and 4 years experience), blinded to all clinical information. CT scoring of the study population commenced after evaluation of a training dataset of 15 cases by both pairs of radiologists to identify and resolve underlying inherent biases in the characterisation of CT patterns. Results that were discrepant by more than two standard deviations for both scorer pairs were consensed by a third scorer (JJ) who has access to the original visual CT scores.

CTs were scored on a lobar basis using a continuous scale. The total interstitial lung disease (ILD) extent was estimated to the nearest 5% and subclassified into 4 patterns: reticular pattern, ground-glass opacification, honeycombing, and consolidation, using definitions from the Fleischner Society glossary of terms for thoracic imaging(47). To derive a lobar percentage for each parenchymal pattern, the total lobar ILD extent was multiplied by individual lobar parenchymal pattern extents and divided by 100. The percentage (to the nearest 5%) of each lobe that contained emphysema was also recorded. For each radiologist, the individual lobar

percentages of each parenchymal pattern were summed and divided by 6 to create an averaged lobar score per pattern, per scorer per case.

Traction bronchiectasis, as defined in the Fleischner Society glossary of terms(47), was assigned with two categorical scores. The first "severity" score took into account the average degree of airway dilatation within areas of fibrosis as well as the extent of dilatation throughout the lobe and was given an overall score of: none=0, mild=1, moderate=2, severe=3 in a 19-point scale (score range=0-18). The second "extent" score summed the numbers of segments of the lung (to a maximum of 3 segments per lobe) containing traction bronchiectasis resulting in a 17-point scale (score range =0-16 as the middle lobes contain only two segments).

CALIPER CT evaluation

Data processing

Data processing steps have been previously described(27, 48, 49). In reconstruction algorithms utilizing an edge-enhancing algorithm (Siemens B70, Phillips L and YC), a median filter was applied to 3x3x1 voxel volumes. The Siemens B80 algorithm was markedly edge-enhancing and resulted in misclassification of parenchymal features such as honeycombing. As a result, CT imaging reconstructed with the Siemens B80 algorithm (n=6) were excluded from the study.

After the median filter pre-processing step for the other edge-enhancing CT algorithms, the lungs were extracted from the surrounding thoracic structures and then segmented into upper, middle, and lower lung zones using the carina as a

landmark. Lung segmentation was performed with an adaptive density-based morphologic approach,(50) whereas airway segmentation involved iterative 3-dimensional region growing, density thresholding (thresholds including 950HU and 960 HU), and connected components analysis. Parenchymal tissue type classification was applied to 15x15x15 voxel volume units using texture analysis, computer vision-based image understanding of volumetric histogram signature mapping features, and 3-dimensional morphology(27).

Pulmonary vessel extraction was achieved using an optimised multiscale tubular structure enhancement filter based on the eigenvalues of the Hessian matrix. More precisely, the second-order derivatives that occurred in the regions that surrounded each pulmonary voxel were calculated and formed the Hessian matrix. The eigenvalues of this Hessian matrix were then computed and used to determine the likelihood that an underlying voxel was connected to a dense tubular structure and therefore represented a vessel or vessel-related tubular structure (27, 51). The pulmonary VRS score (CALIPER VRS) excluded vessels at the lung hilum and expressed the VRS as a percentage of the total lung volume.

List of Supplementary Material:

1. Supplementary Table 1. Comparison of patient characteristics in both IPF cohorts.

2. Supplementary Table 2. C-index values for adjusted models containing examined baseline variables.

3. Supplementary Table 3. Change in C-index values for adjusted models when CALIPER vessel-related structure subdivisions were added to functional indices.

4. Supplementary Figure 1. Kaplan Meier survival curves for patients in both IPF cohorts and for those patients in each cohort with a DLco≥30% predicted.

5. Supplementary Figure 2. Graphs demonstrating log10 p-values for various computer-derived (CALIPER) vessel-related structure scores, the diffusing capacity for carbon monoxide (DLco) and the composite physiologic index (CPI), in all study patients (discovery and validation cohorts) that received antifibrotic medication.

6. Supplementary Figure 3. Graphs demonstrating log10 p- for various computer-derived (CALIPER) vessel-related structure scores, the diffusing capacity for carbon monoxide (DLco) and the composite physiologic index (CPI), in all study patients (discovery and validation cohorts) with a DLco<30% predicted, with a subanalysis in patients that did not receive antifibrotic medication.

CVRS Threshold (Percentage of population)	Setting A (25% effect)		Setting B (40% effect)		Setting C (50% effect)	
	N (size)	Size diff (%)	N (size)	Size diff (%)	N (size)	Size diff (%)
0 (100)	2480	0	970	0	622	0
3.7 (70)	2432	48 (2)	952	18 (2)	610	12 (2)
4.4 (50)	1824	656 (26)	714	256 (26)	458	164 (26)
5.1 (30)	1790	690 (28)	702	268 (28)	450	172 (28)

