ORIGINAL ARTICLE

Predicting Outcomes in Idiopathic Pulmonary Fibrosis Using Automated Computed Tomographic Analysis

Joseph Jacob[1,2], Brian J. Bartholmai[3], Srinivasan Rajagopalan[3], Coline H. M. van Moorsel[4,5], Hendrik W. van Es[6], Frouke T. van Beek[4], Marjolijn H. L. Struik[4,5], Maria Kokosi[7], Ryoko Egashira[8], Anne Laure Brun[9], Arjun Nair[10], Simon L. F. Walsh[11], Gary Cross[12], Joseph Barnett[12], Angelo de Lauretis[13], Eoin P. Judge[14], Sujal Desai[15], Ronald Karwoski[16], Sebastien Ourselin[17], Elisabetta Renzoni[7], Toby M. Maher[7], Andre Altmann[2], and Athol U. Wells[7]

[1]Department of Respiratory Medicine, [2]Centre for Medical Image Computing, and [10]Department of Radiology, University College London, London, United Kingdom; [3]Division of Radiology and [16]Department of Physiology and Biomedical Engineering, Mayo Clinic Rochester, Rochester, Minnesota; [4]St. Antonius ILD Center of Excellence, Department of Pulmonology, and [6]Department of Radiology, St. Antonius Hospital, Nieuwegein, the Netherlands; [5]Division of Heart and Lungs, University Medical Center Utrecht, Utrecht, the Netherlands; [7]Interstitial Lung Disease Unit and [15]Department of Radiology, Royal Brompton Hospital, Royal Brompton and Harefield NHS Foundation Trust, London, United Kingdom; [8]Department of Radiology, Faculty of Medicine, Saga University, Saga City, Japan; [9]Imaging Department, Hôpital Cochin, Paris-Descartes University, Paris, France; [11]Department of Radiology, King's College Hospital NHS Foundation Trust, London, United Kingdom; [12]Department of Radiology, Royal Free Hospital NHS Foundation Trust, London, United Kingdom; [13]Division of Pneumology, "Guido Salvini" Hospital, Garbagnate Milanese, Italy; [14]Department of Respiratory Medicine, Aintree University Hospital, Liverpool, United Kingdom; and [17]Translational Imaging Group, Centre for Medical Image Computing, University College London, London, United Kingdom

ORCID ID: 0000-0002-8054-2293 (J.J.).

Abstract

Rationale: Quantitative computed tomographic (CT) measures of baseline disease severity might identify patients with idiopathic pulmonary fibrosis (IPF) with an increased mortality risk. We evaluated whether quantitative CT variables could act as a cohort enrichment tool in future IPF drug trials.

Objectives: To determine whether computer-derived CT measures, specifically measures of pulmonary vessel–related structures (VRSs), can better predict functional decline and survival in IPF and reduce requisite sample sizes in drug trial populations.

Methods: Patients with IPF undergoing volumetric noncontrast CT imaging at the Royal Brompton Hospital, London, and St. Antonius Hospital, Utrecht, were examined to identify pulmonary function measures (including FVC) and visual and computer-derived (CALIPER [Computer-Aided Lung Informatics for Pathology Evaluation and Rating] software) CT features predictive of mortality and FVC decline. The discovery cohort comprised 247 consecutive patients, with validation of results conducted in a separate cohort of 284 patients, all fulfilling drug trial entry criteria.

Measurements and Main Results: In the discovery and validation cohorts, CALIPER-derived features, particularly VRS scores, were among the strongest predictors of survival and FVC 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 greater than 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 patients with IPF fulfilling drug trial entry criteria—the VRS score—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 FVC decline.

Keywords: idiopathic pulmonary fibrosis; quantitative computed tomographic imaging; pulmonary vessels

(Received in original form November 4, 2017; accepted in final form April 20, 2018)

Correspondence and requests for reprints should be addressed to Joseph Jacob, M.R.C.P., F.R.C.R., M.D.(Res.), Department of Respiratory Medicine, Rayne Institute, University College London, University Street, Fitzrovia, London, WC1E 6JF, UK. E-mail: j.jacob@ucl.ac.uk.

This article has an online supplement, which is accessible from this issue's table of contents at www.atsjournals.org.

Am J Respir Crit Care Med Vol 198, Iss 6, pp 767–776, Sep 15, 2018

Originally Published in Press as DOI: 10.1164/rccm.201711-2174OC on April 23, 2018
Internet address: www.atsjournals.org

At a Glance Commentary

Scientific Knowledge on the Subject: Quantification of computed tomographic (CT) parenchymal patterns in idiopathic pulmonary fibrosis using computer tools has been suggested as a method that can improve mortality prediction using visual CT scoring. As computer technology advances, it is becoming possible to study CT parenchymal features that have no visual correlate.

What This Study Adds to the Field: Our study demonstrates that computer-derived vessel-related structure scores can outperform current gold standard measures of outcome in idiopathic pulmonary fibrosis, such as FVC decline. Specifically, we demonstrate that using thresholds of computer-derived vessel-related structure scores for cohort enrichment can identify patients with idiopathic pulmonary fibrosis who 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.

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 FVC has been used as the primary measure of outcome in IPF (12–17), and it has been correlated with survival in several IPF cohorts (13, 15, 18–21). FVC testing is associated with a degree of measurement variation (22). Evidence from several clinical studies (15, 18–21) confirms that a 10% FVC decline threshold is a surrogate marker of mortality, and in the absence of an alternative explanation, it is an indicator of genuine disease progression (11).

Given the inexorable decline associated with late-stage IPF, characterization of indicators suggesting a poor prognosis at baseline in IPF is essential to guiding optimal management and allowing timely referral for lung transplant (23). In IPF drug trial settings, cohort enrichment strategies aim to create more homogeneous study populations and identify patients who are likely to experience increased clinical events (24), thereby reducing sample sizes and lessening the prohibitive costs associated with clinical trials (25).

