Mortality prediction in idiopathic pulmonary fibrosis: evaluation of computer-based CT analysis with conventional severity measures

Joseph Jacob[1], Brian J. Bartholmai[2], Srinivasan Rajagopalan[3], Maria Kokosi[4], Arjun Nair[1], Ronald Karwoski[3], Simon L.F. Walsh[1], Athol U. Wells[4] and David M. Hansell[1]

Affiliations: [1]Dept of Radiology, Royal Brompton Hospital, Royal Brompton and Harefield NHS Foundation Trust, London, UK. [2]Division of Radiology, Mayo Clinic Rochester, Rochester, MN, USA. [3]Dept of Physiology and Biomedical Engineering, Mayo Clinic Rochester, Rochester, MN, USA. [4]Interstitial Lung Disease Unit, Royal Brompton Hospital, Royal Brompton and Harefield NHS Foundation Trust, London, UK.

Correspondence: Joseph Jacob, 10 Wolsey Road, Northwood, Middlesex, HA6 2HW, UK.
E-mail: akelajacob@gmail.com

@ERSpublications
CALIPER-derived parameters such as pulmonary vessel volume are more accurate prognostically than visual CT scores http://ow.ly/2b5G304exlA

Cite this article as: Jacob J, Bartholmai BJ, Rajagopalan S, *et al.* Mortality prediction in idiopathic pulmonary fibrosis: evaluation of computer-based CT analysis with conventional severity measures. *Eur Respir J* 2017; 49: 1601011 [https://doi.org/10.1183/13993003.01011-2016].

ABSTRACT   Computer-based computed tomography (CT) analysis can provide objective quantitation of disease in idiopathic pulmonary fibrosis (IPF). A computer algorithm, CALIPER, was compared with conventional CT and pulmonary function measures of disease severity for mortality prediction.

CT and pulmonary function variables (forced expiratory volume in 1 s, forced vital capacity, diffusion capacity of the lung for carbon monoxide, transfer coefficient of the lung for carbon monoxide and composite physiologic index (CPI)) of 283 consecutive patients with a multidisciplinary diagnosis of IPF were evaluated against mortality. Visual and CALIPER CT features included total extent of interstitial lung disease, honeycombing, reticular pattern, ground glass opacities and emphysema. In addition, CALIPER scored pulmonary vessel volume (PVV) while traction bronchiectasis and consolidation were only scored visually. A combination of mortality predictors was compared with the Gender, Age, Physiology model.

On univariate analyses, all visual and CALIPER-derived interstitial features and functional indices were predictive of mortality to a 0.01 level of significance. On multivariate analysis, visual CT parameters were discarded. Independent predictors of mortality were CPI (hazard ratio (95% CI) 1.05 (1.02–1.07), p<0.001) and two CALIPER parameters: PVV (1.23 (1.08–1.40), p=0.001) and honeycombing (1.18 (1.06–1.32), p=0.002). A three-group staging system derived from this model was powerfully predictive of mortality (2.23 (1.85–2.69), p<0.0001).

CALIPER-derived parameters, in particular PVV, are more accurate prognostically than traditional visual CT scores. Quantitative tools such as CALIPER have the potential to improve staging systems in IPF.

Introduction

Accurate prognostication is central to the management of patients with idiopathic pulmonary fibrosis (IPF). In addition to informing a patient of their probable life expectancy [1], accurate prediction of a patient's likely clinical course allows the institution of appropriate management, which presently may include antifibrotic medication, referral for transplantation [2] or palliative care pathways.

However, prognostication in IPF is fraught with difficulties. An array of prognostic indicators have been used over the years, but with varying degrees of success in IPF. Pulmonary function tests (PFTs) such as diffusion capacity of the lung for carbon monoxide (D_{LCO}) are perhaps the most sensitive markers of disease severity, but are associated with a measurement variation of 10–15% per test [3]. Visual computed tomographic (CT) evaluation is subject to interobserver variation [4, 5]. Composite indices have been proposed, but are yet to be fully validated [6]. As a result, new tools that may be more accurate in predicting a patient's prognosis are required. A recent recommendation from the Fleischner Society [7] has emphasised computer-based quantitative CT analysis as a potential outcome measure in IPF.

Establishing whether a quantitative tool is a suitable marker of disease outcome requires the evaluation of the tool against other markers of baseline disease severity. A sophisticated quantitative CT algorithm (CALIPER) has been shown to have better correlations with PFTs than semiquantitative visual CT evaluation [8].

The aims of our study were to compare CALIPER, visual CT scoring and PFTs against survival in IPF. Optimal stratification was explored using CALIPER variables, visual variables, PFTs and the Gender, Age, Physiology (GAP) score.

Methods

A retrospective analysis of an interstitial lung disease database identified all consecutive newly attending patients receiving a multidisciplinary team diagnosis of IPF according to published guidelines [9], over a 4.5-year period (January 2007 to July 2011). Patients with a departmental, noncontrast, supine, volumetric CT were included in the study cohort (figure 1). CT, echocardiography and PFT protocols are included in the online supplementary material, as are details of CALIPER CT evaluation. The Digital Imaging and Communications in Medicine images for the CT scans were transferred to the Mayo Clinic (Rochester, MN, USA) for blinded CALIPER processing. Approval for this study of clinically indicated CT and pulmonary function data was obtained from the institutional ethics committees of the Royal Brompton Hospital (London, UK) and Mayo Clinic and informed patient consent was not required.

Visual CT evaluation

Each CT scan was evaluated independently by two radiologists with 5 and 7 years of thoracic imaging experience, blinded to all clinical information. An initial training dataset of 15 nonstudy cases was used to help to identify pre-existing biases. The scores of the test cases were reviewed and the most widely discrepant results discussed with a third radiologist.