Table 1. Impact of cohort enriching an IPF drug trial study population by analysing baseline CT imaging and stratifying patients according to various thresholds of CALIPER total vessel-related structures (CVRS). Statistical modelling was performed to calculate the number of patients that would be required in a study to identify a drug effect of 25%, 40% and 50% reductions in forced vital capacity (FVC) decline. The number of patients in the unenriched cohort are demonstrated (CVRS threshold of 0) and the savings in drug trial sample size when using three CVRS thresholds (3.7%, 4.4%, 5.1% of the lung equating to 70%, 50% and 30% of the study population respectively) are expressed as a number and percentage.
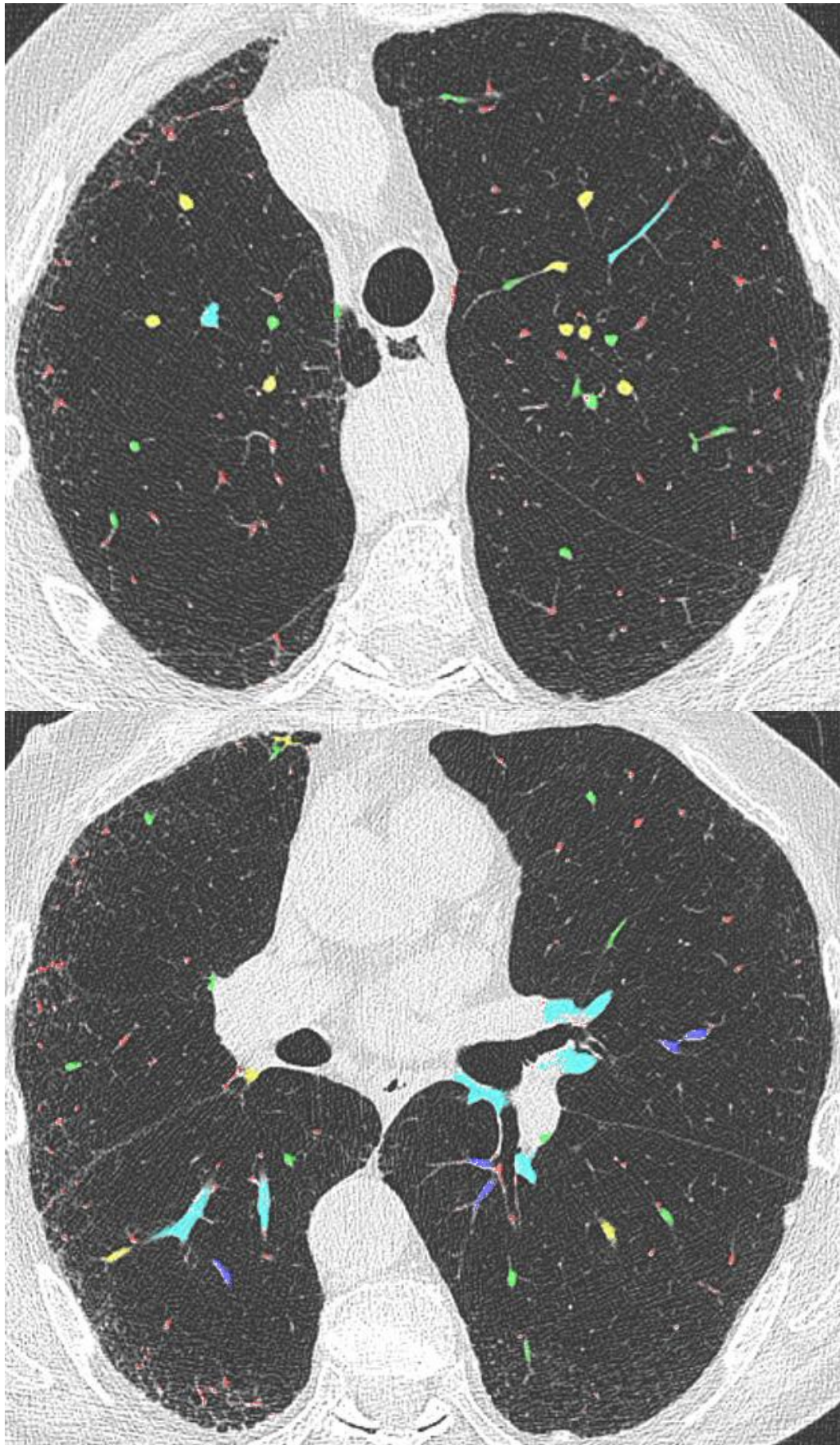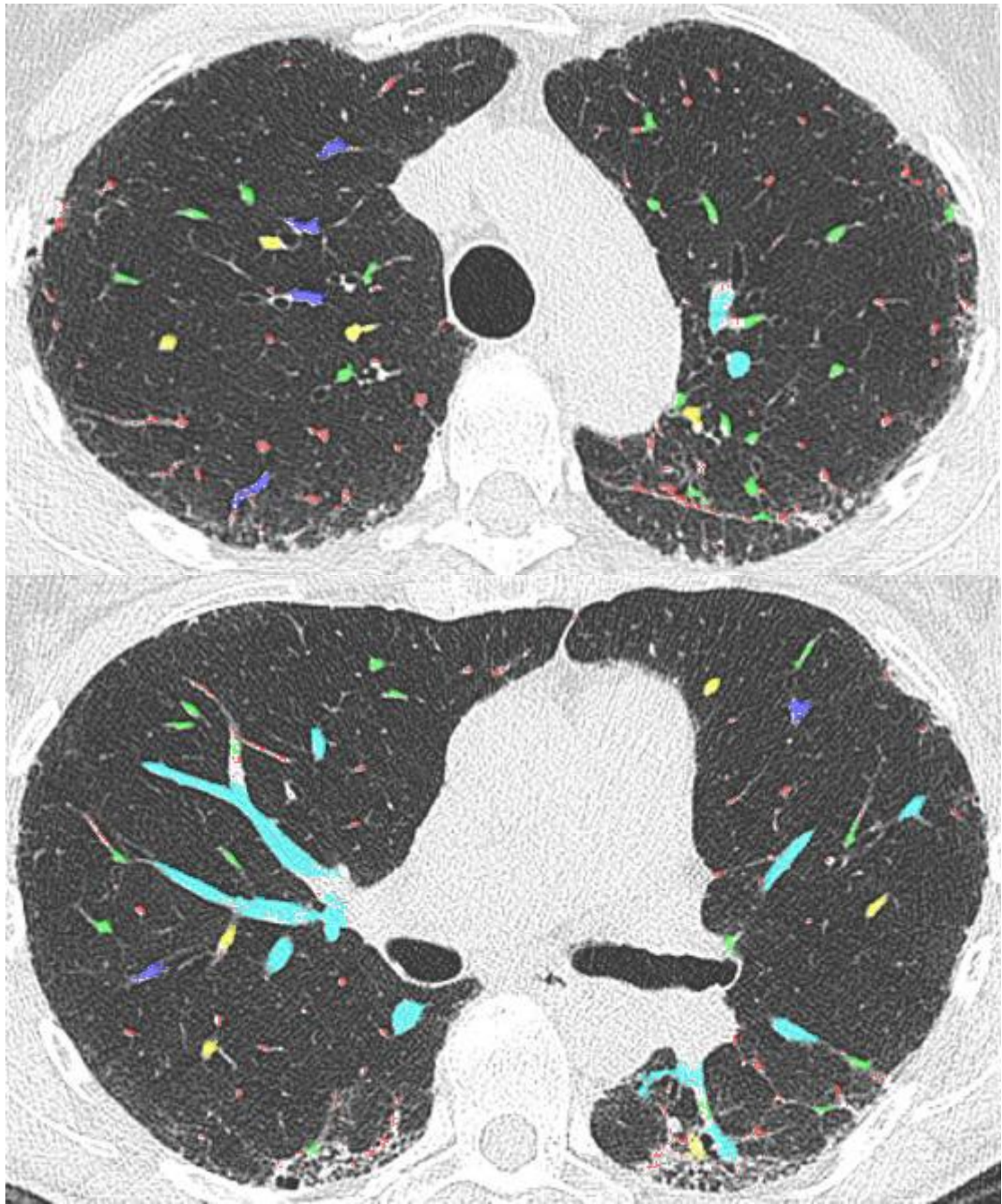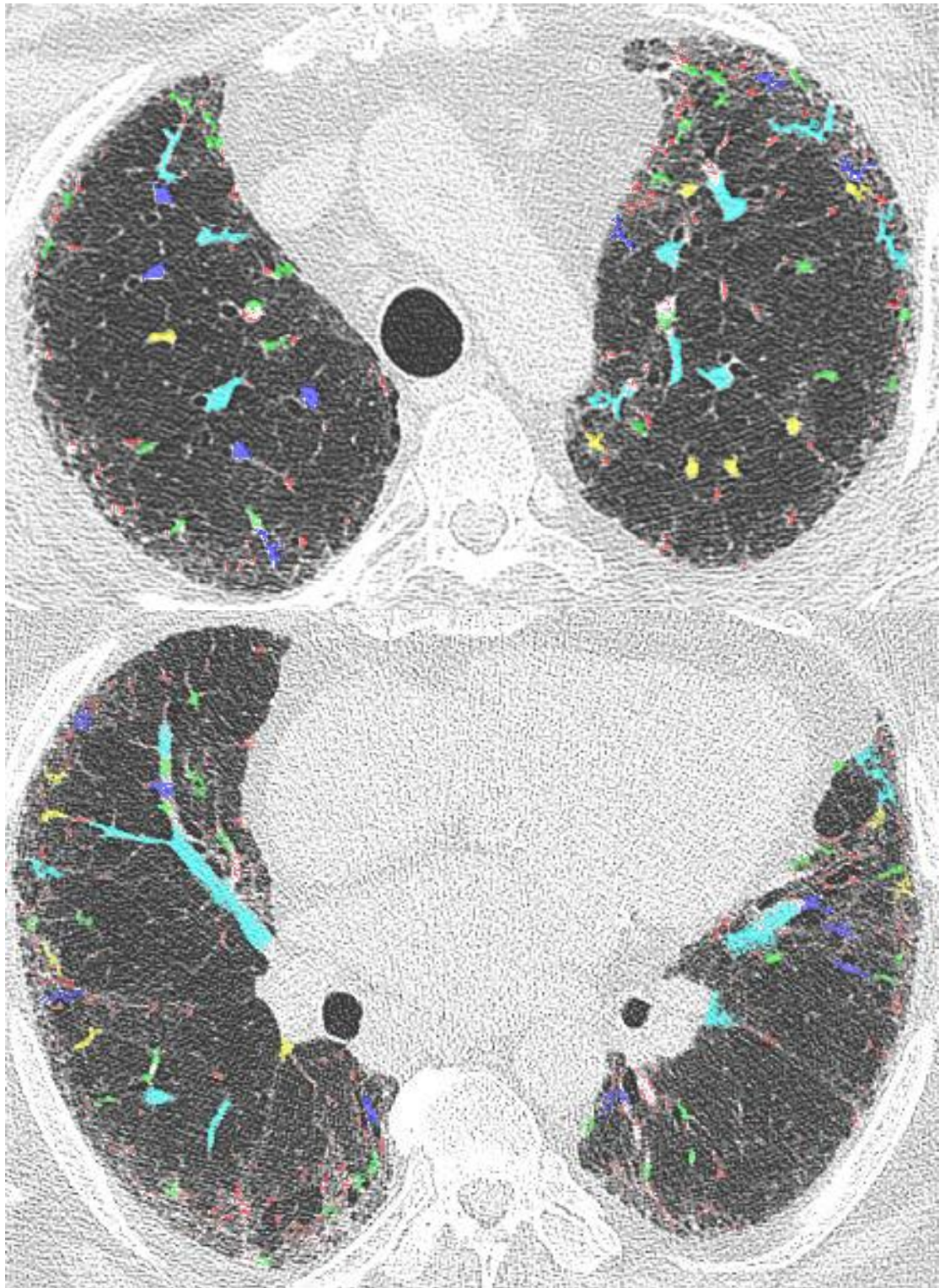
Figure 1. Colour overlay CT images demonstrating CALIPER vessel-related structures (VRS) of different sizes in the upper and middle lung zones in three IPF patients with varying degrees of disease severity. A 78-year-old male ex-smoker (diffusion capacity for carbon monoxide (DLco)=69.7% predicted) demonstrates mild reticulation at the

