# Automated quantification of radiological patterns predicts survival in idiopathic pulmonary fibrosis

Fabien Maldonado[1,5], Teng Moua[1,5], Srinivasan Rajagopalan[2],
Ronald A. Karwoski[2], Sushravya Raghunath[2], Paul A. Decker[3],
Thomas E. Hartman[4], Brian J. Bartholmai[4], Richard A. Robb[2]
and Jay H. Ryu[1]

**Affiliations:**
[1]Division of Pulmonary and Critical Care, Mayo Clinic Rochester, Rochester, MN,
[2]Dept of Physiology and Biomedical Engineering, Mayo Clinic Rochester, Rochester, MN,
[3]Dept of Biomedical Statistics and Informatics, Mayo Clinic Rochester, Rochester, MN, and
[4]Division of Radiology, Mayo Clinic Rochester, Rochester, MN, USA.
[5]These authors contributed equally.

**Correspondence:** F. Maldonado, Division of Pulmonary and Critical Care Medicine, Gonda 18 South Mayo Clinic,
200 1st St. SW, Rochester, MN 55905, USA. E-mail: Maldonado.Fabien@mayo.edu

ABSTRACT   Accurate assessment of prognosis in idiopathic pulmonary fibrosis remains elusive due to significant individual radiological and physiological variability. We hypothesised that short-term radiological changes may be predictive of survival.

We explored the use of CALIPER (Computer-Aided Lung Informatics for Pathology Evaluation and Rating), a novel software tool developed by the Biomedical Imaging Resource Laboratory at the Mayo Clinic Rochester (Rochester, MN, USA) for the analysis and quantification of parenchymal lung abnormalities on high-resolution computed tomography. We assessed baseline and follow-up (time-points 1 and 2, respectively) high-resolution computed tomography scans in 55 selected idiopathic pulmonary fibrosis patients and correlated CALIPER-quantified measurements with expert radiologists' assessments and clinical outcomes.

Findings of interval change (mean 289 days) in volume of reticular densities (hazard ratio 1.91, p=0.006), total volume of interstitial abnormalities (hazard ratio 1.70, p=0.003) and per cent total interstitial abnormalities (hazard ratio 1.52, p=0.017) as quantified by CALIPER were predictive of survival after a median follow-up of 2.4 years. Radiologist interpretation of short-term global interstitial lung disease progression, but not specific radiological features, was also predictive of mortality.

These data demonstrate the feasibility of quantifying interval short-term changes on high-resolution computed tomography and their possible use as independent predictors of survival in idiopathic pulmonary fibrosis.

@ERSpublications
**Short-term quantified CT changes are predictive of survival in IPF** http://ow.ly/qmbjd

## Introduction

Baseline measures of disease severity have been used to predict survival in idiopathic pulmonary fibrosis (IPF) and include clinical features, pulmonary function measures, computed tomography (CT) findings and composite scoring indices [1–4]. While these measures have been validated in several studies, they appear inferior to dynamic surrogates of disease progression, such as longitudinal changes in physiological indices, *i.e.* decline in forced vital capacity (FVC) and diffusing capacity of the lung for carbon monoxide ($D$LCO) over 6–12 months [2, 5, 6]. However, the utility of such longitudinal physiological indices for individual patient management is still significantly limited by substantial intra-individual variability, perhaps with confounding by coexisting conditions, such as emphysema or pulmonary hypertension [7, 8].

CALIPER (Computer-Aided Lung Informatics for Pathology Evaluation and Ratings) is an image analysis tool developed by the Biomedical Imaging Resource Laboratory at the Mayo Clinic Rochester (Rochester, MN, USA) for the characterisation and quantification of lung parenchymal findings on high-resolution computed tomography (HRCT). The detection and quantification of pulmonary parenchyma by CALIPER is based on histogram signature mapping techniques trained through expert radiologist consensus assessment of pathologically confirmed training sets obtained through the Lung Tissue Research Consortium (LTRC) [9]. We hypothesised that this novel computer-aided method of analysing pulmonary parenchymal features on chest HRCT would provide a reproducible and accurate assessment of disease that may correlate with the semi-quantitative assessment of radiologists. We further postulated that the overall extent or short-term changes in parenchymal features as detected by CALIPER or radiologists may correlate with and be independently predictive of survival in patients with IPF.

## Materials and methods

### Patient and CT selection

Selected patients evaluated at the Mayo Clinic from January 2000 to December 2010, with IPF as diagnosed according to the latest international consensus guideline [10] (with or without surgical lung biopsy), and for whom at least two serial HRCTs had been obtained within a 3–15-month interval were included, representing time-points 1 and 2. HRCTs obtained at the time of acute exacerbation, infection, fluid overload or thromboembolic disease were excluded from the study. In addition, the HRCT had to satisfy technical requirements for CALIPER, which include high-resolution images ($\leqslant$2-mm slice thickness). The scans in this study were acquired using LightSpeed Ultra (BONE kernel; GE Healthcare, Cleveland, OH, USA) and Model Sensation 64 (B46 kernel; Siemens, Munich, Germany) scanners.

Collected demographic data included age, sex and smoking history with number of pack-years. Time of diagnosis was defined as the first date of contact with a clinical care provider who made or suspected diagnosis of IPF based on assessment of clinical presentation, HRCT scan and/or biopsy. Patients were followed to end-points of death or last known visit date recorded in the medical record at Mayo Clinic. Patients who underwent lung transplantation were followed until the transplant date and censored, with interval HRCT changes prior to this included in the analysis.

The study protocol was approved by the Mayo Clinic institutional review board (IRB number 10-000320). All included subjects provided informed consent.