CTs were scored on a lobar basis using a continuous scale. The total extent of interstitial lung disease (ILD) was initially estimated to the nearest 5%, then subclassified into four patterns: reticular, ground glass opacification, honeycombing and consolidation, using definitions from the Fleischner Society glossary of terms for thoracic imaging [10]. 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. Furthermore, the percentage (to the nearest 5%) of each lobe that contained mosaicism (decreased attenuation component) or emphysema was recorded. The individual lobar percentages of each parenchymal pattern were summed for each radiologist and divided by six to create an averaged lobar score per pattern, per scorer, per case.

Traction bronchiectasis, as defined in the Fleischner Society glossary of terms [10], was assigned with a categorical "severity" score that 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 a Gestalt score of none (0), mild (1), moderate (2) and severe (3). An index of pulmonary hypertension (main pulmonary artery: ascending aorta ratio) was assessed by two scorers using electronic caliper diameter measurements of the ascending aorta and pulmonary artery diameters at the level of the pulmonary artery bifurcation [11]. Consensus formulation for visual scores is outlined in the online supplementary material.

CALIPER CT evaluation
Data processing

Initial data processing steps involved extraction of the lung from the surrounding thoracic structures and segmentation into upper, middle and lower zones. Lung segmentation was performed using an adaptive density-based morphological approach [12], while airway segmentation involved iterative three-dimensional
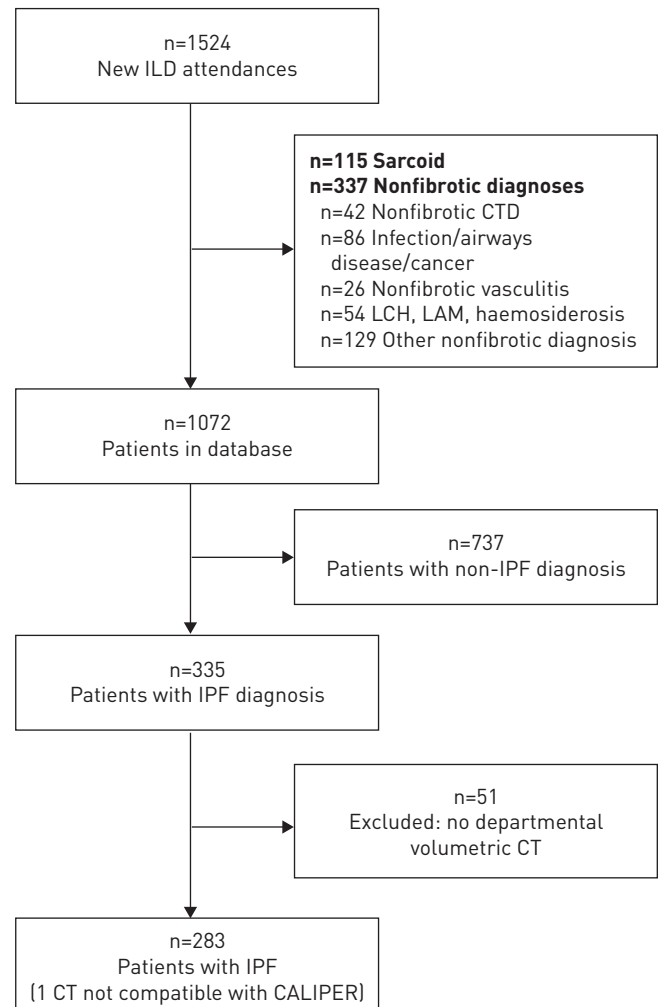
FIGURE 1 Flowchart illustrating the selection of patients for the final study population. ILD: interstitial lung disease; CTD: connective tissue disease; LCH: Langerhans cell histiocytosis; LAM: lymphangioleiomyomatosis; IPF: idiopathic pulmonary fibrosis; CT: computed tomography.

region growing, density thresholding (thresholds including $-950\,HU$ and $-960\,HU$) and connected components analysis. Segmentation of pulmonary vessels, prior to their extraction, was achieved using an optimised multiscale tubular structure enhancement filter based on the eigenvalues of the Hessian matrix. The filters calculated the second-order derivatives that occurred in the regions that surrounded each pulmonary voxel. The eigenvalues of the Hessian matrix that were constructed from the derivatives were then analysed, and from these values, it was possible to determine the likelihood that an underlying voxel was connected to a dense tubular structure and therefore represented a vessel [13, 14].

The pulmonary vessel volume (PVV) score quantified the volumes of pulmonary arteries and veins excluding vessels at the lung hilum as a percentage of lung volume (figure 2). The PVV score was subdivided according to vessel size. All vessels on a single CT image that had a cross-sectional area $<5\,mm^2$ (PVV5) or $<10\,mm^2$ (PVV10) or $>5\,mm^2$ (PVV>5) were summed, and expressed as volume (cm^3) after adjusting for z-axis CT slice thickness. The $5\,mm^2$ and $10\,mm^2$ thresholds were chosen after analysing a range of vessel size thresholds as described in the online supplementary material, and are in line with vessel size thresholds analysed in patients with chronic obstructive pulmonary disease [15]. The PVV>5 threshold was examined to remove the capture of potentially misclassified reticular pattern in the PVV variable in cases with extensive fibrosis.