Given the rapid recent advances in computer technology, the focus of imaging-based disease biomarkers in IPF has fallen primarily on computer analysis of computed tomographic (CT) imaging, which does not have the constraints of interobserver variation associated with visual CT scoring. New computer algorithms using three-dimensional volumetric CT datasets can quantify parenchymal pattern extents (26, 27) and have been shown to better predict survival in various fibrosing lung diseases at baseline than visual CT scores (28, 29). Computer analysis also has the potential to uncover CT features hitherto underrecognized by visual CT analysis that predict mortality in IPF. The quantitation of pulmonary vessels (arteries and veins) and associated structures (perivascular fibrosis), collectively termed vessel-related structures (VRSs) (Figure 1), which cannot be achieved by the human eye, has been shown to associate strongly with survival in a series of patients with IPF (28).

However, the utility of quantitation of VRSs in predicting survival in IPF has not been examined across multicentered patient cohorts or 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. Therefore, in our present study, we evaluated mortality prediction using pulmonary function tests (PFTs), composite indices, and visual and computer-based CT scoring in a discovery cohort of patients with IPF 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 derived 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. Some of the results of these studies were previously reported in the form of an abstract (30).

Methods

Study Design

All research subjects were diagnosed with IPF by a multidisciplinary team and had received a noncontrast volumetric CT scan as part of their clinical care. In this observational study, all consecutive patients with IPF presenting to the Royal Brompton Hospital between January 2007 and June 2011 were included in the discovery cohort. All patients with IPF presenting between July 2011 and December 2014 to the Royal Brompton Hospital and patients with IPF presenting to the St. Antonius Hospital were amalgamated to form the validation dataset. PFT, echocardiography, and CT protocols are outlined in the appendix in the online supplement. CALIPER (Computer-Aided Lung Informatics for Pathology Evaluation and Rating) CT analysis has been described previously (29, 31) and is outlined in the appendix in the online supplement. Approval of this study of clinically indicated computed tomography and pulmonary function data was obtained from the

J.J. was supported by Wellcome Trust Clinical Research Career Development Fellowship 209553/Z/17/Z. A.A. was supported by the Medical Research Council (grant MR/L016311/1). The work was supported by the National Institute of Health Research Respiratory Disease Biomedical Research Unit at the Royal Brompton and Harefield NHS Foundation Trust and Imperial College London. S.O. was partially funded by the National Institute for Health Research, University College London Hospitals Biomedical Research Centre.

Author Contributions: J.J., A.A., F.T.v.B., C.H.M.v.M., M.H.L.S., H.W.v.E., S.O., E.P.J., S.D., A.d.L., G.C., J.B., M.K., R.E., A.L.B., A.N., S.L.F.W., T.M.M., E.R., and A.U.W. were involved in acquisition, analysis, or interpretation of data; J.J. and A.U.W. were also involved in the conception and design of the study; B.J.B., R.K., and S.R. invented and developed CALIPER (Computer-Aided Lung Informatics for Pathology Evaluation and Rating software) and were involved in processing the raw computed tomographic scans and generation of figures but were not involved with the analysis or interpretation of the data. All authors revised the work for important intellectual content, gave final approval for the version to be published, and agree to be accountable for 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.

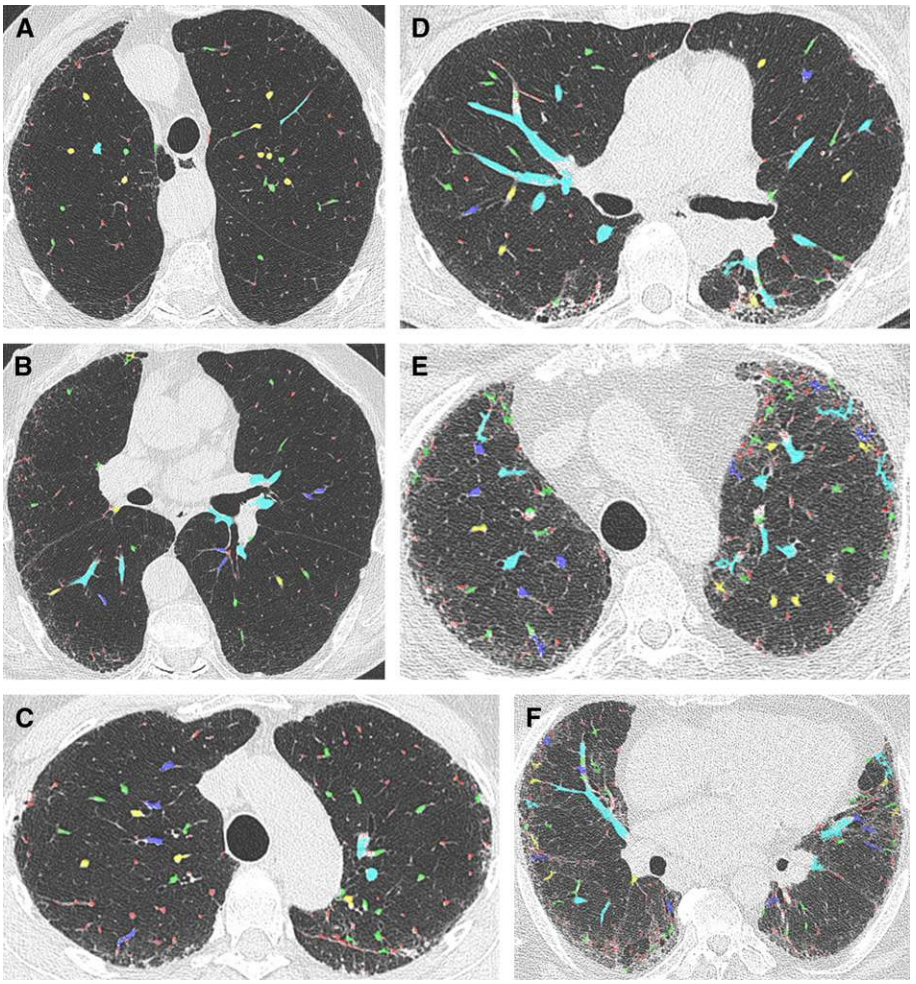

Figure 1. Color overlay computed tomographic images demonstrating CALIPER (Computer-Aided Lung Informatics for Pathology Evaluation and Rating software) vessel-related structures (CAL VRSs) of different sizes in the upper and middle lung zones in three patients with idiopathic pulmonary fibrosis with varying degrees of disease severity. (*A* and *B*) A 78-year-old male ex-smoker ($D_{L_{CO}}$, 69.7% predicted) demonstrates mild reticulation at the anterior and posterior aspects of the lungs. (*C* and *D*) A 56-year-old female ex-smoker ($D_{L_{CO}}$, 57.1% predicted) demonstrates coarse reticulation most marked in the apical segments of the lower lobes bilaterally. (*E* and *F*) A 68-year-old male ex-smoker ($D_{L_{CO}}$, 27.5% predicted) has more extensive reticulation in the lung periphery and at the lung bases. As fibrosis extent increases, so too does the CAL VRS score within the lungs. CAL VRS size key: red = <5 mm^2; green = 5–10 mm^2; yellow = 10–15 mm^2; dark blue = 15–20 mm^2; light blue = >20 mm^2.