### Pulmonary function studies

All included patients had corresponding pulmonary function test (PFT) data obtained within 30 days of the selected scans. Spirometry was performed using a Puritan Bennett Renaissance pneumotachography-based flow spirometer (Mallinckrodt, St Louis, MO, USA) according to standards set by the American Thoracic Society/European Respiratory Society guidelines [11]. Flows were expressed as % predicted using the reference equations of MILLER *et al.* [12].

### Radiologists' semiquantitative assessment of CT

For comparison of radiologists' semiquantitative assessment of HRCT findings with CALIPER quantification, two subspecialty thoracic radiologists (T.E. Hartman and B.J. Bartholmai) independently reviewed the HRCT scans unblinded to time-points 1 and 2 for each patient and scored them in a standardised manner while blinded to the clinical data and CALIPER results. The scoring system consisted of a visual estimation of the percentage of parenchymal involvement, to the nearest 5%, in a particular region for four radiological abnormalities: emphysema, ground-glass opacities, reticular opacities and honeycombing. The distribution of abnormalities was scored separately within each of 12 regions: the central and peripheral zones of right upper, right middle, right lower, left upper, lingula and left lower lobes. The reviewing radiologists also recorded their assessment of interval stability, progression or improvement of overall CT abnormalities between the two scans of each subject. The percentage of regional involvement

for each abnormality and total involvement (summing the radiologist's assessment for each region) of each of the radiological parameters (emphysema, ground glass, reticulation and honeycombing) was used for statistical analysis.

### CALIPER analysis
#### Data processing
The data processing step included lung extraction and segmentation with identification and extraction of central airways and central vascular structures, manual segmentation of the anatomical lobes, and classification of the remaining pulmonary parenchyma. Segmentation of the lungs was achieved using an adaptive density-based morphological approach [13]. Airways were segmented using iterative three-dimensional region growing and connected components analysis. Pulmonary vessels were extracted using an optimised multi-scale tubular structure enhancement filter [14]. Additional sub-segmentation into anatomical lobes by manual tracing technique was performed by a thoracic radiologist (B.J. Bartholmai) and determination of central/peripheral regions was completed by three-dimensional volume erosion techniques, such that the peripheral zone represented ~50% of the total volume of each lung. Separate inclusion of the perihilar regions into the central zone of each lobe was performed by semi-automated determination of the hila through tracheal tree analysis and inclusion of a spherical region of 5 cm, bilaterally, from the hilar points. Parenchymal tissue type detection and quantification of the classes (normal, emphysema, ground glass, reticular and honeycombing) (fig. 1) was performed using a sliding window supervised classification scheme.

During the pre-processing supervised training phase, multiple (n=976) $15 \times 15 \times 15$-pixel volumes of interest were selected through independent analysis by four subspecialty thoracic radiologists from HRCT scans of 14 subjects with proven pathological diagnosis of diffuse pulmonary disease (interstitial lung disease (ILD) of emphysema) or known control subjects without ILD. The exemplar candidates with agreement on the class of abnormality by all four radiologists were used to determine canonical histogram signatures for each of the classes of visual abnormality with automatic cluster affinity techniques and those
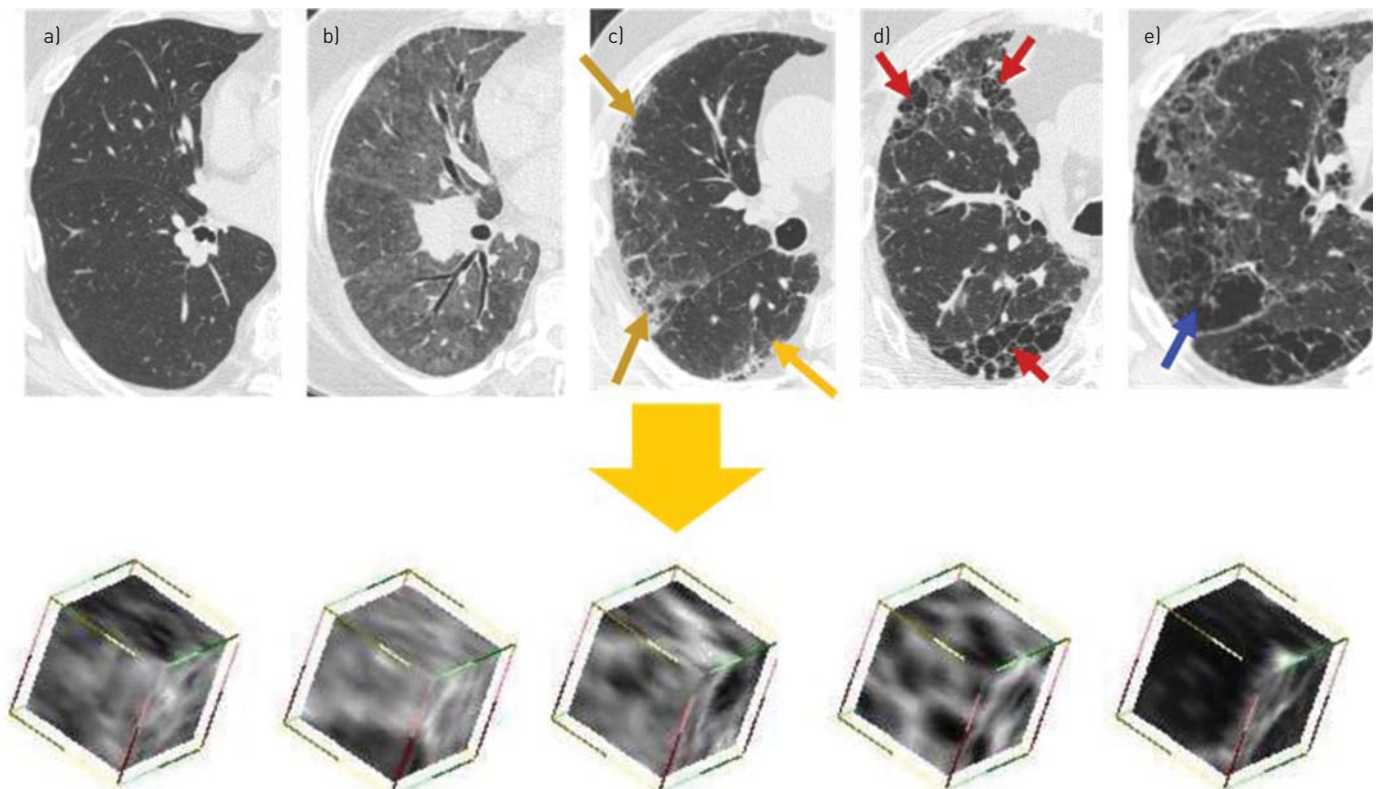