Parenchymal tissue type classification was applied to 15×15×15-voxel volume units using texture analysis, computer vision-based image understanding of volumetric histogram signature mapping features and 3D morphology [13]. The CALIPER tool was trained by subspecialty thoracic radiologist consensus assessment of pathologically confirmed datasets [13, 16].
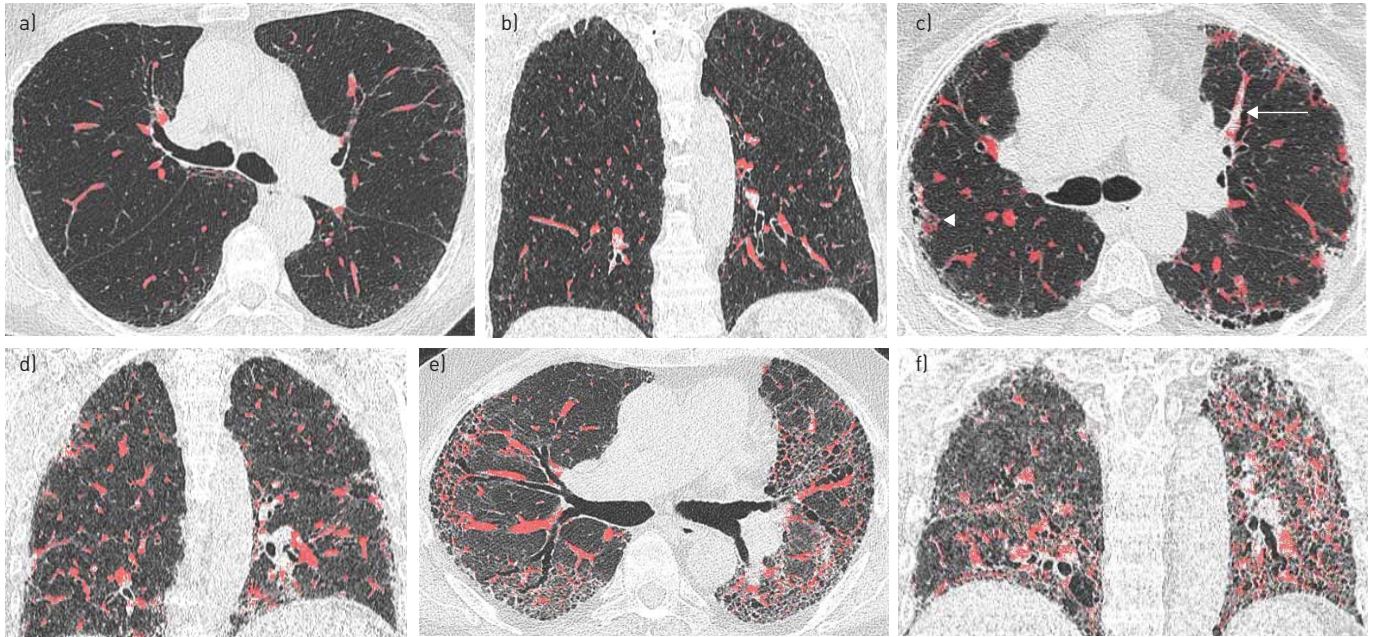
FIGURE 2 Three examples of delineation of pulmonary arteries and veins (red), quantified by CALIPER as a percentage of the total lung volume and expressed as the pulmonary vessel volume (PVV). Note that segmentation is incomplete, for example a left upper lobe vessel in part c) (arrow). In addition, the CALIPER software misclassifies small volumes of reticulation as vessels (arrowhead). Case 1: a 62-year-old female ex-smoker, with a 15 pack-year smoking history: a) mean visual scores characterised the lungs as consisting of 9% interstitial lung disease (ILD) (ground glass opacity 4.0% and reticular pattern 5.0%); b) CALIPER identified a total ILD extent of 2% of the lung (ground glass opacity 0.7% and reticular pattern 0.9%). PVV represented 2% of the lung volume. Case 2: a 61-year-old female ex-smoker, with a 25 pack-year smoking history: c) visual scores characterised the lungs as consisting of 29% ILD (ground glass opacity 3.3%, reticular pattern 8.4% and honeycombing 17.1%); d) CALIPER identified a total ILD extent of 27% (ground glass opacity 15.2% and reticular pattern 11.4%). PVV represented 6% of the lung volume. Case 3: a 64-year-old male ex-smoker, with a 40 pack-year smoking history: e) visual scores characterised the lungs as consisting of 76% total ILD extent (ground glass opacity 15.9%, reticular pattern 34.9% and honeycombing 25.4%); f) CALIPER identified a total ILD extent of 55% (ground glass opacity 41.4%, reticular pattern 12.0% and honeycombing 1.3%). PVV represented 9% of the lung volume.

Pattern evaluation

CALIPER evaluation of CT data involved algorithmic identification and volumetric quantification of every voxel volume unit into one of eight radiological parenchymal features: normal lung, three grades of decreased lung attenuation (grade 1: mild, 2: moderate and 3: marked), ground glass opacification, reticular pattern, honeycombing and the pulmonary vessels (figure 2). Volumes for all eight parenchymal features were converted into a percentage using the total lung volume, also measured by CALIPER. Total extent of ILD represented the sum of ground glass, reticular and honeycomb percentages. The sum of grade 2 and 3 decreased lung attenuation represented emphysema [8].

Statistical analysis

Data are presented as median, mean±SD or n (%). Interobserver variation for visual scores was assessed using the single-determination standard deviation. Linear regression analyses were performed to explore relationships between the PVV and CALIPER ILD extent, visual ILD and reticular pattern extents and right ventricular systolic pressure (RVSP). Linear regression was performed to evaluate relationships between PVV subdivisions and PVV and D_{LCO}. Univariate and multivariate Cox regression analyses were used to investigate relationships within and between the three datasets: CALIPER CT evaluation, visual CT evaluation and PFTs. In all study analyses, $p < 0.01$ was considered significant.

The hazard ratios for those parameters that were independent predictors of mortality on multivariate Cox regression analysis were used to generate a formula that represented an estimate of mortality for each patient. The hazard ratios for CALIPER parameters alone that were independent predictors of mortality on multivariate Cox regression analysis were also used to generate a separate formula that represented an estimate of mortality for each patient.

The mortality estimates derived from the hazard ratio scores were converted into categorical scores and compared to mortality estimates derived from the GAP index staging system [17], using univariate and bivariate Cox regression analyses, Kaplan–Meier survival plots and the log-rank test. Robustness of results were confirmed using bootstrapping and resampling of the dataset up to 1000 times. Goodness of fit of the survival models was calculated using Harrell's concordance index. Assumptions of linearity and

proportional hazards were tested by visual inspection of Martingale residuals and scaled Schoenfeld residuals. Statistical analyses were performed using Stata (version 12; StataCorp, College Station, TX, USA).