Liverpool Research Ethics Committee (reference number 14/NW/0028) and the institutional ethics committees of the Royal Brompton Hospital (London, UK), the Mayo Clinic (Rochester, MN), and St. Antonius Hospital (Nieuwegein, the Netherlands). Informed patient consent was not required.

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 also summed ground-glass opacification. Visual scores also quantified traction bronchiectasis extent and severity. CALIPER also quantified pulmonary VRSs (*see* details in the appendix in the online supplement). Volumes for all CALIPER parenchymal features were converted into percentages using the total lung volume measured by CALIPER. The total pulmonary VRS score (CAL VRS) 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; $D_{L_{CO}}$; composite physiologic index [CPI]; gender, age, physiology [GAP] score; and GAP index), five visual CT scores (ILD extent, fibrosis extent, honeycombing extent, traction bronchiectasis severity, and traction bronchiectasis extent), four CALIPER CT parenchymal scores (ILD extent, fibrosis extent, honeycombing, and CALIPER total VRSs [CAL VRSs]), and 18 detailed VRS variables (the five different cross-sectional areas [<5 mm^2, 5–10 mm^2, 10–15 mm^2, 15–20 mm^2, and >20 mm^2] occurring in the upper, middle, or lower zones of the lungs, as well as 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 with more routinely acquired clinical and functional variables.

In our primary analyses, to compare the predictor variables, the $-\log_{10} P$ values for each measure from the discovery and validation sets were plotted on the *x*- and *y*-axes, respectively, with the axes labeled using the *P* value scale, and including both horizontal and vertical lines marking the Li and Ji corrected cutoffs 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). The Bonferroni method 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 in the present study (e.g., total pulmonary VRS [CAL VRS] and upper zone [UZ] VRS scores are strongly correlated; $R^2 = 0.77$). We therefore used the method of Li and Ji to estimate the effective number of independent tests (M_{eff}), which is derived

from the eigenvalues of the correlation matrix of all tested variables. We then used M_{eff} to adjust the significance threshold: $\alpha = 0.05/M_{eff}$. Essentially, the method of 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 PFTs. Percent change from baseline FVC (raw) was modeled 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 imaging, 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 of analyzing how the predictor affects the longitudinal percentage decline in FVC. The lmer function of the R package lme4 (33) was used for the analysis. We present the $-\log_{10} P$ values of the predictor-by-time interaction for the discovery and validation cohorts.

In addition to the linear mixed effects analysis, we calculated for each subject the presence or absence of a 10% decline in FVC at 12 months. More precisely, for each subject, we estimated a 10% FVC loss within 12 months based on best linear unbiased predictions from the longitudinal mixed model (a minimum of 4 mo of follow-up data was required). The presence of a 10% FVC decline at 12 months was defined on the basis of fitted trajectory (i.e., the predicted value at 12 mo was 10% less than the baseline FVC measurement).

We combined the endpoint of a 10% FVC decline at 12 months with a second endpoint of death within 12 months, and using logistic regression, we 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 described above, we present the $-\log_{10} P$ values of the predictor for the 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 mentioned above, we present the $-\log_{10} P$ values of the predictor for the 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 mo) and Cox mortality models. We further computed C-indices in the validation cohort by 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 contained FVC, D_{LCO}, or CPI.

Drug Trial Power Calculation

We conducted a power analysis to explore possible sample size reductions in clinical trials using CAL VRS scores as an enrichment parameter. The power analysis was based on a two-sample *t* test with a *P* value threshold of 0.05 and assuming 90% power. Furthermore, we assumed that the control group in an IPF drug trial would be receiving antifibrotic medication rather than a placebo and would decline, on average, by 120 ml/yr 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 antifibrotic trials (12, 13).

Three different drug effects were investigated: 25%, 40%, and 50% effect, leading to annual average declines in FVC of 90, 72, and 60 ml, respectively. To derive the effect size (Cohen's *d*), we computed the SD in the cohort with D_{LCO} greater than or equal to 30. We used the stable estimator for SD based on median absolute deviation, resulting in a value of 230.27. Thus, in the unenriched design, the effect sizes (Cohen's *d*) are 0.13, 0.21, and 0.26, respectively. For the enrichment design, we considered CAL VRS thresholds that retain 70%, 50%, and 30% of subjects with D_{LCO} greater than or equal to 30% predicted: 3.7, 4.4, and 5.1, respectively. Again, we used the stable estimator to compute the SD of annual FVC decline in these enriched subgroups: 228.1, 197.4, and 195.6, respectively. 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 antifibrotic studies in which these data are available (16, 37), SDs of FVC decline graphed or derived from graphed SEs did not differ materially between treatment and nontreatment arms.

The C-index was calculated for the logistic regression and mortality models that contained the CAL VRS thresholds in both the discovery and validation cohorts when patients fulfilled drug trial entry criteria (D_{LCO}, $\geqslant$30% predicted). We finally also examined the effect on study outcome measures of 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 patients with IPF fulfilling drug trial inclusion criteria (17). Accordingly, patients with a D_{LCO} percent predicted between 30% and 90% were analyzed in the discovery ($n = 163$) and validation ($n = 200$) cohorts. Two secondary analyses were also performed. First, after combining the discovery and validation cohorts, we separately examined patients with a D_{LCO} greater than or equal to 30% predicted who were not exposed to antifibrotic medication ($n = 200$) and those who had received antifibrotics ($n = 159$) to characterize the effects of medication on CT measures predicting FVC decline and survival. Antifibrotic use was categorized on an intention-to-treat basis. Second, we examined patients with a D_{LCO} less than 30% predicted in both the discovery ($n = 84$) and validation ($n = 84$) cohorts to evaluate the performance of computer tools in predicting the various study outcome measures in patients with severe disease.