FIGURE 1 Computed tomography images demonstrating a) normal appearance and various visual manifestations of diffuse lung disease: b) ground glass, c) reticular changes (arrows), d) honeycombing (arrows) and e) emphysema (arrow)). In training datasets, the consensus of four thoracic radiologists was used to identify multiple volumes of interest of $15 \times 15 \times 15$ pixels corresponding to normal, ground-glass density, reticular abnormalities, honeycombing and emphysema.

FIGURE 2 Representative results from the quantitative analysis of CALIPER. The two rows (a–c) and (d–f) correspond to the two time-points from one of the pati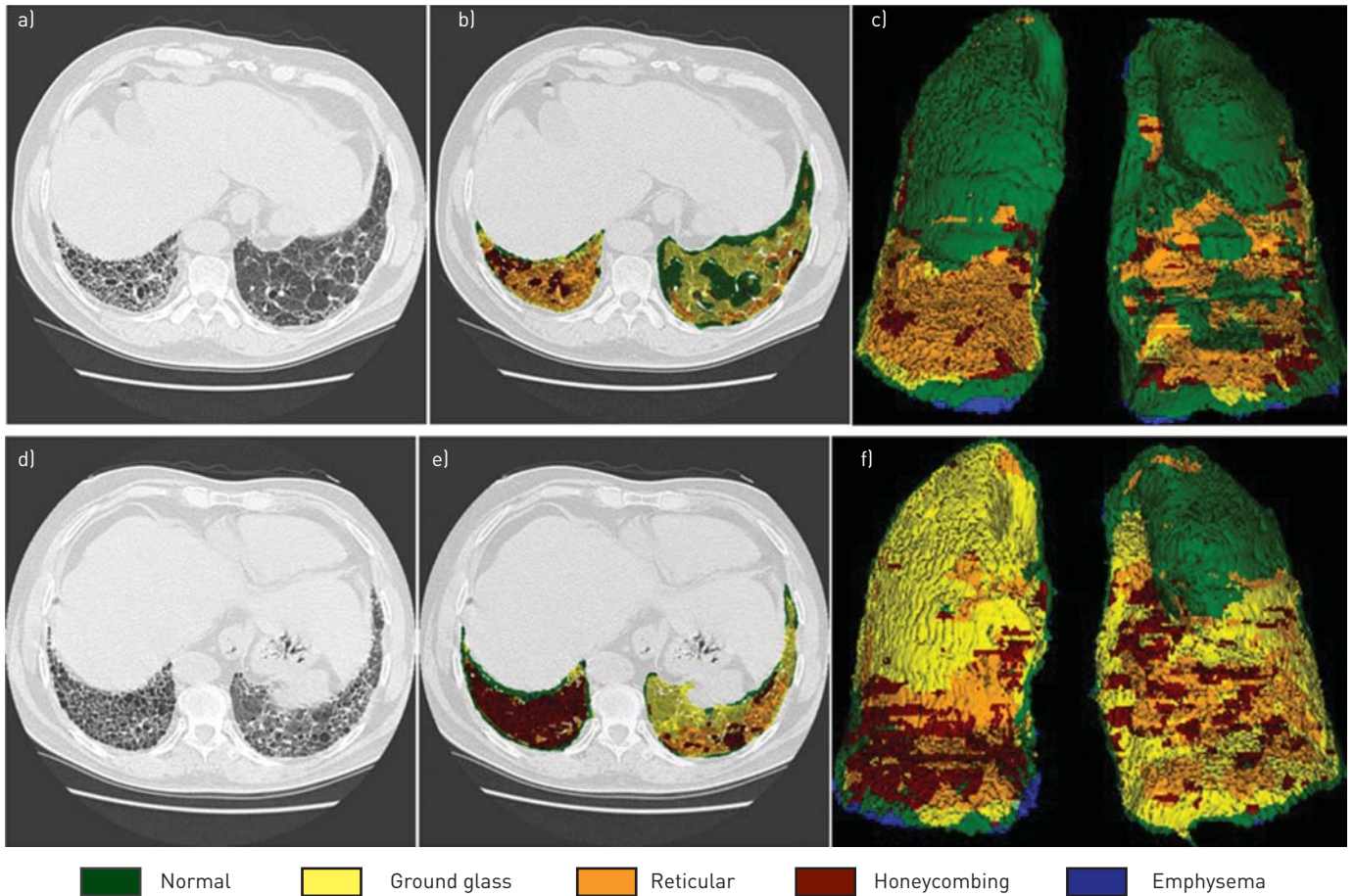ents. The individual voxels of the lung regions in the original sections (a and d) are classified and colour-coded into one of the classes of visible abnormalities (b and e). c and f) Three-dimensional maximum feature projections. CALIPER quantification and radiologist reviews were both indicative of an overall increase in ground-glass opacities, reticular abnormalities and honeycombing.

signatures of each of the visual classes were used for the volumetric classification of the HRCT data of test subjects (fig. 2).