anterior and posterior aspects of the lungs (A+B). A 56-year-old female ex-smoker (DLco=57.1% predicted) demonstrates coarse reticulation most marked in the apical segments of the lower lobes bilaterally (C+D). A 68-year-old male ex-smoker (DLco=27.5% predicted) has more extensive reticulation in the lung periphery and at the lung bases. As fibrosis extent increases, so too does the VRS score within the lungs. VRS size key: red=<5mm^2, green=5-10mm^2, yellow=10-15mm^2, dark blue=15-20mm^2, light blue=>20mm^2.

Figure 2. Scatterplots demonstrating -log10 p-values for variables (various computer-derived (CALIPER), visual CT-derived and pulmonary function indices) in the discovery cohort (x-axis) and validation cohort (y-axis). Horizontal and vertical dotted lines represent the Li and Ji corrected cutoff for statistical significance. Graphs A-C represent all subjects, graphs D-F represent patients with a DLco≥30% predicted, graphs G-I represent patients with a DLco≥30% predicted who were not exposed to antifibrotic medication. The first column of graphs (A,D,G) represent variables predicting FVC decline. The second column of graphs (B,E,H) represent variables predicting a 10% FVC decline or death within 12 months. The third column of graphs (C,F,I) represent variables predicting survival. In graph C, to allow visualization of all the points on the graph, values that were infinite (due to a p-value of 0) were set to 20. DLco=diffusing capacity for carbon monoxide, CPI=composite physiologic index, TxBx=traction bronchiectasis (extent and severity), CFibrosis=CALIPER fibrosis extent, CAL VRS=CALIPER vessel-related structure, UZ=upper zone, MZ=middle zone, UZ VRS=upper zone vessel-related structure. The pulmonary vessel-related structure score was subdivided according to structure cross-sectional area ($<5mm^2$, $5\text{-}10\ mm^2$, $10\text{-}15\ mm^2$, $15\text{-}20\ mm^2$, $>20\ mm^2$).