Patients with a D_{LCO} greater than or equal to 30% predicted in the validation cohort had a marginally higher mean D_{LCO} than patients in the discovery cohort, with

no statistically significant difference identified in other baseline variables (*see* Table E1 in the online supplement). No statistically significant difference in survival curves was found between cohorts (Figures E1A and E1B).

FVC Decline

Across the entire discovery ($n = 181$) and validation ($n = 207$) cohorts, CALIPER variables (particularly UZ VRS and CALIPER fibrosis extents) 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 D_{LCO} greater than or equal to 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).

Ten Percent FVC Decline/Death at 12 Months

When all patients in the discovery ($n = 224$) and validation ($n = 251$) cohorts were evaluated, CAL 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 D_{LCO} greater than or equal to 30% 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, D_{LCO} and CPI were the most powerful predictors of survival. Total CAL VRS and UZ VRS scores also strongly predicted survival across both cohorts (Figure 2C). In patients fulfilling drug trial entry criteria (D_{LCO}, $\geq$30% predicted) ($n = 363$), functional indices (CPI and D_{LCO}), total and UZ VRS scores, and large VRS scores (>20 mm) in the middle and UZs best predicted survival (Figure 2F).

Subanalysis Related to Antifibrotic Administration

Analyses of patients who had never received antifibrotic medication but fulfilled drug trial entry criteria (D_{LCO}, $\geq$30% predicted) in the discovery ($n = 128$) and validation ($n = 72$) cohorts are demonstrated in Figures 2G–2I and Figures 3A–3C. UZ VRS variables were the strongest predictors of FVC decline (Figures 2G and 3A), 10% FVC decline/death at 12 months (Figures 2H and 3B), and survival (Figures 2I and 3C). Upper and midzone VRS variables outperformed functional indices for all three study outcome measures. Results were maintained when patients were stratified on the basis of median cohort D_{LCO} of 40.15% predicted (below median D_{LCO}, *see* Figures 3B, 3E, and 3H; above median D_{LCO}, *see* Figures 3C, 3F, and 3I).

In patients with a D_{LCO} greater than or equal to 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 (Figure E2D) and 10% FVC decline/death at 12 months (Figure E2E), and they outperformed D_{LCO} and CPI. UZ VRS subdivisions were similar to D_{LCO} and CPI in their ability to predict survival (Figure E2F).

C-Index Analyses

The relative strength of adjusted models predicting the combined endpoint of 10% FVC decline/death at 12 months was compared using the C-index for models containing CAL VRS subdivision scores, functional indices, and visual CT variables. All models were adjusted for patient age, sex, smoking status, and reconstruction algorithm in patients with a D_{LCO} greater than or equal to 30% predicted who had never received antifibrotics ($n = 175$). Models were examined separately in the combined cohorts, the discovery cohort, and the validation cohort (Table E2). Models in the combined cohorts containing CAL 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 CAL VRS subdivisions demonstrated higher C-indices than models containing FVC, D_{LCO}, and CPI in the combined cohort (Table E2). 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 (Table E3).

Patients with Severe Disease

When patients with severe disease (D_{LCO}, <30% predicted) were evaluated ($n = 168$), CPI and D_{LCO} were clearly the strongest predictors of survival (Figure E3C). These results were maintained in subanalysis of patients who received antifibrotic medication ($n = 29$) (Figure E2C). No variables strongly predicted FVC decline or a 10% FVC decline/death at 12 months in all study patients with a D_{LCO} less than 30% predicted (Figures E3A and E3B) or when those receiving antifibrotic medication were subanalyzed ($n = 27$) (Figures E2A and E2B).

When patients with a D_{LCO} less than 30% predicted who had never received antifibrotics were subanalyzed ($n = 139$), lower-zone VRS subdivisions were better predictors of FVC decline than D_{LCO} or CPI (Figure E3D). UZ VRS subdivisions demonstrated stronger relationships with 10% FVC decline/death at 12 months and survival than D_{LCO} or CPI in patients with severe disease who were not exposed to antifibrotics (Figures E3E and E3F).

Cohort Enrichment Using CAL 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 CAL VRS thresholds (Table 1). Using patients with IPF with a D_{LCO} greater than or equal to 30% predicted, we modeled sample size savings for a drug with three potential effect sizes on FVC decline (25%, 40%, and 50% reductions in FVC decline). At each drug effect size, we looked at cohort enrichment using three CAL VRS thresholds (representing the total VRS score throughout the entire lungs) corresponding to 70% (CAL VRSs = 3.7% of the lung), 50% (CAL VRSs = 4.4% of the lung), and 30% (CAL VRSs = 5.1% of the lung) of the original IPF study population.

As shown in Table 1, restricting a clinical trial cohort to patients with IPF with a CAL VRS threshold of 4.4% of the lung or greater, it would be possible to reduce the sample size by 26% to identify the same drug treatment effect size. Importantly, half of all patients with IPF studied were included in the CAL VRS threshold of greater than 4.4% of the lung. The model C-indices for the CAL VRS threshold of 4.4% in the discovery cohort of patients with a D_{LCO} greater than or equal to 30% predicted and not on antifibrotics