### Tissue quantification

CALIPER analysis involved algorithmic identification and volumetric quantification of five radiological parenchymal features: normal lung, emphysema, ground-glass density, reticular abnormalities and honeycombing measured in total litres for the whole lung. Total ILD was defined as the volumetric sum of total ground-glass density, reticular abnormalities and honeycombing. Percentage ILD was defined as the ratio of the sum of the total ILD divided by the CALIPER segmented total lung parenchymal volume.

### Statistical analysis

Associations between CALIPER measures, PFTs and radiologists' estimation of percentage involvement were assessed using Spearmans' correlation coefficient. Changes in CALIPER measures between the two CT scans of each subject were calculated for total volume of normal lung, ground-glass density, reticular abnormalities, honeycombing, total lung volume, total ILD and percentage ILD. The association of change in CALIPER measures and change in radiologists' assessments of regional semi-quantitative findings and overall impression of progression, regression or stability of disease with survival was assessed using Cox proportional hazards regression. A landmark survival analysis was performed with survival indexed from the date of the second CT, a method we chose in order to avoid survival bias conditioned on living to the second CT. Both univariable and multivariable analyses including sex, pack-years, baseline FVC % pred, baseline $D$LCO % pred and time between CTs were performed. In all cases, two-sided p-values <0.05 were considered statistically significant.

| TABLE 1 Demographic data[#] | |
|---|---|
| Age years | 72.4±6.9 |
| Females/males | 28 (51)/27 (49) |
| Height cm | 168.0±9.4 |
| Smoking history | |
|   Active | 1 (2) |
|   Previous | 34 (62) |
|   Never | 20 (36) |
| Transplanted | 2 (4) |
| Surgical lung biopsy confirmation | 31 (56) |
| Mean interval between HRCT scans days | 289.2±109.7 (55–540) |
| Median follow-up years (range) | 2.4 (0.1–8.5) |
| Deaths | 37 (67) |
| Scan type 110 HRCT scans | |
|   GE 83 | 75 |
|   Siemens 27 | 25 |

Data are presented as mean± SD, n (%), mean± SD (range) or n, unless otherwise stated. HRCT: high-resolution computed tomography. [#]: total subjects n=55.

## Results

### Patient and HRCT characteristics

55 patients were included in the analysis with a mean± SD age of 72.4±6.9 years and a sex distribution of 51% males (n=28) (table 1). 31 (56%) of the 55 patients had undergone surgical lung biopsy to confirm usual interstitial pneumonia pathology, with the remainder diagnosed on the basis of typical usual interstitial pneumonia pattern on HRCT and clinical exclusion of known causes of pulmonary fibrosis. The median transplant-free survival was 2.1 years (range 1.1–3.4 years) from the second HRCT time-point. For the 18 subjects not known to be deceased, the median follow-up time was 2.4 years (range 0.1–8.5 years). Mean± SD time between serial HRCT was 289.2±109 days. Although the study did not exclude subjects based on differences in the brand or model of scanner between the two scans, 75% of subjects had studies performed on the same brand of scanner at both time-points (Siemens and GE Healthcare).

### PFTs

PFT values (absolute and % pred values) corresponding to the two CT scan time-points are presented in table 2 along with % pred differences over the interval.

### CALIPER analysis

Mean volumetric quantification by CALIPER in litres for each specific radiological parameter at time-points 1 and 2 are presented in table 3 along with respective differences over the time interval. As CALIPER quantifies the total volumetric proportion of ground glass, reticulation and honeycombing in patients presenting with varying stages of IPF severity, mean total and percentage changes over time rather than single time-point volumes were used in survival analysis. Radiologist interpretation of radiological findings at time-points 1 and 2 are presented in table 4. Mean percentages represent scoring to the closest 5% for

| TABLE 2 Raw and % predicted pulmonary function test values corresponding with computed tomography scans at time-points 1 and 2 | | | |
|---|---|---|---|
| Parameter | Time-point 1 | Time-point 2 | Difference± SD |
| TLC % pred | 67.3±9.6 | 65.9±9.7 | -3.4±5.6 |
| TLC L | 3.9±0.8 | 3.9±0.8 | -0.2±0.3 |
| FVC % pred | 67.5±14.1 | 64.1±16.8 | -3.4±8.2 |
| FVC L | 2.4±0.8 | 2.3±0.8 | -0.1±0.3 |
| FEV1 % pred | 72.0±14.7 | 68.9±16.9 | -3.1±8.5 |
| FEV1 L | 2.0±0.6 | 1.9±0.6 | -0.1±0.2 |
| $D_{LCO}$ % pred | 50.6±14.5 | 45.7±12.4 | -7.0±11.6 |
| $D_{LCO}$ mL·min$^{-1}$·mmHg$^{-1}$ | 11.7±4.0 | 10.7±3.0 | -1.5±2.6 |

Data are presented as mean± SD. TLC: total lung capacity; FVC: forced vital capacity; FEV1: forced expiratory volume in 1 s; $D_{LCO}$: diffusing capacity of the lung for carbon monoxide.