Results

Demographic data

The study population consisted of 283 consecutive patients with a multidisciplinary diagnosis of IPF. The median age at presentation was 67 years; the mean follow-up time was 30±21.5 months and 210 (74%) patients died during the study period. Data on vital status were completed in 98.6% of cases, with four (1.4%) patients censored. Demographic data and average visual score, CALIPER score and PFT data are presented in table 1. Interobserver variation values for the visual scores are provided in the online supplementary material.

Mortality analyses

All visual and CALIPER-derived interstitial features were predictive of mortality. The CALIPER-derived PVV (shown in figure 2) and PVV5, PVV10 and PVV>5 were highly significant on univariate analysis, as were all PFTs, the RVSP and the GAP score (table 2). When RVSP and CALIPER ILD extent were each placed alongside PVV in a bivariate mortality analysis, only PVV remained independently predictive of mortality. On linear regression analysis, major co-linearity was demonstrated between PVV and CALIPER ILD extent (R^2 0.76, p<0.0001) and between PVV and visual ILD extent (R^2 0.57, p<0.0001) [8]. A lesser

TABLE 1 Patient demographics and pulmonary function tests, CALIPER and visually scored computed tomography parameters and echocardiography data

Subjects	283
Median age years	67
Male/female	219/64
Smoking history	
Never-smokers	97 (34)
Ex-smokers	178 (63)
Current smokers	8 (3)
Lung biopsy	60 (21)
Follow-up months	30.0±21.5
FEV1[#] % pred	70.8±19.1
FVC[#] % pred	68.8±20.5
D_{LCO}[¶] % pred	36.1±12.9
K_{CO}[¶] % pred	69.0±19.2
TLC[+] % pred	63.7±15.9
CPI[§] %	55.1±11.7
CALIPER total extent of ILD %	26.5±18.1
Ground glass opacity	17.0±14.7
Reticular pattern	8.5±6.0
Honeycombing	1.0±1.7
CALIPER PVV %	5.1±1.7
CALIPER grade 1 decreased attenuation %	20.8±20.7
CALIPER grade 2 decreased attenuation %	0.8±2.6
CALIPER grade 3 decreased attenuation %	0.5±2.8
Visual extent of ILD %	43.1±17.8
Ground glass opacity	10.4±11.4
Reticular pattern	21.7±10.9
Honeycombing	9.8±12.6
Consolidation	1.1±3.3
Visual emphysema %	4.7±10.9
Visual mosaic attenuation %	0.8±2.1
Visual TxBx severity (maximum score 18)	7.0±3.3
Main pulmonary artery diameter mm	30.3±4.8
Ascending aorta diameter mm	34.8±4.2
Echocardiography RVSP[f] mmHg	45.1±16.8

Data are presented as n, n (%) or mean±SD, unless otherwise stated. FEV1: forced expiratory volume in 1 s; FVC: forced vital capacity; D_{LCO}: diffusing capacity of the lung for carbon monoxide; K_{CO}: transfer coefficient of the lung for carbon monoxide; TLC: total lung capacity; CPI: composite physiologic index; ILD: interstitial lung disease; PVV: pulmonary vessel volume; TxBx: traction bronchiectasis; RVSP: right ventricular systolic pressure. [#]: n=257; [¶]: n=254; [+]: n=241; [§]: n=249; [f]: n=150.

TABLE 2 Univariate Cox regression analysis demonstrating mortality according to CALIPER indices, visually derived high-resolution computed tomography indices, pulmonary function tests (PFTs) and echocardiography and the Gender, Age, Physiology (GAP) score

	Subjects n	Hazard ratio (95% CI)	p-value
CALIPER score	283		
Total ILD extent		1.03 (1.03–1.04)	<0.0001
Ground glass opacity		1.03 (1.02–1.04)	<0.0001
Reticular pattern		1.10 (1.07–1.12)	<0.0001
Honeycombing		1.11 (1.03–1.20)	0.006
Grade 1 DA		0.97 (0.97–0.98)	<0.0001
Grade 2 DA			NS
Grade 3 DA			NS
Normal lung			NS
PVV		1.52 (1.40–1.65)	<0.0001
PVV5		3.02 (2.36–3.87)	<0.0001
PVV10		2.29 (1.91–2.75)	<0.0001
PVV>5		1.64 (1.48–1.82)	<0.0001
Visual score	283		
ILD extent		1.03 (1.02–1.04)	<0.0001
Ground glass opacity		1.01 (1.00–1.03)	0.03
Reticular pattern		1.02 (1.01–1.03)	0.002
Honeycombing		1.03 (1.03–1.07)	<0.0001
Consolidation		1.08 (1.04–1.12)	<0.0001
Total emphysema			NS
Mosaicism			NS
TxBx severity (max score 18)		1.11 (1.07–1.16)	<0.0001
Main pulmonary artery diamter mm			NS
Pulmonary artery:aorta ratio			NS
PFTs			
FEV1	257	0.97 (0.97–0.98)	<0.0001
FVC	257	0.97 (0.96–0.98)	<0.0001
TLC	241	0.96 (0.95–0.97)	<0.0001
D_{LCO}	254	0.94 (0.93–0.95)	<0.0001
K_{CO}	254	0.99 (0.98–1.00)	0.001
CPI	249	1.07 (1.05–1.09)	<0.0001
Echocardiography RVSP mmHg	150	1.02 (1.02–1.03)	<0.0001
GAP score (max score 8)	249	1.45 (1.30–1.62)	<0.0001

ILD: interstitial lung disease; DA: decreased attenuation; PVV: pulmonary vessel volume; PVV5: PVV <5 mm^2 on computed tomography (CT) cross-sectional area; PVV10: PVV <10 mm^2 on CT cross-sectional area; PVV>5: PVV >5 mm^2 on CT cross-sectional area; TxBx: traction bronchiectasis; FEV1: forced expiratory volume in 1 s; FVC: forced vital capacity; TLC: total lung capacity, D_{LCO}: diffusing capacity of the lung for carbon monoxide; K_{CO}: transfer coefficient of the lung for carbon monoxide; CPI: composite physiologic index; RVSP: right ventricular systolic pressure; NS: nonsignificant.

degree of co-linearity was demonstrated between PVV and RVSP (R^2 0.20, p<0.0001) and between PVV and visual reticular pattern extent (R^2 0.16, p<0.0001). Relationships between PVV subdivisions and both functional indices and survival are shown in the online supplementary material.