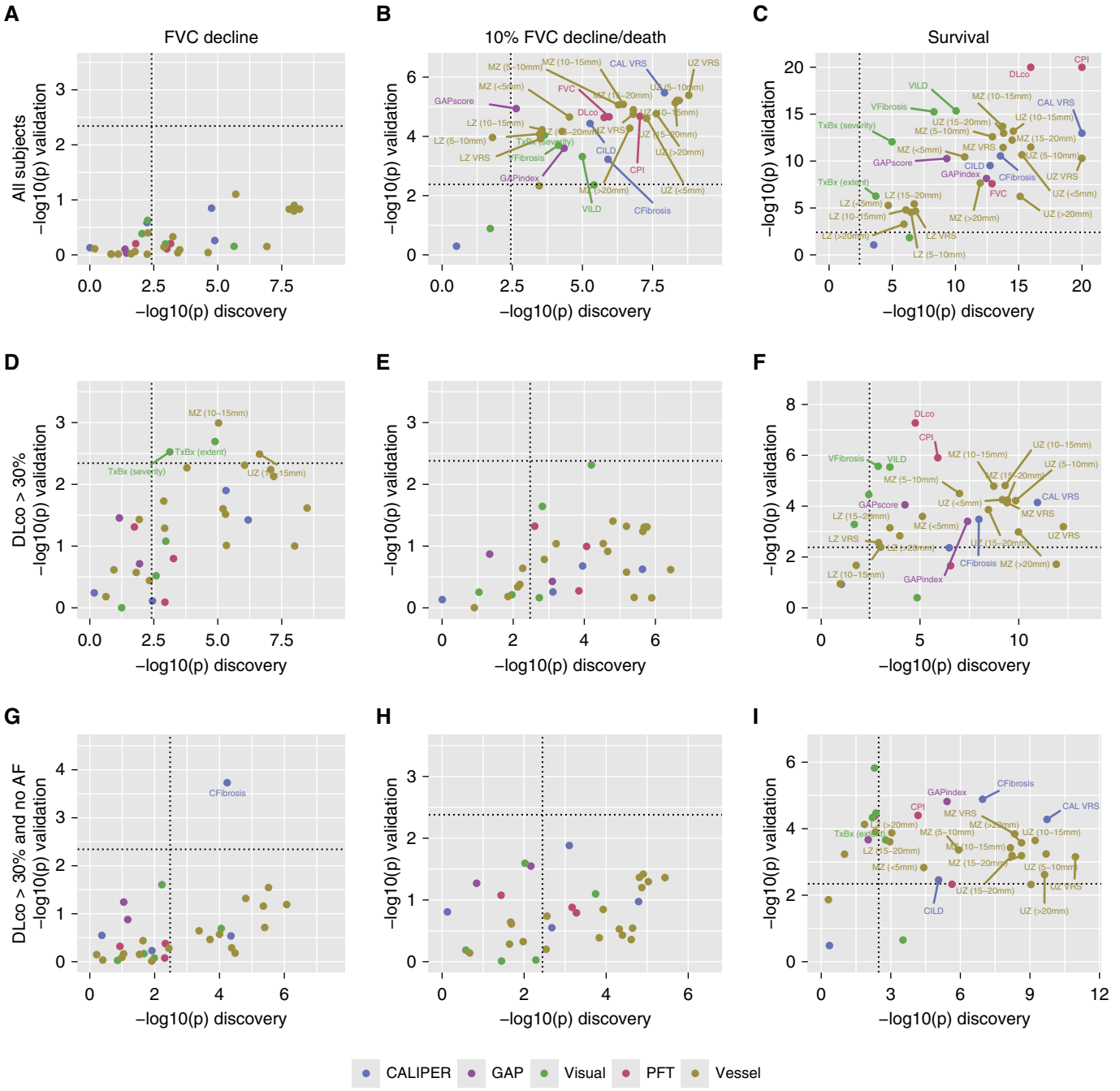

Figure 2. Scatterplots demonstrating $-\log_{10} P$ values for variables (various computer-derived CALIPER [Computer-Aided Lung Informatics for Pathology Evaluation and Rating software], visual computed tomography–derived, and pulmonary function indices) in the discovery cohort (*x*-axis) and validation cohort (*y*-axis). Horizontal and vertical dotted lines represent the Li and Ji corrected cutoffs for statistical significance. (*A–C*) All subjects. (*D–F*) Patients with D_{LCO} greater than or equal to 30% predicted. (*G–I*) Patients with D_{LCO} greater than or equal to 30% predicted who were not exposed to antifibrotic medication. The first column of *A*, *D*, and *G* represents variables predicting FVC decline. The second column of *B*, *E*, and *H* represents variables predicting a 10% FVC decline or death within 12 months. The third column of *C*, *F*, and *I* represents variables predicting survival. In *C*, to allow visualization of all the points on the figure, values that were infinite (owing to a *P* value of 0) were set to 20. The pulmonary vessel-related structure score was subdivided according to structure cross-sectional area (<5 mm^2, 5–10 mm^2, 10–15 mm^2, 15–20 mm^2, and >20 mm^2). AF = antifibrotic therapy; CAL VRS = CALIPER total vessel-related structure scores; CFibrosis = CALIPER fibrosis extent; CILD = CALIPER interstitial lung disease; CPI = composite physiologic index; GAP = gender, age, physiology; LZ = lower zone; MZ = middle zone; MZ VRS = middle zone vessel-related structure scores; PFT = pulmonary function test; TxBx = traction bronchiectasis (extent and severity); UZ = upper zone; UZ VRS = upper zone vessel-related structure scores; VFibrosis = visual fibrosis extent; VILD = visual interstitial lung disease.