TABLE 3 Mean volumetric measurements and changes for computed tomography findings at time-points 1 and 2

| Finding | Time-point 1 | Time-point 2 | Difference $\pm$ SD |
|---|---|---|---|
| Total normal L | 2.5 $\pm$ 1.0 | 2.2 $\pm$ 1.1 | -0.3 $\pm$ 0.6 |
| Total emphysema L | 0.02 $\pm$ 0.04 | 0.01 $\pm$ 0.03 | -0.002 $\pm$ 0.03 |
| Total emphysema % | 0.4 $\pm$ 1.2 | 0.4 $\pm$ 1.0 | 0.0 $\pm$ 0.9 |
| Total GGO L | 0.3 $\pm$ 0.3 | 0.4 $\pm$ 0.4 | 0.08 $\pm$ 0.3 |
| Total GGO[#] % | 10.2 $\pm$ 10.6 | 13.3 $\pm$ 12.8 | 3.2 $\pm$ 9.8 |
| Total reticular L | 0.5 $\pm$ 0.3 | 0.6 $\pm$ 0.4 | 0.06 $\pm$ 0.2 |
| Total reticular[#] % | 15.1 $\pm$ 9.9 | 18.1 $\pm$ 12.4 | 3.0 $\pm$ 6.4 |
| Total honeycombing L | 0.2 $\pm$ 0.3 | 0.2 $\pm$ 0.3 | -0.01 $\pm$ 0.2 |
| Total honeycombing[#] % | 6.0 $\pm$ 6.9 | 6.0 $\pm$ 8.8 | -0.1 $\pm$ 6.3 |
| Total ILD L | 1.1 $\pm$ 0.6 | 1.2 $\pm$ 0.6 | 0.1 $\pm$ 0.4 |
| Total ILD[#] % | 31.3 $\pm$ 16.8 | 37.4 $\pm$ 22.1 | 6.1 $\pm$ 13.0 |
| Total lung volume L | 3.6 $\pm$ 0.8 | 3.5 $\pm$ 0.9 | -0.2 $\pm$ 0.4 |

Data are presented as mean $\pm$ SD, unless otherwise stated. GGO: ground-glass opacities; ILD: interstitial lung disease (ground glass + reticular + honeycombing). [#]: % of total lung volume.

each region and averaged across 12 regions for a percentage of that radiological finding in the whole lung. Results of univariable analysis adjusted for sex, pack-years, baseline FVC % pred and baseline $D$LCO % pred are presented in table 5. A statistically significant association with survival was noted for changes in $D$LCO % pred (hazard ratio (HR) 2.14, 95% CI 1.27–3.60; p=0.004) and total lung capacity (HR 4.17, 95% CI 1.42–12; p=0.009) on analysis of pulmonary function testing. Adjusted analysis of changes over time for CALIPER measured percent ILD (HR 1.52, 95% CI 1.08–2.15; p=0.017), total ILD volume (HR 1.70, 95% CI 1.19–2.43; p=0.003) and total reticulation volume (HR 1.91, 95% CI 1.21–3.0; p<0.006) were associated with survival (table 4).

The correlation between total lung volume as measured by CALIPER and total lung capacity measured by PFT was very good (r=0.77 (p<0.001) and r=0.87 (p<0.001) for time-points 1 and 2, respectively).

*Radiologist correlation with CALIPER and survival*

Interobserver agreement between two expert radiologists and CALIPER is reported in the online supplementary tables S3 and S4. Overall, correlation between the two radiologists for ILD scoring (ground-glass opacities, reticular and honeycombing) was moderate to substantial in all regions (range 0.33, p<0.001 to 0.73, p<0.001 for time-point 1, and 0.33, p<0.001 to 0.77, p<0.001 for time-point 2). Correlations were overall best for ground glass and reticular findings in all lobes and poorest for honeycombing. Correlation between radiologist 1 and CALIPER for ILD scoring was mild to moderate (range 0.29, p=0.001 to 0.63, p<0.005 for time-point 1 and 0.24, p<0.001 to 0.64, p<0.001 for time-point 2) and mild to moderate (range 0.29, p<0.001 to 0.48, p<0.001 for time-point 1 and 0.28, p<0.001 to 0.64, p<0.001 for time-point 2) for radiologist 2. After adjustment for smoking history (pack-years), sex, FVC and $D$LCO % pred, and time between HRCT scans, change in individual ILD parenchymal findings as assessed by radiologists were

TABLE 4 Radiologists' assessment of computed tomography findings at time-points 1 and 2

| Finding | Time-point 1 | Time-point 2 | Difference $\pm$ SD |
|---|---|---|---|
| **Radiologist 1** | | | |
| Total emphysema % | 1.3 $\pm$ 4.0 | 1.3 $\pm$ 3.9 | 0.0 $\pm$ 0.7 |
| Total GGO % | 14.9 $\pm$ 17.5 | 17.5 $\pm$ 20.7 | 2.6 $\pm$ 8.5 |
| Total reticular % | 17.6 $\pm$ 9.6 | 19.9 $\pm$ 10.9 | 2.3 $\pm$ 4.6 |
| Total honeycombing % | 3.5 $\pm$ 7.9 | 3.7 $\pm$ 8.4 | 0.2 $\pm$ 1.0 |
| **Radiologist 2** | | | |
| Total emphysema % | 1.3 $\pm$ 3.9 | 1.4 $\pm$ 3.9 | 0.1 $\pm$ 0.5 |
| Total GGO % | 22.2 $\pm$ 20.3 | 24.8 $\pm$ 24.4 | 2.6 $\pm$ 11.0 |
| Total reticular % | 18.2 $\pm$ 8.2 | 18.6 $\pm$ 8.5 | 0.4 $\pm$ 1.9 |
| Total honeycombing % | 1.7 $\pm$ 6.2 | 1.7 $\pm$ 6.3 | 0.1 $\pm$ 0.5 |

Data are presented as mean $\pm$ SD, unless otherwise stated. GGO: ground-glass opacities.