Figure 3. Graphs A-C demonstrate -log10 p-values for variables (various computer-derived (CALIPER), visual CT-derived, echocardiography-derived and pulmonary function indices) in all study patients (discovery and validation cohorts) with a DLco ≥30% predicted who were not exposed to antifibrotic medication. The patients were stratified with regard to disease severity, based on the median DLco value for the combined cohort (40.15% predicted), with results for patients below the median

DLco shown in graphs (D-F) and patients above the median DLco shown in graphs (G-I). Graphs A,D and G represent variables predicting FVC decline, graphs B,E and H represent variables predicting a 10% FVC decline or death within 12 months, and graphs C,F and I represent variables predicting survival. DLco=diffusing capacity for carbon monoxide, CPI=composite physiologic index, CVRS=CALIPER vessel-related structure, UZ=upper zone, MZ=middle zone, LZ=lower zone, UZVRS=upper zone vessel-related structure, MZVRS=middle zone vessel-related structure, LZVRS=lower zone vessel-related structure, The pulmonary vessel-related structure score was subdivided according to structure cross-sectional area ($<5mm^2$, 5-10 mm^2, 10-15 mm^2, 15-20 mm^2, >20 mm^2).

Supplementary Materials:

Population	Variable Units are percentage unless stated	Discovery Cohort (n=247 unless stated)	Validation Cohort (n=284 unless stated)	Cohort Comparison
All IPF patients	Median Age (years)	67	69	0.14*
	Male/female (ratio)	192/55	225/59	0.68*
	Survival (alive/dead)	63/184	93/191	0.07*
	FVC % predicted	69.1 ± 20.1 (242)	74.2 ± 19.5 (283)	0.003
	DLco % predicted	36.3 ± 12.9	39.1 ± 14.4	0.02
	CPI	55.2 ± 11.5 (242)	52.8 ± 12.6 (283)	0.02
	CALIPER ILD extent	25.4 ± 17.0	23.5 ± 16.3	0.19
	CALIPER VRS	5.0 ± 1.6	5.3 ± 1.7	0.10
IPF patients with a DLco<30% predicted	FVC % predicted	75.3 ± 18.7 (159)	79.0 ± 17.8 (199)	0.06
	DLco % predicted	43.0 ± 10.2 (163)	45.8 ± 11.4 (200)	0.01
	CPI	49.5 ± 9.3 (159)	47.4 ± 10.4 (199)	<0.05
	CALIPER ILD extent	19.8 ± 13.2 (163)	18.4 ± 13.1 (200)	0.32
	CALIPER VRS	4.5 ± 1.3 (163)	4.7 ± 1.5 (200)	0.08

Supplementary Table 1. Comparison of patient characteristics in both IPF cohorts.

Key characteristics of all IPF patients in the discovery and validation cohorts with a

subanalysis in patients with a DLco≥30% predicted. Data represent median values for

patient age and mean values with standard deviations for CALIPER and lung function

variables. Statistically significant differences between the two groups were

calculated using the Chi-Square test for categorical independent variables (*) and the

T test for continuous variables. FVC = forced vital capacity, DLco=diffusing capacity

for carbon monoxide, CPI=composite physiologic index, ILD=interstitial lung disease,

VRS= vessel-related structure.

Variable	Combined Cohort		Discovery Cohort		Validation Cohort	
	Survival	10% FVC decline or death in 12mths	Survival	10% FVC decline or death in 12mths	Survival	10% FVC decline or death in 12mths
CAL VRS	0.68	0.73	0.69	0.75	0.63	0.67
UZ VRS	0.67	0.76	0.69	0.77	0.61	0.69
UZ (5-10mm)	0.67	0.75	0.69	0.76	0.58	0.69
UZ (10-15mm)	0.67	0.76	0.67	0.76	0.60	0.72
MZ VRS	0.67	0.70	0.68	0.72	0.63	0.64
UZ (>20mm)	0.66	0.74	0.67	0.75	0.62	0.71
MZ (15-20mm)	0.66	0.70	0.67	0.72	0.62	0.63
MZ (10-15mm)	0.66	0.70	0.67	0.70	0.61	0.66
MZ (>20mm)	0.66	0.69	0.68	0.73	0.63	0.65
UZ (<5mm)	0.65	0.73	0.68	0.75	0.54	0.57
CFibrosis	0.65	0.73	0.65	0.70	0.64	0.73
UZ (15-20mm)	0.64	0.75	0.65	0.76	0.59	0.69
CILD	0.64	0.69	0.64	0.70	0.57	0.64
MZ (5-10mm)	0.64	0.68	0.65	0.69	0.58	0.61
CPI	0.64	0.68	0.63	0.66	0.66	0.62
GAPindex	0.63	0.68	0.64	0.62	0.64	0.76
FVC	0.63	0.68	0.63	0.66	0.61	0.63
MZ (<5mm)	0.62	0.65	0.63	0.64	0.56	0.53
LZ (>20mm)	0.62	0.64	0.60	0.59	0.60	0.72
LZ (15-20mm)	0.61	0.65	0.60	0.65	0.58	0.65
DLco	0.61	0.65	0.59	0.57	0.65	0.62
LZ VRS	0.61	0.64	0.57	0.59	0.59	0.63
TxBx (extent)	0.60	0.65	0.61	0.68	0.57	0.62
LZ (10-15mm)	0.60	0.64	0.58	0.60	0.57	0.59
TxBx (severity)	0.60	0.64	0.57	0.61	0.60	0.67
VILD	0.60	0.63	0.60	0.61	0.57	0.55
VFibrosis	0.60	0.62	0.58	0.57	0.56	0.57
LZ (5-10mm)	0.56	0.60	0.54	0.58	0.52	0.53
VHC	0.55	0.61	0.58	0.52	0.48	0.66
LZ (<5mm)	0.53	0.60	0.51	0.51	0.48	0.48
CHC	0.49	0.60	0.49	0.51	0.45	0.67
GAPscore	0.48	0.59	0.56	0.52	0.33	0.41

Supplementary Table 2. C-index values for CALIPER, functional and visual CT

variables predicting survival or a combined endpoint of a 10% FVC decline/death

within 12 months, in all study patients with a DLco≥30% not receiving antifibrotic

medication. Results are separately shown for the combined endpoint in the combined population (n=175), the discovery cohort (n=118) and the validation cohort (n=57) and mortality analyses in the combined population (n=200), the discovery cohort (n=128) and the validation cohort (n=72). All models were adjusted for patient age, gender, smoking status (never versus ever) and CT reconstruction algorithm. FVC = forced vital capacity, DLco=diffusing capacity for carbon monoxide, CPI=composite physiologic index, GAP=gender, age and physiology (score and index), CILD=CALIPER interstitial lung disease extent, CFibrosis=CALIPER fibrosis extent, CHC=CALIPER honeycombing extent, CAL VRS=CALIPER vessel-related structure, UZ=upper zone, MZ=middle zone, LZ=lower zone, UZVRS=upper zone vessel-related structure, MZVRS=middle zone vessel-related structure, LZVRS=lower zone vessel-related structure, VILD=Visual interstitial lung disease extent, VFibrosis=Visual fibrosis extent, VHC=Visual honeycombing extent, TxBx=traction bronchiectasis. The pulmonary vessel-related structure score was subdivided according to structure cross-sectional area (<5mm^2, 5-10 mm^2, 10-15 mm^2, 15-20 mm^2, >20 mm^2).