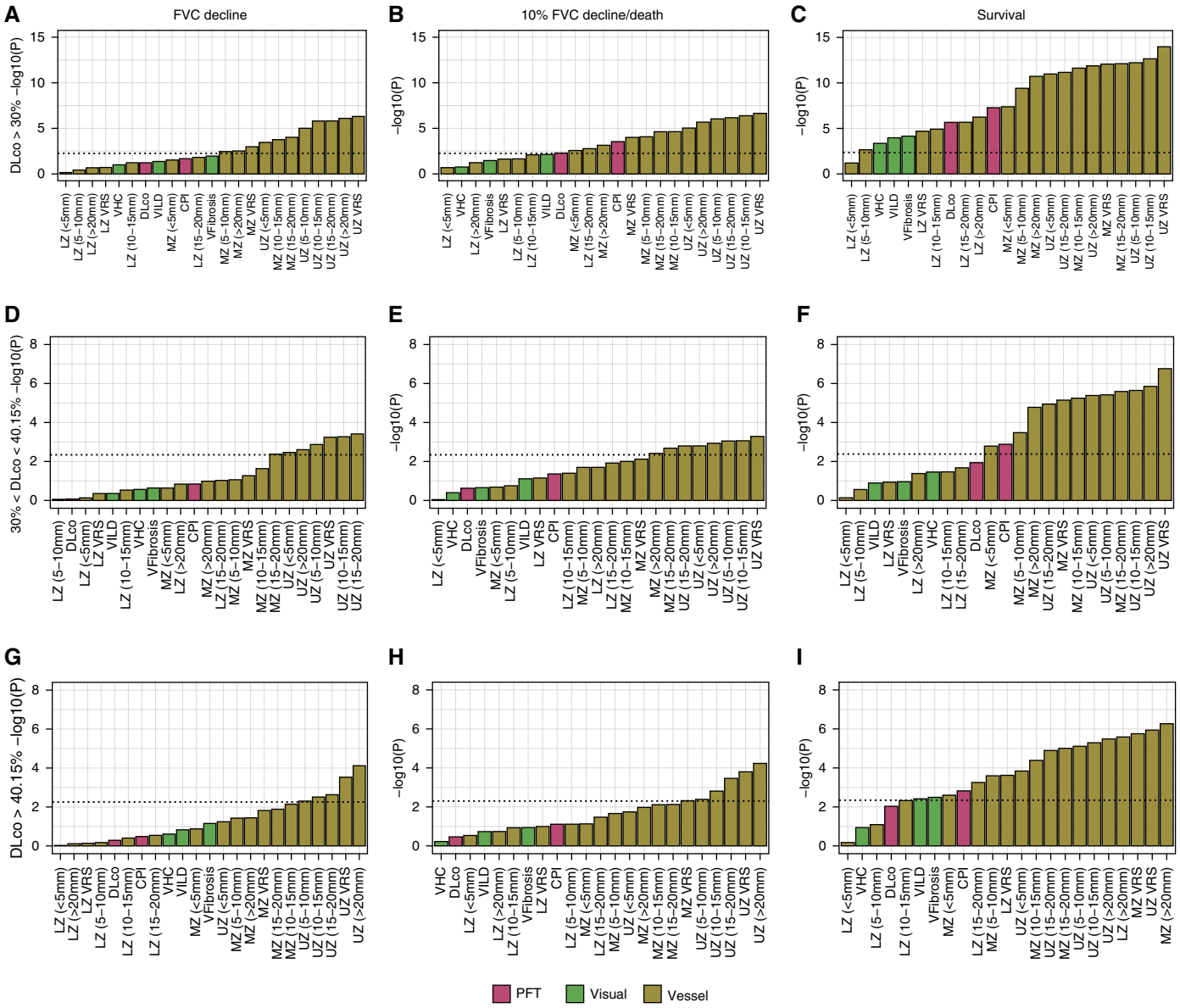

Figure 3. (*A–C*) $-\text{Log}_{10}$ *P* values for variables (various computer-derived CALIPER [Computer-Aided Lung Informatics for Pathology Evaluation and Rating software], visual computed tomography–derived, echocardiography-derived, and pulmonary function indices) in all study patients (discovery and validation cohorts) with a D_{LCO} greater than or equal to 30% predicted who were not exposed to antifibrotic medication. The patients were stratified with regard to disease severity on the basis of median D_{LCO} value for the combined cohort (40.15% predicted), with results for patients below the median D_{LCO} shown in *D–F* and patients above the median D_{LCO} shown in *G–I*. *A*, *D*, and *G* represent variables predicting FVC decline. *B*, *E*, and *H* represent variables predicting a 10% FVC decline or death within 12 months. *C*, *F*, and *I* represent variables predicting survival. The pulmonary vessel–related structure score was subdivided according to structure cross-sectional area (<5 mm^2, 5–10 mm^2, 10–15 mm^2, 15–20 mm^2, >20 mm^2). CPI = composite physiologic index; LZ = lower zone; LZ VRS = lower zone vessel-related structure scores; MZ = middle zone; MZ VRS = middle zone vessel-related structure scores; PFT = pulmonary function test; UZ = upper zone; UZ VRS = upper zone vessel-related structure scores; VFibrosis = visual fibrosis extent; VHC = visual honeycombing extent; VILD = visual interstitial lung disease.

($n = 128$) were 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 ($n = 72$ after exclusion of patients on antifibrotics and those with a D_{LCO} <30% predicted) were 0.56 for survival and 0.64 for the combined endpoint.

As a final analysis, to exclude the possibility that the CAL VRS threshold of 4.4% was identifying patients who were likely to progress regardless of any medical intervention, we examined the effect of antifibrotic medication in all study patients with a D_{LCO} greater than or equal to 30% predicted with a CAL VRS threshold greater than 4.4% ($n = 190$). Our results demonstrated that antifibrotic use in this subpopulation increased life expectancy (mortality analysis odds ratio, 0.437; 95% confidence interval [CI], 0.298–0.641; $Z = -4.239$; $P = 2.25 \times 10^{-5}$; 10% FVC decline/death in 12 months, odds ratio, 0.246; 95% CI, 0.122–0.498; $Z = -3.89$;

Table 1. Impact of Cohort Enriching an Idiopathic Pulmonary Fibrosis Drug Trial Study Population

CAL VRS Threshold (Percentage of Population)	Setting A (25% Effect)		Setting B (40% Effect)		Setting C (50% Effect)	
	n (Size)	Size Difference (%)	*n* (Size)	Size Difference (%)	*n* (Size)	Size Difference (%)
0 (100)	2,480	0	970	0	622	0
3.7 (70)	2,432	48 (2)	952	18 (2)	610	12 (2)
4.4 (50)	1,824	656 (26)	714	256 (26)	458	164 (26)
5.1 (30)	1,790	690 (28)	702	268 (28)	450	172 (28)

Definition of abbreviation: CAL VRS = Computer-Aided Lung Informatics for Pathology Evaluation and Rating software total vessel-related structures. Data were derived by analyzing baseline computed tomographic imaging and stratifying patients according to various thresholds of CAL VRSs. Statistical modeling 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 FVC decline. The number of patients in the unenriched cohort is demonstrated (CAL VRS threshold of 0), and the savings in drug trial sample size when using three CAL VRS thresholds (3.7%, 4.4%, and 5.1% of the lung equating to 70%, 50%, and 30% of the study population, respectively) are expressed as number and percentage.