TABLE 5 Survival analysis of pulmonary function tests and CALIPER measurement

| Difference | Univariable | | | Adjusted | | |
|---|---|---|---|---|---|---|
| | HR | 95% CI | p-value | HR | 95% CI | p-value |
| TLC % pred[#] | 1.19 | 0.67–2.12 | 0.55 | 4.17 | 1.42–12.0 | 0.009 |
| FVC % pred[#] | 1.10 | 0.76–1.62 | 0.60 | 1.54 | 0.95–2.51 | 0.081 |
| FEV1 % pred[#] | 1.21 | 0.82–1.79 | 0.33 | 1.46 | 0.93–2.29 | 0.099 |
| $D_{LCO}$ % pred[#] | 1.42 | 0.93–2.16 | 0.11 | 2.14 | 1.27–3.60 | 0.004 |
| Total emphysema[¶] | 1.19 | 0.87–1.65 | 0.28 | 1.29 | 0.92–1.79 | 0.14 |
| Total emphysema[¶] % | 1.10 | 0.85–1.41 | 0.48 | 1.16 | 0.87–1.54 | 0.32 |
| Total GG[¶] | 1.40 | 1.05–1.85 | 0.022 | 1.30 | 0.88–1.90 | 0.19 |
| Total GG[¶] % | 1.29 | 0.98–1.70 | 0.072 | 1.21 | 0.85–1.73 | 0.28 |
| Total reticular[¶] | 1.35 | 0.91–1.99 | 0.14 | 1.91 | 1.21–3.0 | 0.006 |
| Total reticular[¶] % | 1.38 | 1.03–1.84 | 0.032 | 1.93 | 1.30–2.89 | 0.001 |
| Total honeycombing[¶] | 0.95 | 0.74–1.21 | 0.68 | 1.03 | 0.80–1.31 | 0.84 |
| Total honeycombing[¶] % | 0.94 | 0.74–1.20 | 0.62 | 1.02 | 0.80–1.29 | 0.90 |
| Total ILD[¶] | 1.47 | 1.08–2.0 | 0.016 | 1.70 | 1.19–2.43 | 0.003 |
| Total ILD[¶] % | 1.30 | 0.99–1.71 | 0.062 | 1.52 | 1.08–2.15 | 0.017 |
| Emphysema radiologist 1[¶] | 1.32 | 0.89–1.97 | 0.17 | 1.32 | 0.88–1.98 | 0.18 |
| Emphysema radiologist 2[¶] | 0.82 | 0.52–1.29 | 0.38 | 1.48 | 0.79–2.77 | 0.22 |
| GG radiologist 1[¶] | 1.16 | 0.86–1.57 | 0.32 | 1.40 | 0.97–2.02 | 0.070 |
| GG radiologist 2[¶] | 1.02 | 0.77–1.35 | 0.88 | 1.24 | 0.92–1.67 | 0.16 |
| Honeycombing radiologist 1[¶] | 1.07 | 0.74–1.54 | 0.73 | 1.12 | 0.72–1.75 | 0.61 |
| Honeycombing radiologist 2[¶] | 1.10 | 0.79–1.54 | 0.56 | 1.11 | 0.75–1.65 | 0.60 |
| Reticular radiologist 1[¶] | 1.11 | 0.84–1.47 | 0.46 | 1.20 | 0.87–1.64 | 0.27 |
| Reticular radiologist 2[¶] | 1.14 | 0.70–1.84 | 0.60 | 1.03 | 0.63–1.68 | 0.92 |
| **Change radiologist 1** | | | | | | |
| Improved/stable | Ref. | | | Ref. | | |
| Worse | 2.1 | 1.09–4.04 | 0.026 | 2.22 | 1.08–4.56 | 0.029 |
| **Change radiologist 2** | | | | | | |
| Improved/stable | Ref. | | | Ref. | | |
| Worse | 1.86 | 0.96–3.61 | 0.066 | 2.50 | 1.12–5.54 | 0.025 |

CALIPER: Computer-Aided Lung Informatics for Pathology, Evaluation and Ratings; HR: hazard ratio; TLC: total lung capacity; % pred; % predicted; FVC: forced vital capacity; FEV1: forced expiratory volume in 1 second; $D_{LCO}$: diffusing capacity of the lung for carbon monoxide; ILD: interstitial lung disease (ground glass (GG) + reticular + honeycombing); Ref: reference.[#]: HRs are for a 15-unit change in $D_{LCO}$ % pred and a 10-unit change in FEV1 % pred, FVC % pred and TLC % pred; [¶]: HRs are for a SD change.

not predictive of survival (table 5). Radiologist interpretation of overall global change in terms of progression of ILD was predictive of survival.

## Discussion

CALIPER represents an automated volumetric quantification tool for assessing specific parenchymal radiological features on HRCT. In our study of IPF patients, CALIPER measured short-term (3–15 months) reticular changes, and percentage and total ILD changes were predictive of survival. Correlation between radiologists was moderate to substantial regarding ground glass and reticular findings when estimating to the nearest 5%, and mild to moderate between radiologists and CALIPER for those same ILD findings. No specific parenchymal estimates of ILD were predictive of survival with radiologist assessment, although overall global assessment of disease progression or change by radiologists was predictive in our study.

The absence of a gold standard to validate accuracy of CALIPER volumetric measurements to specific regions in a pathological specimen led us to attempt to correlate longitudinal changes in quantitative estimates of radiological fibrosis to patient outcomes, specifically mortality. As the quantitative assessment of fibrosis by CALIPER is not influenced by confounding conditions such as emphysema or pulmonary hypertension that may affect pulmonary function measures, we felt radiological features may reflect more directly the progression of fibrotic processes and could represent a novel and promising tool in the assessment and management of patients with IPF.