On stepwise proportional hazards analysis, the independent CALIPER-derived predictors of mortality were honeycombing and the PVV (table 3). When inserted separately into multivariate models, PVV5 and PVV10 did not retain significance. However, PVV>5 was independently predictive of mortality (online supplementary table S4). On multivariate analysis of pulmonary function indices, the composite physiologic index (CPI) was the strongest predictor of mortality. A multivariate Cox regression analysis combining CALIPER scores, visual scores, PFTs and the GAP score demonstrated the CPI (which has been previously validated using CALIPER [8]) to be the variable that best quantified the severity of ILD. The two remaining CT variables independently predictive of mortality were derived by CALIPER: honeycombing and PVV. No visually scored CT parameters were independently associated with mortality (table 3). When the GAP score was substituted for the CPI in the final multivariate model, although it was retained as an independent predictor of mortality, it was not as strong as the CPI in the model (table 3). These findings were maintained when PVV>5 was substituted for PVV.

TABLE 3 Multivariate Cox regression analysis demonstrating mortality separately for CALIPER indices, visually derived high-resolution computed tomography (HRCT) indices and combined CALIPER, visually derived HRCT indices and pulmonary function tests (model 1)

	Hazard ratio (95% CI)	p-value
CALIPER analysis		
PVV	1.53 (1.41–1.66)	<0.0001
Honeycombing	1.12 (1.04–1.21)	0.004
Visual score analysis		
Ground glass opacity	1.03 (1.01–1.04)	<0.0001
Reticular pattern	1.03 (1.01–1.04)	<0.0001
Honeycombing	1.04 (1.03–1.05)	<0.0001
Consolidation	1.10 (1.06–1.15)	<0.0001
Combined model 1		
CALIPER PVV	1.23 (1.08–1.40)	0.001
CALIPER honeycombing	1.18 (1.06–1.32)	0.002
CPI	1.05 (1.02–1.07)	<0.0001
Combined model 2		
CALIPER PVV	1.39 (1.25–1.54)	<0.0001
CALIPER honeycombing	1.21 (1.09–1.34)	0.0004
GAP score (max score 8)	1.20 (1.06–1.35)	0.004

The Gender, Age, Physiology (GAP) score was substituted for composite physiologic index (CPI) in the final combined multivariate model (model 2), but was a weaker predictor of mortality than the CPI. PVV: pulmonary vessel volume.

Derivation of composite variables

A further mortality evaluation compared the GAP index staging system [17] to two mortality estimates derived from the hazard ratios of two multivariate models. One formula was derived from the hazard ratios of the final three independent predictors of mortality (CALIPER–CPI score) as follows:

$$\text{CALIPER–CPI score} = (\text{CALIPER PVV} \times 23.0904) + (\text{CALIPER honeycombing} \times 18.3795) + (\text{CPI} \times 4.5065)$$

The second formula was derived from the hazard ratios of the two variables that on multivariate analysis of CALIPER scores were independent predictors of mortality (CALIPER-only score):

$$\text{CALIPER-only score} = (\text{CALIPER PVV} \times 52.9004) + (\text{CALIPER honeycombing} \times 12.0524)$$

Both the CALIPER–CPI and CALIPER-only scores were converted into categorical scores by aligning the individual scores in ascending numerical order and dividing the respective cohorts into three equally sized groups (n=83).

Comparison of composite variables

Univariate Cox regression analyses demonstrated that the CALIPER–CPI and CALIPER-only categories were not only of similar prognostic strength to the GAP index staging system, but demonstrated improved goodness of fit as models (table 4).

On bivariate Cox regression analysis, the GAP index staging system did not retain significance against the CALIPER–CPI categories, a result confirmed with bootstrapping of 1000 samples (table 4). The GAP index staging system was also shown to be a weaker predictor of mortality than the CALIPER-only categories, again confirmed with bootstrapping of 1000 samples. These findings were maintained when PVV>5 was substituted for PVV (online supplementary table S5). Kaplan–Meier survival curves demonstrated similar separation of the groups using the CALIPER–CPI and CALIPER-only categories when compared to the GAP index staging system (log-rank test p<0.0001 for the GAP index staging system, CALIPER–CPI and CALIPER-only categories) (figure 3).

When the CALIPER–CPI and CALIPER-only scores were adjusted such that patient numbers in each of the groups were identical to the patient numbers in the GAP staging system groups, the relationships on univariate analysis and bivariate analysis with bootstrapping did not change. Goodness of fit of the two models was similar to that of the GAP index staging system (Harrell's c-index 0.66 for the CALIPER–CPI score and 0.65 for the CALIPER-only score). Kaplan–Meier survival curves again demonstrated similar separation of the groups using the new adjusted CALIPER–CPI and CALIPER-only categories when compared to the GAP index staging system (log-rank test p<0.0001 for the adjusted CALIPER–CPI and CALIPER-only categories) (figure 3).