	FVC	DLco	CPI
UZ VRS	0.044	0.065	0.039
MZ VRS	0.045	0.054	0.032
LZ VRS	0.011	0.011	0.003
UZ (<5mm)	0.019	0.045	0.025
UZ (5-10mm)	0.037	0.059	0.032
UZ (10-15mm)	0.036	0.058	0.031
UZ (15-20mm)	0.021	0.043	0.020
UZ (>20mm)	0.035	0.058	0.032
MZ (<5mm)	0.017	0.022	0.008
MZ (5-10mm)	0.024	0.033	0.016
MZ (10-15mm)	0.038	0.050	0.026
MZ (15-20mm)	0.037	0.052	0.028
MZ (>20mm)	0.039	0.051	0.035
LZ (<5mm)	-0.006	-0.004	0.003
LZ (5-10mm)	-0.004	-0.002	-0.001
LZ (10-15mm)	0.008	0.006	0.001
LZ (15-20mm)	0.009	0.013	0.007
LZ (>20mm)	0.024	0.030	0.018
CAL VRS	0.051	0.069	0.044

Supplementary Table 3. Change in C-index values when CALIPER VRS variables were added to adjusted survival models containing functional indices (forced vital capacity (FVC); diffusing capacity for carbon monoxide (DLco) and composite physiologic index (CPI). All study patients with a DLco≥30% predicted not receiving antifibrotic medication were examined (n=200) and all models were adjusted for patient age, gender, smoking status (never versus ever) and CT reconstruction algorithm. A positive value indicates that the addition of a VRS variable improves the model. VILD=Visual interstitial lung disease extent, VFibrosis=Visual fibrosis extent, VHC=Visual honeycombing extent, CAL VRS=CALIPER vessel-related structure, UZ=upper zone, MZ=middle zone, LZ=lower zone, UZVRS=upper zone vessel-related structure, MZVRS=middle zone vessel-related structure, LZVRS=lower zone vessel-related structure. The pulmonary vessel-related structure score was subdivided

according to structure cross-sectional area ($<5mm^2$, 5-10 mm^2, 10-15 mm^2, 15-20 mm^2, >20 mm^2).
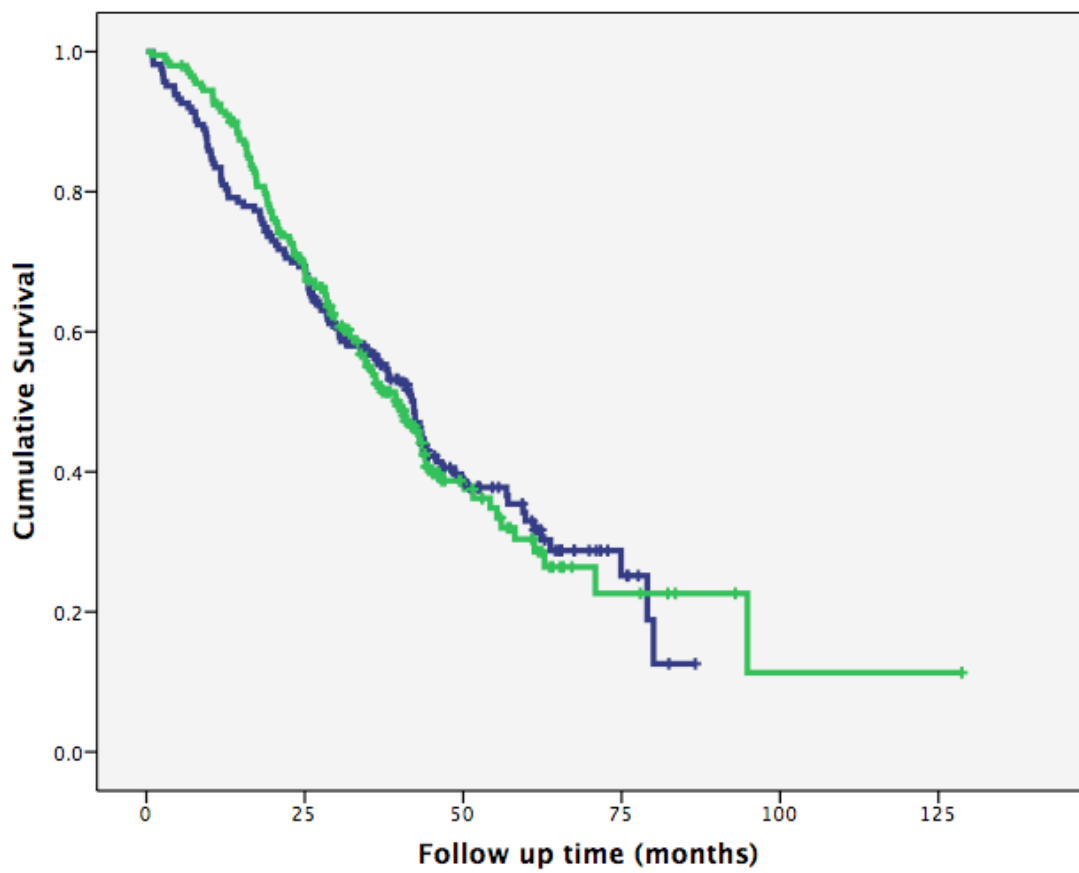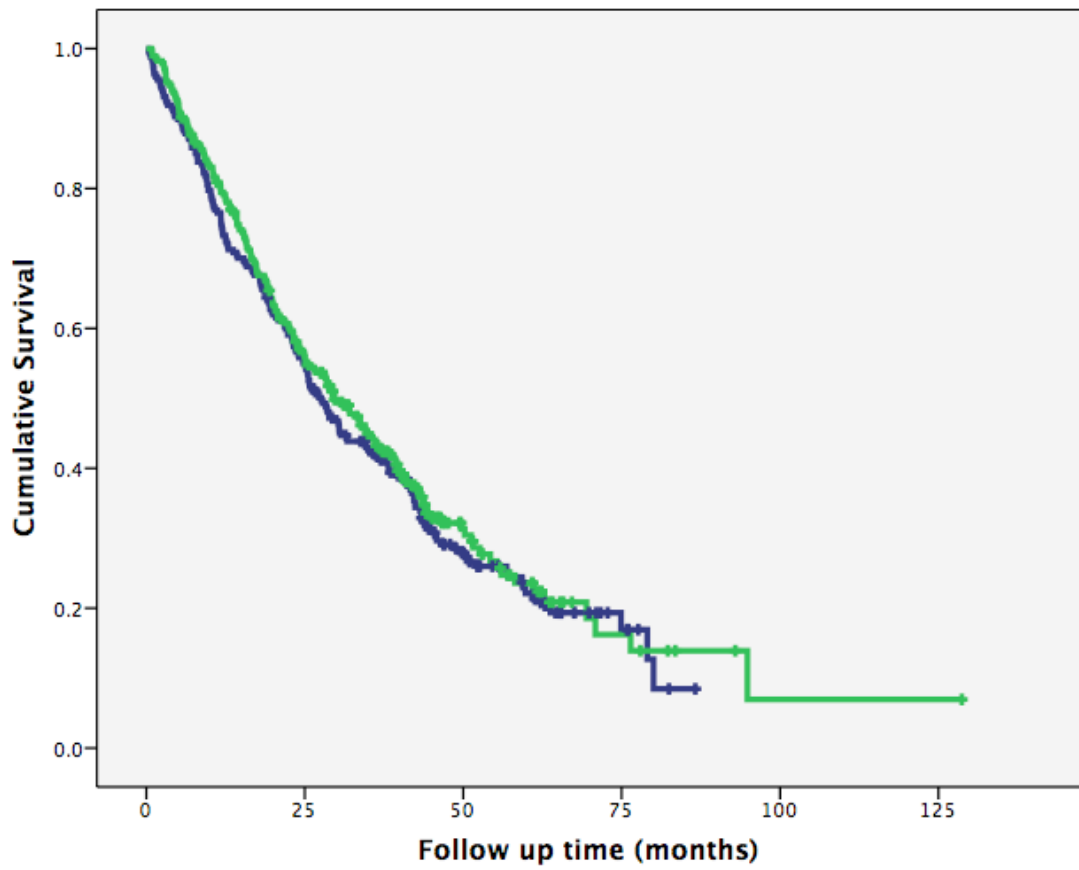
Supplementary Figure 1. Kaplan Meier survival curves for IPF patients in the discovery (a) and validation (b) cohorts and for those patients in each cohort with a DLco≥30% predicted. All IPF patients in discovery cohort: blue, n=247, restricted median survival=27.6±1.8; validation cohort: green, n=284, restricted median survival=29.7±2.9. Log rank test p=0.52. IPF patients with a DLco≥30% predicted in discovery cohort: blue, n=163, restricted median survival=42.2±2.6; validation cohort: green, n=200, restricted median survival=39.7±3.0. Log rank test p=0.98