$P = 9.73 \times 10^{-5}$) 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 CAL VRS thresholds are patients who would respond to antifibrotic medication and represent a good population with which to enrich an IPF drug trial.

Discussion

Our study across two cohorts of patients with IPF has demonstrated that computer-derived CT variables, particularly the CAL VRSs, predict two separate disease endpoints (survival and FVC decline at 12 months) and demonstrate an enhanced effect size in patients with IPF with less extensive disease. Most importantly, VRS scores can be used in cohort enrichment of an IPF population. Patients selected using a CAL VRS threshold of over 4.4% were shown to reduce an IPF drug trial sample size by 26% and were demonstrated to have an increased life expectancy and reduced rate of FVC decline when receiving antifibrotic medication compared with patients not receiving antifibrotics.

Existing computer algorithms that have approached the analysis of CT imaging in IPF have done so by characterizing and quantifying the standard library of CT parenchymal patterns recognized visually by radiologists for over 20 years (27, 38–40). The advent of volumetric CT datasets, however, has allowed three-dimensional structural information within the lungs to be analyzed across several hundred CT images. Furthermore, features such as the pulmonary CAL VRS scores 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 nontraditional 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.

PFTs, specifically $D_{L_{CO}}$ and CPI, were the strongest predictors of survival in patients with IPF with severe disease ($D_{L_{CO}}$, <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, was 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).

Although traction bronchiectasis has been recognized 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.

UZ VRS scores were among the strongest predictors of mortality in most analyses, whereas no powerful mortality signal was obtained from VRSs in the lower lung zones, regardless of the severity of underlying fibrosis. These results argue for the selective evaluation of VRSs in discrete lung subunits.

The potential utility of CALIPER as a cohort enrichment tool could help to reduce the prohibitive cost of modern drug trials in IPF. With patients in the control arms of new trials receiving antifibrotics and therefore demonstrating a slower decline in FVC when compared with the earlier placebo-controlled nintedanib (16) and pirfenidone (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 patients with IPF who 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 noncontrast volumetric imaging acquired using appropriate reconstruction algorithms. In the present study, most of the CT scans were obtained 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 centers. First, scrupulous attention would need to be paid to the quality of the CT scan acquired. An optimal full inspiratory breath at TLC is essential and could be aided with a form of spirometric gating. The desired scanner algorithms used to reconstruct the CT scan would be clearly specified to avoid severely edge-enhancing algorithms.

Once acquired, the anonymized volumetric noncontrast digital imaging and communications in medicine axial CT files would be electronically sent by a computed tomography technician to a central processing hub, such as the Biomedical Imaging Resource of the Mayo Clinic in Rochester, MN. Some preprocessing of CT data might be necessary before computer analysis to reduce image noise, and though this step can be inadvertently overlooked, its importance to ensure optimal parenchymal characterization cannot be overstated. Once ready for analysis, CALIPER processing of an entire CT examination takes less than one minute, though it 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, CAL VRS quantitation remained predictive of mortality across a spectrum of CT algorithms in the validation cohort, suggesting that CAL VRS measures may have a role in future multicenter studies.

The present study had limitations. 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, when fibrosis is least extensive in IPF, and retained utility across independent patient populations and a range of reconstruction algorithms, suggesting a robustness of the measure. Last, although the time intervals of FVC measurement in our study were not standardized 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 shows that in IPF, computer analysis of CT imaging, in particular quantitation of pulmonary VRS, can strongly predict survival and likelihood of FVC decline with effects enhanced over functional indices in patients with less extensive disease. Importantly, CAL VRS scores can selectively identify patients with IPF who will reach drug trial endpoints and respond to antifibrotic medication. CAL VRS scores may therefore have a major role in drug trial cohort enrichment, reducing the prohibitive costs of current IPF trials. ■

Author disclosures are available with the text of this article at www.atsjournals.org.

References

1. Navaratnam V, Fleming KM, West J, Smith CJP, Jenkins RG, Fogarty A, *et al*. The rising incidence of idiopathic pulmonary fibrosis in the U.K. *Thorax* 2011;66:462–467.
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;174:810–816.
3. Fernández Pérez ER, Daniels CE, Schroeder DR, St Sauver J, Hartman TE, Bartholmai BJ, *et al*. Incidence, prevalence, and clinical course of idiopathic pulmonary fibrosis: a population-based study. *Chest* 2010; 137:129–137.
4. Hutchinson J, Fogarty A, Hubbard R, McKeever T. Global incidence and mortality of idiopathic pulmonary fibrosis: a systematic review. *Eur Respir J* 2015;46:795–806.
5. Harari S, Madotto F, Caminati A, Conti S, Cesana G. Epidemiology of idiopathic pulmonary fibrosis in northern Italy. *PLoS One* 2016;11: e0147072.
6. Schwartz DA, Helmers RA, Galvin JR, Van Fossen DS, Frees KL, Dayton CS, *et al*. Determinants of survival in idiopathic pulmonary fibrosis. *Am J Respir Crit Care Med* 1994;149:450–454.
7. Flaherty KR, Thwaite EL, Kazerooni EA, Gross BH, Toews GB, Colby TV, *et al*. Radiological versus histological diagnosis in UIP and NSIP: survival implications. *Thorax* 2003;58:143–148.
8. Ley B, Collard HR, King TE Jr. Clinical course and prediction of survival in idiopathic pulmonary fibrosis. *Am J Respir Crit Care Med* 2011;183: 431–440.
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;42:483–489.
10. Nathan SD, Shlobin OA, Weir N, Ahmad S, Kaldjob JM, Battle E, *et al*. Long-term course and prognosis of idiopathic pulmonary fibrosis in the new millennium. *Chest* 2011;140:221–229.
11. Raghu G, Collard HR, Egan JJ, Martinez FJ, Behr J, Brown KK, *et al*.; ATS/ERS/JRS/ALAT Committee on Idiopathic Pulmonary Fibrosis. 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, *et al*. Relative versus absolute change in forced vital capacity in idiopathic pulmonary fibrosis. *Thorax* 2012;67:407–411.
13. du Bois RM, Weycker D, Albera C, Bradford WZ, Costabel U, Kartashov A, *et al*. Forced vital capacity in patients with idiopathic pulmonary fibrosis: test properties and minimal clinically important difference. *Am J Respir Crit Care Med* 2011;184:1382–1389.
14. Zappala CJ, Latsi PI, Nicholson AG, Colby TV, Cramer D, Renzoni EA, *et al*. Marginal decline in forced vital capacity is associated with a poor outcome in idiopathic pulmonary fibrosis. *Eur Respir J* 2010;35:830–836.
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 Respir Crit Care Med* 2003;168: 538–542.
16. Richeldi L, du Bois RM, Raghu G, Azuma A, Brown KK, Costabel U, *et al*.; INPULSIS Trial Investigators. Efficacy and safety of nintedanib in idiopathic pulmonary fibrosis. *N Engl J Med* 2014;370:2071–2082.
17. King TE Jr, Bradford WZ, Castro-Bernardini S, Fagan EA, Glaspole I, Glassberg MK, *et al*.; ASCEND Study Group. A phase 3 trial of pirfenidone in patients with idiopathic pulmonary fibrosis. *N Engl J Med* 2014;370:2083–2092.
18. Jegal Y, Kim DS, Shim TS, Lim C-M, Do Lee S, Koh Y, *et al*. Physiology is a stronger predictor of survival than pathology in fibrotic interstitial pneumonia. *Am J Respir Crit Care Med* 2005;171:639–644.
19. King TE Jr, Safrin S, Starko KM, Brown KK, Noble PW, Raghu G, *et al*. Analyses of efficacy end points in a controlled trial of interferon-gamma1b for idiopathic pulmonary fibrosis. *Chest* 2005;127:171–177.
20. Latsi PI, du Bois RM, Nicholson AG, Colby TV, Bisirtzoglou D, Nikolakopoulou A, *et al*. Fibrotic idiopathic interstitial pneumonia: the prognostic value of longitudinal functional trends. *Am J Respir Crit Care Med* 2003;168:531–537.
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; 168:543–548.