TABLE 4 Univariate Cox regression analyses comparing the Gender, Age, Physiology (GAP) index staging system with scores derived from two separate hazard ratio formulae

Composite score (max score 3)	Hazard ratio (95% CI)	p-value	Harrell's c-index
Univariate analysis			
GAP index staging system	2.00 (1.61–2.48)	<0.0001	0.62
CALIPER–CPI	2.23 (1.85–2.69)	<0.0001	0.67
CALIPER only	2.14 (1.78–2.58)	<0.0001	0.67
Bivariate analysis			
GAP index staging system	1.25 (0.97–1.62)	0.09	
CALIPER–CPI	1.99 (1.58–2.50)	<0.0001	
GAP index staging system[#]	(-0.03–0.48)	0.09	
CALIPER–CPI[#]	(0.46–0.95)	0.001	
GAP index staging system	1.41 (1.11–1.80)	0.005	
CALIPER only	1.87 (1.51–2.30)	<0.0001	
GAP index staging system[#]	(0.10–0.59)	0.007	
CALIPER only[#]	(0.41–0.86)	0.001	

CALIPER–CPI: derived from the hazard ratios of the three independent predictors of mortality in a combined visual, CALIPER and pulmonary function index multivariate model. CALIPER only: derived from the hazard ratios of CALIPER variables that were independent predictors of mortality. Bivariate analyses compared the hazard ratio derived scores to the GAP index staging system with the results confirmed on bootstrapping of 1000 samples. CPI: composite physiologic index. [#]: bootstrapped results.
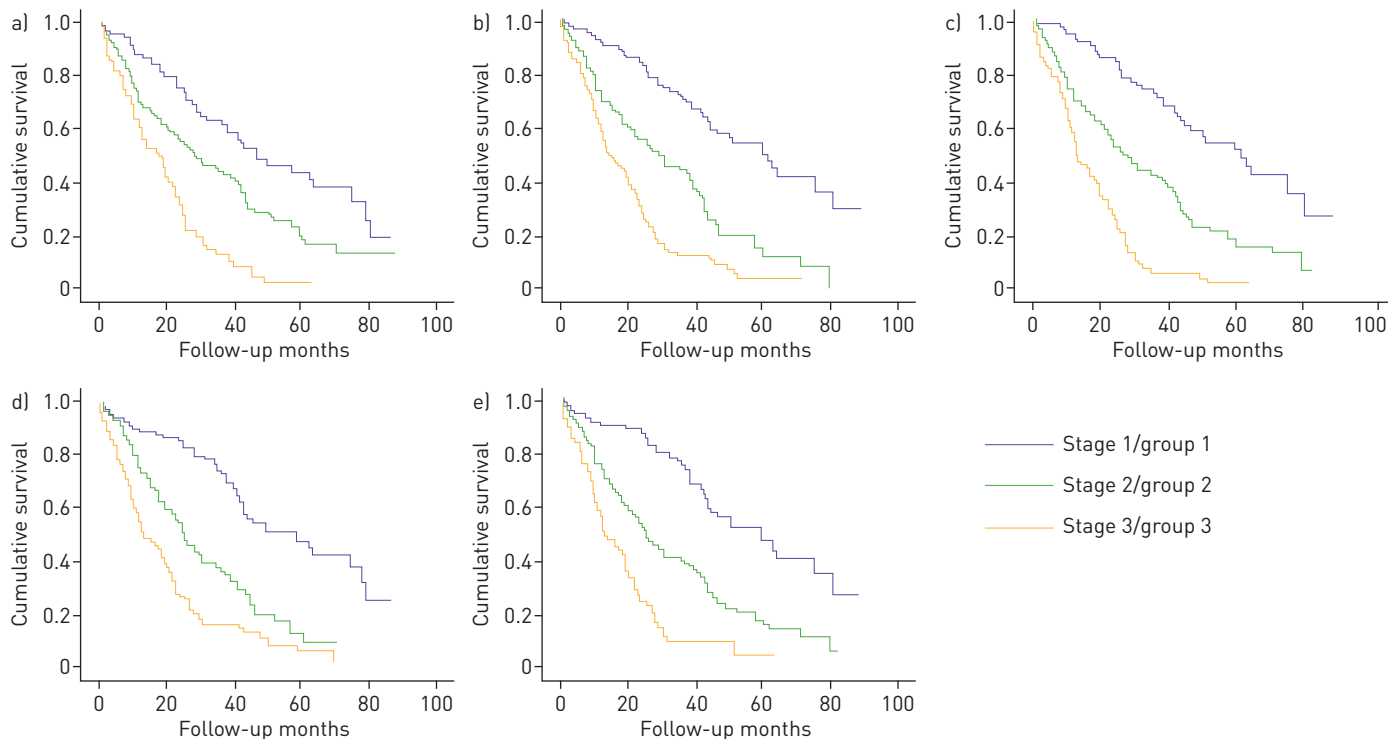

FIGURE 3 Survival curves for the idiopathic pulmonary fibrosis population in the current study. a) Cases have been separated into three groups according to the Gender, Age, Physiology (GAP) index staging system. Mean±SE survival: stage 1 (n=69) 50.2±3.7; stage 2 (n=125) 35.5±2.6; stage 3 (n=55) 19.0±1.9. b, c) Stratification of patients into three groups was achieved using a formula derived from the hazard ratios of the three independent predictors of mortality in a combined visual, CALIPER and pulmonary function index multivariate model (CALIPER-CPI (composite physiologic index)). b) Three groups of equal size. Mean±SE survival: group 1 (n=83) 56.0±3.4; group 2 (n=83) 31.9±2.7; group 3 (n=83) 19.2±1.8. c) Three groups, each equal in size to the GAP groups. Mean±SE survival: group 1 (n=69) 56.0±3.6; group 2 (n=125) 32.8±2.4; group 3 (n=55) 16.7±1.8. d, e) Stratification of patients into three groups was achieved using a formula derived from the hazard ratios of CALIPER variables that were independent predictors of mortality (CALIPER-only). d) Three groups of equal size. Mean±SE survival: group 1 55.4±3.3; group 2 30.3±2.3; group 3 20.1±2.1. e) Three groups, each equal in size to the GAP groups. Mean±SE survival: group 1 (n=69) 55.5±3.6; group 2 (n=125) 32.1±2.3; group 3 (n=55) 18.9±2.1.