Supplementary Figure 2. Graphs A-C demonstrate -log10 p-values for CALIPER vessel-related structure subdivisions, composite physiologic index (CPI) and diffusing capacity for carbon monoxide (DLco), in all study patients with a DLco<30% predicted (discovery and validation cohorts) that received antifibrotic medication (n=29). Graphs D-F represent similar analyses in study patients with a DLco≥30% predicted, who received antifibrotic medication (n=159). Graphs A and D represent variables predicting FVC decline, graphs B and E represent variables predicting a 10% FVC decline or death within 12 months, and graphs C and F represent variables predicting survival. DLco=diffusing capacity for carbon monoxide, CPI=composite physiologic index, CAL VRS=CALIPER vessel-related structure, UZ=upper zone, MZ=middle zone, LZ=lower zone, UZVRS=upper zone vessel-related structure, MZVRS=middle zone vessel-related structure, LZVRS=lower zone vessel-related structure. The pulmonary vessel-related structure score was subdivided according to structure cross-sectional area (<5mm^2, 5-10 mm^2, 10-15 mm^2, 15-20 mm^2, >20 mm^2).

Supplementary Figure 3. Graphs A-C demonstrate -log10 p-values for CALIPER vessel-related structure subdivisions, composite physiologic index (CPI) and diffusing capacity for carbon monoxide (DLco), in all study patients (discovery and validation cohorts) with a DLco<30% predicted (n=168). Graphs D-F represent a subanalysis in study patients with a DLco<30% predicted, who were not exposed to antifibrotic medication (n=139). Graphs A and D represent variables predicting FVC decline, graphs B and E represent variables predicting a 10% FVC decline or death within 12 months, and graphs C and F represent variables predicting survival. DLco=diffusing capacity for carbon monoxide, CPI=composite physiologic index, CAL VRS=CALIPER vessel-related structure, UZ=upper zone, MZ=middle zone, LZ=lower zone, UZVRS=upper zone vessel-related structure, MZVRS=middle zone vessel-related structure, LZVRS=lower zone vessel-related structure. The pulmonary vessel-related structure score was subdivided according to structure cross-sectional area (<5mm^2, 5-10 mm^2, 10-15 mm^2, 15-20 mm^2, >20 mm^2).

References:
1.   Navaratnam V, Fleming KM, West J, Smith CJP, Jenkins RG, Fogarty A, Hubbard RB. The rising incidence of idiopathic pulmonary fibrosis in the UK. Thorax. 2011;66:462-7.
2.   Raghu G, Weycker D, Edelsberg J, Bradford WZ, Oster G. Incidence and Prevalence of Idiopathic Pulmonary Fibrosis. Am J Respir Crit Care Med. 2006;174(7):810-6.
3.   Pérez ERF, Daniels CE, Schroeder DR, St. Sauver J, Hartman TE, Bartholmai BJ, Yi ES, Ryu JH. Incidence, prevalence, and clinical course of idiopathic pulmonary fibrosis: A population-based study. Chest. 2010;137(1):129-37.
4.   Hutchinson J, Fogarty A, Hubbard R, McKeever T. Global incidence and mortality of idiopathic pulmonary fibrosis: a systematic review. Eur Respir J. 2015;46(3):795-806.
5.   Harari S, Madotto F, Caminati A, Conti S, Cesana G. Epidemiology of Idiopathic Pulmonary Fibrosis in Northern Italy. PLoS One. 2016;11(2):e0147072.