The mild-to-moderate correlation between CALIPER measurements and estimates of extent of fibrosis by two expert radiologists in interstitial lung diseases is reassuring and suggests that CALIPER may indeed

allow both accurate and reproducible assessments, though CALIPER's semiquantitative measurements of specific parenchymal ILD features may prove more advantageous when subtle changes are not readily recognised between serial scans. Estimation using a scoring system to the nearest 5% allowed for moderate-to-substantial correlation between two experienced radiologists in our study, but only had mild-to-moderate correlation with CALIPER. We note that while algorithmic quantitative analysis by CALIPER may detect subtle abnormalities that radiologists may not, misclassification of abnormalities with similar density but different morphology (such as honeycombing *versus* emphysema) may occur with CALIPER or differ slightly from the radiologist assessment (ground glass *versus* reticular abnormality), explaining a difference in correlation.

Interestingly, radiologists' global assessment of short-term ILD progression was also predictive of survival in our study. We know, in general, that radiologists' quantitative assessments and diagnostic interpretation are often inconsistent with one another, although our degree of radiologist correlation was higher. Perhaps this may be explained by our reviewing radiologists being subspecialists in thoracic radiology at the same institution, and both having reviewed and standardised cases and terminology as training for other ILD research studies. Nonblinding to time-points 1 and 2 may have increased radiologist vigilance for expected change between interval scans and biased interpretation of disease progression. Nonetheless, this conclusion is gratifying in regards to experienced radiologists detecting progression over shorter time intervals as being statistically predictive of survival, although detection of such subtle changes may not be reproducible or consistent across all practices and institutions. A quantitative method such as CALIPER may provide this consistency, particularly as detection of subtle progression over shorter time intervals may be valuable in estimating survival and perhaps be used as a marker of treatment response in future clinical trials. We are encouraged that correlation of our reproducible quantitative assessments of disease with outcomes is significant, independent of correlation with the subjective descriptions of disease features by a radiologist.

Prognostication of IPF remains challenging due to the paucity of validated surrogate markers of disease progression and an unpredictable natural history. The use of physiological data as surrogate markers has significant limitations. First, physiological measures, such as FVC and $D_{LCO}$, are characterised by significant intra-individual variability. In fact, the threshold of 10% decline in FVC is below the degree of normal intra-individual variability, as suggested by recent guidelines [15]. Secondly, physiological measures provide only indirect estimates of the progression of fibrosis, and may be affected by coexisting emphysema and/or pulmonary hypertension that are frequently associated with IPF [7, 8]. Lastly, PFTs may not be sensitive enough to detect subclinical progression of fibrosis: FVC decline as defined by current thresholds is a relatively rare event in IPF clinical trials, which has led some authors to suggest that marginal declines in FVC may be more sensitive, though less specific [16].

A number of other studies suggest that the extent of fibrosis assessed semiquantitatively on HRCT is a strong predictor of outcomes in IPF [1, 4, 17]. Correlation of quantitative estimates of pulmonary fibrosis by automated analysis to mortality has been the object of few reports [1, 18, 19]. This line of research has been hampered by the lack of validation methodology and poor access to high-quality HRCT allowing for volumetric assessment of the lung parenchyma. Most of the available data on quantitative analysis of lung fibrosis have focused on the predictive power of baseline HRCT abnormalities.

The use of longitudinal quantitative indices of lung fibrosis as identified on HRCT could represent an appealing alternative to physiological measures. An accurate and reproducible method allowing monitoring of fibrosis progression on HRCT would be a valuable surrogate marker of disease. Unfortunately, assessment of fibrosis volumes by expert radiologists has been hampered by substantial intra- and interobserver variability, and quantitative CT indices using fractal analysis and global histogram-based methods have not been validated or found helpful in clinical practice thus far [1, 4, 19–24]. CALIPER is based on a texture-sensitive volumetric analysis that allows automated classification of lung parenchyma according to a database of HRCT volumes of interest validated by radiologists using data from the LTRC [9]. The majority of existing expert systems and associated quantitative tools depend on strictly controlled image acquisition protocols to provide consistent results. We believe careful selection of training sets for CALIPER enables more reproducible classification, whose greyscale local histogram-based algorithms are less affected by image noise, reconstruction kernel and other scan parameters.

While our preliminary results are promising, we recognise the limitations of our study. First, the CALIPER technology was developed based on "lung pattern signatures" derived from standardised CT analyses and acquisition protocols used in the LTRC database [9]. Scan parameters used for CT in this retrospective study were different to those used in the LTRC database and were not always identical for the scans at the two time-points. Despite differences in acquisition parameters for some of the data sets, the mild-to-moderate correlation between radiologist semiquantitative assessment and CALIPER analysis supports the validity of our results. We postulate that our histogram signature-based method may be more robust and less sensitive to specific slice thickness or image reconstruction parameters than other texture-based or pixel

counting techniques. Secondly, the requirement of two serial HRCT obtained solely for follow-up purposes led us to exclude patients with HRCT obtained for coexisting illnesses (heart failure, infection and acute exacerbation) and others with only one HRCT available. This exclusion criteria arguably limits the external validity of our study, as included patients were more likely to represent a subset of IPF patients with gradual decline rather than stable (less likely to have repeat HRCT) or unstable patients (more likely to be lost to follow-up, die or experience acute exacerbation). However, the fact that the median survival of included patients was similar to that typically seen for patients with IPF is reassuring in this regard. Finally, the number of patients included was small, due to stringent eligibility criteria. While small, the numbers are comparable to those used in prior studies evaluating the value of longitudinal trends in physiological measures. We recognise these limitations and believe that further validation of our preliminary results is warranted, including prospective analysis of standardised HRCT with equivalent time intervals, comparison of IPF abnormalities to those of other ILD and application of short-term changes found with CALIPER to predicting acute exacerbation or the presence of related complications such as pulmonary hypertension.