Discussion

Our study has demonstrated that computer-derived quantitative CT parameters are better predictors of mortality in IPF than visually scored parameters. When patients are stratified on the basis of CALIPER variables and the CPI, mortality prediction is improved when compared to stratification using the GAP index staging system. Central to the strength of mortality prediction using CALIPER was a single variable, the PVV. The PVV is a novel CALIPER variable with no visually scored equivalent. Accordingly, the PVV may represent a new parameter in the evaluation of patients with IPF.

D_{LCO} has long been considered the parameter that best reflects disease severity at baseline in IPF [18], but is handicapped by measurement noise in the range 5–15% [3, 19]. Consequently, interest is increasing in exploring other potential markers of disease severity or worsening in IPF, such as peripheral blood [20] and imaging [7] biomarkers. Given the rapid technological advances of computer-based quantitative CT [13], an exploration of their potential accuracy in assessing IPF is timely.

In the current study, on univariate analysis, visually scored reticular pattern, honeycombing and traction bronchiectasis were predictive of mortality, confirming the conclusions of a previous study which evaluated the prognostic significance of the same patterns in patients with a histopathologically proven usual interstitial pneumonia pattern [21].

The final combined visual, CALIPER and CPI multivariate model identified three independent predictors of survival. One of these was the CPI, and the finding confirms previous reports highlighting the strength of signal of the CPI in predicting outcome in fibrosing lung disease [22, 23]. Similarly, the prognostic signal associated with honeycombing when scored by CALIPER confirms results from visual scoring of the same pattern in IPF [24].

The co-linearity demonstrated between PVV and CALIPER/visual ILD extent and between PVV and RVSP suggests that the PVV might represent a variable that simultaneously captures disease within the interstitial and vascular compartments. It is noteworthy that CALIPER software was originally designed to segment out and discard vascular structures (rather than quantify them) and in so doing optimise the classification of parenchymal patterns. In cases with more severe fibrosis on CT, vessel segmentation inevitably captures a minor degree of peripheral reticular pattern (figure 2); however, there was notably only a minor degree of co-linearity between visual reticular pattern and PVV. Furthermore, the ability of PVV to predict mortality when vessels <5 mm^2 were excluded from analysis argues against the likelihood that a significant part of the PVV signal is misclassified reticular pattern. In any event, as software algorithms improve, there remains the potential to refine vessel delineation by CALIPER, or a similar quantitative tool, and thus improve the prognostic signal of this new CT parameter.

The basis of the signal provided by PVV is obscure. PVV has not previously been studied as a prognostic marker in IPF or considered as a prognostic indicator independent of pulmonary hypertension. Recent IPF studies assessing the complex interactions between angiogenic and angiostatic mediators in disease pathogenesis [25] have primarily considered the vasculature through the prism of overt pulmonary hypertension [26].

Blood perfusion in areas of fibrosis has been shown to be reduced [27], but conversely increased in spared lung adjacent to areas of fibrosis [28, 29]. It follows that the strong correlations between ILD extent and vessels may reflect regional, subclinically elevated local pulmonary arterial pressures within mildly fibrotic lung, or capillary bed destruction in more advanced disease, which may produce a preferential diversion of blood flow to relatively spared or nonfibrotic lung. The vascular capacitance of spared lung (the upper and middle lobes in patients with IPF, a predominantly basal disease) may result in an increase in vessel volume in more advanced disease. The identification of greater numbers of vessels, of a size that could be detected by CALIPER, could therefore act as a surrogate marker for the extent and severity of parenchymal disease in IPF.

Another possible explanation for the relationship between PVV and ILD extent relates to the increased negative intrathoracic pressure that noncompliant fibrotic lungs need to generate during inspiration. The transmission of high negative pressures through the pleural space into the parenchyma could in turn be exerted on the vasculature, resulting in dilatation throughout the lung and an increase in capacitance. A third possible mechanism relates to pleuroparenchymal and/or bronchial-pulmonary artery anastomoses described histopathologically in patients with fibrosing lung disease [30]. While the clinical importance of shunting within the lungs is yet to be established [31], the development of shunts could theoretically increase the PVV as fibrosis progresses.

The evaluation of CALIPER in our previous study highlighted the close correlations between parenchymal patterns scored using an automated computer system and pulmonary function indices [8]. The results suggested that a tool such as CALIPER, when used alone, could be a viable alternative to PFTs in predicting outcome in IPF. Our mortality analyses have also underlined the importance of integrating

structural and functional parameters for prognostication in IPF. The CPI, when combined with CALIPER variables, produced a stronger model with which to predict mortality than achieved by CALIPER or functional indices alone.

The GAP score is a multidimensional continuous score that aims to predict mortality in IPF by utilising commonly measured clinical and physiological variables [17]. The GAP score was shown to be a strong univariate predictor of mortality in our study, and was enhanced when combined with CALIPER variables in a multivariate model. However, the model was not as powerful as the combination of CALIPER variables with CPI. The findings highlight a relative weakness of the GAP score consequent upon its relatively coarse nine-point gradation, whereas the continuous nature of the CPI allows the CPI to be more discriminatory.

The GAP staging system represents a categorical version of the GAP score [17]. A consequence of stratification according to the GAP staging system is that only 50% of patients in the current study constituted GAP groups 1 and 3 (patients with mild and severe disease, respectively). Yet it is precisely such patients that require identification in IPF cohorts. Those identified as likely to have limited disease can be monitored using a watch-and-wait policy, while those with more severe disease might be referred earlier for transplantation [2]. When stratification with the CALIPER–CPI categories was performed, despite the arbitrary division of the patients into evenly sized groups, prognostication and goodness of fit was improved when compared to the categorical GAP staging system.

Emphysema when scored visually or by CALIPER was not found to be predictive of mortality in the current study. There have been conflicting reports in the literature regarding the impact on survival when emphysema coexists with fibrosis. Whilst some authors have argued that emphysema is a poor prognostic determinant in IPF [32–35], other studies have refuted this observation [36–40]. While our findings concur with the latter view, more detailed evaluation of emphysema subtypes and distribution within the lung, as well as correction for baseline disease severity in patients with IPF may help clarify the primary effect of emphysema on survival in IPF.