6.   Schwartz DA, Helmers RA, Galvin JR, van Fossen DS, Frees KL, Dayton, Cs, Burmeister LF, Hunninghake GW. Determinants of survival in idiopathic pulmonary fibrosis. Am J Respir Crit Care Med. 1994;149(2 Pt 1):450-4.

7.   Flaherty KR, Thwaite EL, Kazerooni EA, Gross BH, Toews GB, Colby TV, Travis WD, Mumford JA, Murray S, Flint A, Lynch JP, III, Martinez FJ. Radiological versus histological diagnosis in UIP and NSIP: survival implications. Thorax. 2003;58:143-8.

8.   Ley B, Collard HR, King TE, Jr. Clinical course and prediction of survival in idiopathic pulmonary fibrosis. AmJRespirCrit Care Med. 2011;183(4):431-40.

9.   Araki T, Katsura H, Sawabe M, Kida K. A Clinical Study of Idiopathic Pulmonary Fibrosis Based on Autopsy Studies in Elderly Patients. Intern Med. 2003;42(6):483-9.

10.   Nathan SD, Shlobin OA, Weir N, Ahmad S, Kaldjob JM, Battle E. Long-term course and prognosis of idiopathic pulmonary fibrosis in the new millennium. Chest. 2011;140:221-9.

11.   Raghu G, Collard HR, Egan JJ, Martinez FJ, Behr J, Brown KK, Colby TV, Cordier JF, Flaherty KR, Lasky JA, Lynch DA, Ryu JH, Swigris JJ, Wells AU, Ancochea J, Bouros D, Carvalho C, Costabel U, Ebina M, Hansell DM, Johkoh T, Kim DS, King TEJ, Kondoh Y, Myers J, Müller NL, Nicholson AG, Richeldi L, Selman M, Dudden RF, Griss BS, Protzko SL, Schünemann HJ, Fibrosis. AEJACoIP. 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.

12.   Richeldi L, Ryerson CJ, Lee JS, Wolters PJ, Koth LL, Ley B, Elicker BM, Jones KD, King TE, Ryu JH, Collard HR. Relative versus absolute change in forced vital capacity in idiopathic pulmonary fibrosis. Thorax. 2012;67(5):407-11.

13.   du Bois RM, Weycker D, Albera C, Bradford WZ, Costabel U, Kartashov A, King TE, Lancaster L, Noble PW, Sahn SA, Thomeer M, Valeyre D, Wells AU. Forced Vital Capacity in Patients with Idiopathic Pulmonary Fibrosis. Am J Respir Crit Care Med. 2011;184(12):1382-9.

14.   Zappala CJ, Latsi PI, Nicholson AG, Colby TV, Cramer D, Renzoni EA, Hansell DM, Du Bois RM, Wells AU. Marginal decline in forced vital capacity is associated with a poor outcome in idiopathic pulmonary fibrosis. Eur Respir J. 2010;35(4):830-6.

15.   Collard HR, King TE, Jr., Bartelson BB, Vourlekis JS, Schwarz MI, Brown KK. Changes in clinical and physiologic variables predict survival in idiopathic pulmonary fibrosis. Am J RespirCrit Care Med. 2003;168:538-42.

16.   Richeldi L, du Bois RM, Raghu G, Azuma A, Brown KK, Costabel U, Cottin V, Flaherty KR, Hansell DM, Inoue Y, Kim DS, Kolb M, Nicholson AG, Noble PW, Selman M, Taniguchi H, Brun M, Le Maulf F, Girard M, Stowasser S, Schlenker-Herceg R, Disse B, Collard HR. Efficacy and Safety of Nintedanib in Idiopathic Pulmonary Fibrosis. N Engl J Med. 2014;370(22):2071-82.

17.   King TE, Bradford WZ, Castro-Bernardini S, Fagan EA, Glaspole I, Glassberg MK, Gorina E, Hopkins PM, Kardatzke D, Lancaster L, Lederer DJ, Nathan SD, Pereira CA, Sahn SA, Sussman R, Swigris JJ, Noble PW. A Phase 3 Trial of Pirfenidone in Patients with Idiopathic Pulmonary Fibrosis. N Engl J Med. 2014;370(22):2083-92.

18.   Jegal Y, Kim DS, Shim TS, Lim C-M, Do Lee S, Koh Y. Physiology is a stronger predictor of survival than pathology in fibrotic interstitial pneumonia. Am J Respir Crit Care Med. 2005;171:639-44.

19.   King TE, Safrin S, Starko KM, Brown KK, Noble PW, Raghu G, DA S. Analyses of efficacy end points in a controlled trial of interferon-gamma1b for idiopathic pulmonary fibrosis. Chest. 2005;127:171-7.

20.   Latsi PI, Du Bois RM, Nicholson AG, Colby TV, Bisirtzoglou D, Nikolakopoulou A, Veeraraghavan S, Hansell DM, Wells AU. Fibrotic idiopathic interstitial pneumonia: the prognostic value of longitudinal functional trends. Am J RespirCrit Care Med. 2003;168:531-7.

21.   Flaherty KR, Mumford JA, Murray S, Kazerooni EA, Gross BH, Colby TV. Prognostic implications of physiologic and radiographic changes in idiopathic interstitial pneumonia. Am J Respir Crit Care Med. 2003;168.

22.   Pellegrino R, Viegi G, Brusasco V, Crapo RO, Burgos F, Casaburi R, Coates A, van der Grinten CPM, Gustafsson P, Hankinson J, Jensen R, Johnson DC, MacIntyre N, McKay R, Miller MR, Navajas D, Pedersen OF, Wanger J. Interpretative strategies for lung function tests. Eur Respir J. 2005;26(5):948-68.

23.   Flaherty KR, Mumford JA, Murray S, Kazerooni E, Gross BH, Colby TV, Travis WD, Flint A, Toews GB, Lynch JP, Martinez FJ. Prognostic implications of physiologic and radiographic changes in idiopathic interstitial pneumonia. Am J Respir Crit Care Med. 2003;168:543-8.

24.   Oldham JM, Noth I. Idiopathic pulmonary fibrosis: Early detection and referral. Respir Med. 2014;108(6):819-29.

25.   Blackwell TS, Tager AM, Borok Z, Moore BB, Schwartz DA, Anstrom KJ, Bar-Joseph Z, Bitterman P, Blackburn MR, Bradford W, Brown KK, Chapman HA, Collard HR, Cosgrove GP, Deterding R, Doyle R, Flaherty KR, Garcia CK, Hagood JS, Henke CA, Herzog E, Hogaboam CM, Horowitz JC, Talmadge E. King J, Loyd JE, Lawson WE, Marsh CB, Noble PW, Noth I, Sheppard D, Olsson J, Ortiz LA, O'Riordan TG, Oury TD, Raghu G, Roman J, Sime PJ, Sisson TH, Tschumperlin D, Violette SM, Weaver TE, Wells RG, White ES, Kaminski N, Martinez FJ, Wynn TA, Thannickal VJ, Eu JP. Future Directions in Idiopathic Pulmonary Fibrosis Research. An NHLBI Workshop Report. Am J Respir Crit Care Med. 2014;189(2):214-22.