In conclusion, we have shown that CALIPER characterisation and quantification of lung parenchyma on HRCT correlates with visual assessment by expert radiologists, and that quantitative short-term volumetric longitudinal changes on serial HRCT correlate with IPF mortality. We believe that quantitative features of ILD on HRCT determined by CALIPER may represent an accurate and reproducible biomarker for IPF that warrants further validation studies and application.

## References

1   Best AC, Meng J, Lynch AM, *et al.* Idiopathic pulmonary fibrosis: physiologic tests, quantitative CT indexes, and CT visual scores as predictors of mortality. *Radiology* 2008; 246: 935–940.

2   Flaherty KR, Mumford JA, Murray S, *et al.* Prognostic implications of physiologic and radiographic changes in idiopathic interstitial pneumonia. *Am J Respir Crit Care Med* 2003; 168: 543–548.

3   Lama VN, Flaherty KR, Toews GB, *et al.* Prognostic value of desaturation during a 6-minute walk test in idiopathic interstitial pneumonia. *Am J Respir Crit Care Med* 2003; 168: 1084–1090.

4   Wells AU, Desai SR, Rubens MB, *et al.* Idiopathic pulmonary fibrosis: a composite physiologic index derived from disease extent observed by computed tomography. *Am J Respir Crit Care Med* 2003; 167: 962–969.

5   Collard HR, King TE Jr, Bartelson BB, *et al.* Changes in clinical and physiologic variables predict survival in idiopathic pulmonary fibrosis. *Am J Respir Crit Care Med* 2003; 168: 538–542.

6   Latsi PI, du Bois RM, Nicholson AG, *et al.* Fibrotic idiopathic interstitial pneumonia: the prognostic value of longitudinal functional trends. *Am J Respir Crit Care Med* 2003; 168: 531–537.

7   Cottin V, Cordier JF. The syndrome of combined pulmonary fibrosis and emphysema. *Chest* 2009; 136: 1–2.

8   Nathan SD, Noble PW, Tuder RM. Idiopathic pulmonary fibrosis and pulmonary hypertension: connecting the dots. *Am J Respir Crit Care Med* 2007; 175: 875–880.

9   Zavaletta VA, Bartholmai BJ, Robb RA. High resolution multidetector CT-aided tissue analysis and quantification of lung fibrosis. *Acad Radiol* 2007; 14: 772–787.

10  Raghu G, Collard HR, Egan JJ, *et al.* An Official ATS/ERS/JRS/ALAT Statement: Idiopathic Pulmonary Fibrosis: Evidence-based Guidelines for Diagnosis and Management. *Am J Respir Crit Care Med* 2011; 183: 788–824.

11  Miller MR, Hankinson J, Brusasco V, *et al.* Standardisation of spirometry. *Eur Respir J* 2005; 26: 319–338.

12  Miller A, Thornton JC, Warshaw R, *et al.* Single breath diffusing capacity in a representative sample of the population of Michigan, a large industrial state. Predicted values, lower limits of normal, and frequencies of abnormality by smoking history. *Am Rev Respir Dis* 1983; 127: 270–277.

13  Hu S, Hoffman EA, Reinhardt JM. Automatic lung segmentation for accurate quantitation of volumetric X-ray CT images. *IEEE Trans Med Imaging* 2001; 20: 490–498.

14  Shikata H, McLennan G, Hoffman EA, *et al.* Segmentation of pulmonary vascular trees from thoracic 3D CT images. *Int J Biomed Imaging* 2009; 2009: 636240.

15  Pellegrino R, Viegi G, Brusasco V, *et al.* Interpretative strategies for lung function tests. *Eur Respir J* 2005; 26: 948–968.

16  Zappala CJ, Latsi PI, Nicholson AG, *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.

17  Sumikawa H, Johkoh T, Colby TV, *et al.* Computed tomography findings in pathological usual interstitial pneumonia: relationship to survival. *Am J Respir Crit Care Med* 2008; 177: 433–439.

18  Best AC, Lynch AM, Bozic CM, *et al.* Quantitative CT indexes in idiopathic pulmonary fibrosis: relationship with physiologic impairment. *Radiology* 2003; 228: 407–414.

19  Iwasawa T, Asakura A, Sakai F, *et al.* Assessment of prognosis of patients with idiopathic pulmonary fibrosis by computer-aided analysis of CT images. *J Thorac Imaging* 2009; 24: 216–222.

20  Coxson HO, Hogg JC, Mayo JR, *et al.* Quantification of idiopathic pulmonary fibrosis using computed tomography and histology. *Am J Respir Crit Care Med* 1997; 155: 1649–1656.

21  Kauczor HU, Heitmann K, Heussel CP, *et al.* Automatic detection and quantification of ground-glass opacities on high-resolution CT using multiple neural networks: comparison with a density mask. *AJR Am J Roentgenol* 2000; 175: 1329–1334.

22  Kazerooni EA, Martinez FJ, Flint A, *et al.* Thin-section CT obtained at 10-mm increments *versus* limited three-level thin-section CT for idiopathic pulmonary fibrosis: correlation with pathologic scoring. *AJR Am J Roentgenol* 1997; 169: 977–983.

23  Lynch DA, Godwin JD, Safrin S, *et al.* High-resolution computed tomography in idiopathic pulmonary fibrosis: diagnosis and prognosis. *Am J Respir Crit Care Med* 2005; 172: 488–493.

24  Uppaluri R, Hoffman EA, Sonka M, *et al.* Interstitial lung disease: a quantitative study using the adaptive multiple feature method. *Am J Respir Crit Care Med* 1999; 159: 519–525.