22. Pellegrino R, Viegi G, Brusasco V, Crapo RO, Burgos F, Casaburi R, *et al*. Interpretative strategies for lung function tests. *Eur Respir J* 2005;26:948–968.
23. Oldham JM, Noth I. Idiopathic pulmonary fibrosis: early detection and referral. *Respir Med* 2014;108:819–829.
24. Blackwell TS, Tager AM, Borok Z, Moore BB, Schwartz DA, Anstrom KJ, *et al*. Future directions in idiopathic pulmonary fibrosis research. An NHLBI workshop report. *Am J Respir Crit Care Med* 2014;189: 214–222.
25. Collard HR. Improving survival in idiopathic pulmonary fibrosis: the race has just begun. *Chest* 2017;151:527–528.
26. Bartholmai BJ, Raghunath S, Karwoski RA, Moua T, Rajagopalan S, Maldonado F, *et al*. Quantitative computed tomography imaging of interstitial lung diseases. *J Thorac Imaging* 2013;28:298–307.
27. Kim HJ, Brown MS, Chong D, Gjertson DW, Lu P, Kim HJ, *et al*. 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;22:70–80.
28. Jacob J, Bartholmai B, Rajagopalan S, Kokosi M, Nair A, Karwoski R, *et al*. Mortality prediction in idiopathic pulmonary fibrosis: evaluation of computer-based CT analysis with conventional severity measures. *Eur Respir J* 2017;49:1601011.
29. 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.
30. Jacob J, Bartholmai B, Altmann A, de Lauretis A, Rajagopalan S, Kokosi M, *et al*. Predicting time to decline in FVC using baseline visual and computer-based CT analysis and baseline functional indices [abstract]. *Clin Radiol* 2017;72(Suppl 1):S24.
31. Jacob J, Bartholmai BJ, Rajagopalan S, Kokosi M, Nair A, Karwoski R, *et al*. Automated quantitative computed tomography versus visual computed tomography scoring in idiopathic pulmonary fibrosis: validation against pulmonary function. *J Thorac Imaging* 2016;31: 304–311.
32. Li J, Ji L. Adjusting multiple testing in multilocus analyses using the eigenvalues of a correlation matrix. *Heredity (Edinb)* 2005;95: 221–227.
33. Bates D, Mächler M, Bolker B, Walker S. Fitting linear mixed-effects models using lme4. *J Stat Softw* 2015;67:1–48.
34. Therneau TM. A package for survival analysis in S. Version 2.38. 2015.
35. R Core Team. A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing; 2015.
36. Harrell FE Jr. Regression modeling strategies: with applications to linear models, logistic regression, and survival analysis. New York: Springer-Verlag; 2001.
37. Richeldi L, Costabel U, Selman M, Kim DS, Hansell DM, Nicholson AG, *et al*. Efficacy of a tyrosine kinase inhibitor in idiopathic pulmonary fibrosis. *N Engl J Med* 2011;365:1079–1087.
38. Salisbury ML, Lynch DA, van Beek EJ, Kazerooni EA, Guo J, Xia M, *et al*.; IPFnet Investigators. Idiopathic pulmonary fibrosis: the association between the adaptive multiple features method and fibrosis outcomes. *Am J Respir Crit Care Med* 2017;195:921–929.
39. Park HJ, Lee SM, Song JW, Lee SM, Oh SY, Kim N, *et al*. Texture-based automated quantitative assessment of regional patterns on initial CT in patients with idiopathic pulmonary fibrosis: relationship to decline in forced vital capacity. *AJR Am J Roentgenol* 2016;207:976–983.
40. Iwasawa T, Ogura T, Sakai F, Kanauchi T, Komagata T, Baba T, *et al*. CT analysis of the effect of pirfenidone in patients with idiopathic pulmonary fibrosis. *Eur J Radiol* 2014;83:32–38.
41. Salisbury ML, Xia M, Zhou Y, Murray S, Tayob N, Brown KK, *et al*. Idiopathic pulmonary fibrosis: gender-age-physiology index stage for predicting future lung function decline. *Chest* 2016;149: 491–498.
42. 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; 177:433–439.
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;21:1586–1593.