26.   Collard HR. Improving survival in idiopathic pulmonary fibrosis: The race has just begun. Chest. 2017;151(3):527-8.

27.   Bartholmai BJ, Raghunath S, Karwoski RA, Moua T, Rajagopalan S, Maldonado F, Decker PA, Robb RA. Quantitative CT imaging of interstitial lung diseases. J Thorac Imaging. 2013;28(5):298-307.

28.   Kim HJ, Brown MS, Chong D, Gjertson DW, Lu P, Kim HJ, Coy H, Goldin JG. Comparison of the Quantitative CT Imaging Biomarkers of Idiopathic Pulmonary Fibrosis at Baseline and Early Change with an Interval of 7 Months. Acad Radiol. 2015;22(1):70-80.

29.   Jacob J, Bartholmai B, Rajagopalan S, Kokosi M, Nair A, Karwoski R, Walsh S, Wells AU, Hansell DM. 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.

30.   Jacob J, Bartholmai BJ, Rajagopalan S, Brun AL, Egashira R, Karwoski R, Kokosi M, Wells AU, Hansell DM. 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.

31.   Jacob J, Bartholmai B, Rajagopalan S. Automated quantitative CT versus visual CT scoring in idiopathic pulmonary fibrosis: validation against pulmonary function. J Thorac Imaging. 2016;31.

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

33.   Bates D, Mächler M, Bolker B, Walker S. Fitting Linear Mixed-Effects Models Using lme4. Journal of Statistical Software. 2015;67:1-48.

34.   Therneau T. A package for survival analysis in S_. 2.38 ed2015.

35.   Team RC. R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing; 2015.

36.   Harrell FEJ. Regresssion modelling strategies. 1 ed: Springer-Verlag, New York; 2001.

37.   Richeldi L, Costabel U, Selman M, Kim DS, Hansell DM, Nicholson AG, Brown KK, Flaherty KR, Noble PW, Raghu G, Brun M, Gupta A, Juhel N, Kluglich M, Du Bois RM. Efficacy of a tyrosine kinase inhibitor in idiopathic pulmonary fibrosis. N Engl J Med. 2011;365(12):1079-87.

38.   Salisbury ML, Lynch DA, Beek EJRv, Kazerooni EA, Guo J, Xia M, Murray S, Anstrom KJ, Yow E, Martinez FJ, Hoffman EA, Flaherty KR. Idiopathic Pulmonary Fibrosis: The Association between the Adaptive Multiple Features Method and Fibrosis Outcomes. Am J Respir Crit Care Med. 2017;195(7):921-9.

39.   Park HJ, Lee SM, Song JW, Lee SM, Oh SY, Kim N, Seo JB. Texture-Based Automated Quantitative Assessment of Regional Patterns on Initial CT in Patients With Idiopathic Pulmonary Fibrosis: Relationship to Decline in Forced Vital Capacity. American Journal of Roentgenology. 2016;207(5):976-83.

40.   Iwasawa T, Ogura T, Sakai F, Kanauchi T, Komagata T, Baba T, Gotoh T, Morita S, Yazawa T, Inoue T. CT analysis of the effect of pirfenidone in patients with idiopathic pulmonary fibrosis. Eur J Radiol. 2014;83(1):32-8.

41.   Salisbury ML XM, Zhou Y, Murray S, Tayob N, Brown KK, Wells AU, Schmidt SL, Martinez FJ, Flaherty KR. Idiopathic Pulmonary Fibrosis: Gender-Age-Physiology Index Stage for Predicting Future Lung Function Decline. Chest. 2015. Chest. 2015;149:491-8.

42.   Sumikawa H, Johkoh T, Colby TV, Ichikado K, Suga M, Taniguchi H, Kondoh Y, Ogura T, Arakawa H, Fujimoto K, Inoue A, Mihara N, Honda O, Tomiyama N, Nakamura H, Muller NL. Computed tomography findings in pathological usual interstitial pneumonia: relationship to survival. Am J Respir Crit Care Med. 2008;177(4):433-9.

43.   Edey AJ, Devaraj AA, Barker RP, Nicholson AG, Wells AU, Hansell DM. Fibrotic idiopathic interstitial pneumonias: HRCT findings that predict mortality. Eur Radiol. 2011;21(8):1586-93.

44.   Quanjer PH. Standardized lung function testing. Eur Respir J - Suppl. 1993;6:1-100.

45.   Wells AU, Desai SR, Rubens MB, Goh NS, Cramer D, Nicholson AG, Colby TV, Du Bois RM, Hansell DM. Idiopathic pulmonary fibrosis: a composite physiologic index derived from disease extent observed by computed tomography. Am J Respir Crit Care Med. 2003;167:962-9.

46.   Ley B, Ryerson CJ, Vittinghoff E, Ryu JH, Tomassetti S, Lee JS, Poletti V, Buccioli M, Elicker BM, Jones KD, King Jr TE, Collard HR. A multidimensional

index and staging system for idiopathic pulmonary fibrosis. Ann Intern Med. 2012;156:684-91.

47.   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.

48.   Maldonado F, Moua T, Rajagopalan S, Karwoski RA, Raghunath S, Decker PA, Hartman TE, Bartholmai BJ, Robb RA, Ryu JH. Automated quantification of radiological patterns predicts survival in idiopathic pulmonary fibrosis. Eur Respir J. 2014;43(1):204-12.

49.   Jacob J, Bartholmai B, Rajagopalan S, Kokosi M, Nair A, Karwoski R, Raghunath S, S.L.F. W, Wells AU, Hansell DM. Automated quantitative CT versus visual CT scoring in idiopathic pulmonary fibrosis: validation against pulmonary function. J Thorac Imaging. 2016;31:304-11.

50.   Hu S, Hoffman EA, Reinhardt JM. Automatic lung segmentation for accurate quantitation of volumetric X-ray CT images. IEEE Trans Med Imaging. 2001;20(6):490-8.

51.   Shikata H, McLennan G, Hoffman EA, Sonka M. Segmentation of pulmonary vascular trees from thoracic 3D CT images. Int J Biomed Imaging. 2009:11.

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