A limitation of our study lies in the lack of an external validation cohort with which to confirm our study findings. However, the scarcity of well-characterised populations of IPF patients, even in tertiary centres, is well recognised. A potential solution lies in the pooling of international multicentre populations, similar to the COPDGene study population [41], but such an undertaking will require close collaborative efforts between tertiary centres.

In conclusion we have shown that quantitative computer-derived CT variables in IPF are superior predictors of mortality than any visually scored CT parameter. Stratification using CALIPER variables and CPI provides a stronger mortality signal than stratification using the GAP index. But one CALIPER variable in particular, the PVV, has the strongest link with mortality and could be a new index in the evaluation of IPF.

References

1   Schoenheit G, Becattelli I, Cohen AH. Living with idiopathic pulmonary fibrosis: an in-depth qualitative survey of European patients. *Chron Respir Dis* 2011; 8: 225–231.

2   Lamas DJ, Kawut SM, Bagiella E, *et al.* Delayed access and survival in idiopathic pulmonary fibrosis: a cohort study. *Am J Respir Crit Care Med* 2011; 184: 842–847.

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

4   Lynch DA, David GJ, 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.

5   Watadani T, Sakai F, Johkoh T, *et al.* Interobserver variability in the CT assessment of honeycombing in the lungs. *Radiology* 2013; 266: 936–944.

6   Ley B, Collard HR. Risk prediction in idiopathic pulmonary fibrosis. *Am J Respir Crit Care Med* 2012; 185: 6–7.

7   Hansell DM, Goldin JG, King TE Jr, *et al.* CT staging and monitoring of fibrotic interstitial lung diseases in clinical practice and treatment trials: a position paper from the Fleischner Society. *Lancet Respir Med* 2015; 3: 483–496.

8   Jacob J, Bartholmai BJ, Rajagopalan S, *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.

9   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.

10   Hansell DM, Bankier AA, MacMahon H, *et al.* Fleischner Society: glossary of terms for thoracic imaging. *Radiology* 2008; 246: 697–722.

11   Ng CS, Wells AU, Padley SP. A CT sign of chronic pulmonary arterial hypertension: the ratio of main pulmonary artery to aortic diameter. *J Thorac Imaging* 1999; 14: 270–278.

12   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.

13   Bartholmai BJ, Raghunath S, Karwoski RA, *et al.* Quantitative computed tomography imaging of interstitial lung diseases. *J Thorac Imaging* 2013; 28: 298–307.

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   Matsuoka S, Washko GR, Dransfield MT, *et al.* Quantitative CT measurement of cross-sectional area of small pulmonary vessel in COPD: correlations with emphysema and airflow limitation. *Acad Radiol* 2010; 17: 93–99.

16   Maldonado F, Moua T, Rajagopalan S, *et al.* Automated quantification of radiological patterns predicts survival in idiopathic pulmonary fibrosis. *Eur Respir J* 2014; 43: 204–212.

17   Ley B, Ryerson CJ, Vittinghoff E, *et al.* A multidimensional index and staging system for idiopathic pulmonary fibrosis. *Ann Intern Med* 2012; 156: 684–691.

18   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.

19   Hathaway EH, Tashkin DP, Simmons MS. Intraindividual variability in serial measurements of D_{LCO} and alveolar volume over one year in eight healthy subjects using three independent measuring systems. *Am Rev Respir Dis* 1989; 140: 1818–1822.

20   Ley B, Brown KK, Collard HR. Molecular biomarkers in idiopathic pulmonary fibrosis. *Am J Physiol Lung Cell Mol Physiol* 2014; 307: L681–L691.

21   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.

22   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.

23   Corte TJ, Wort SJ, Gatzoulis MA, *et al.* Pulmonary vascular resistance predicts early mortality in patients with diffuse fibrotic lung disease and suspected pulmonary hypertension. *Thorax* 2009; 64: 883–888.

24   Flaherty KR, Toews GB, Travis WD, *et al.* Clinical significance of histological classification of idiopathic interstitial pneumonia. *Eur Respir J* 2002; 19: 275–283.

25   Hanumegowda C, Farkas L, Kolb M. Angiogenesis in pulmonary fibrosis: too much or not enough? *Chest* 2012; 142: 200–207.

26   Farkas L, Farkas D, Ask K, *et al.* VEGF ameliorates pulmonary hypertension through inhibition of endothelial apoptosis in experimental lung fibrosis in rats. *J Clin Invest* 2009; 119: 1298–1311.

27   Renzoni EA, Walsh DA, Salmon M, *et al.* Interstitial vascularity in fibrosing alveolitis. *Am J Respir Crit Care Med* 2003; 167: 438–443.

28   Ebina M, Shimizukawa M, Shibata N, *et al.* Heterogeneous increase in CD34-positive alveolar capillaries in idiopathic pulmonary fibrosis. *Am J Respir Crit Care Med* 2004; 169: 1203–1208.

29   Cosgrove GP, Brown KK, Schiemann WP, *et al.* Pigment epithelium-derived factor in idiopathic pulmonary fibrosis: a role in aberrant angiogenesis. *Am J Respir Crit Care Med* 2004; 170: 242–251.

30   Turner-Warwick M. Precapillary systemic-pulmonary anastomoses. *Thorax* 1963; 18: 225–237.

31   Miller WC, Heard JG, Unger KM, *et al.* Anatomical lung shunting in pulmonary fibrosis. *Thorax* 1986; 41: 208–209.

32   Cottin V, Nunes H, Brillet PY, *et al.* Combined pulmonary fibrosis and emphysema: a distinct underrecognised entity. *Eur Respir J* 2005; 26: 586–593.

33   Lee CH, Kim HJ, Park CM, *et al.* The impact of combined pulmonary fibrosis and emphysema on mortality. *Int J Tuberc Lung Dis* 2011; 15: 1111–1116.

34   Sugino K, Ishida F, Kikuchi N, *et al.* Comparison of clinical characteristics and prognostic factors of combined pulmonary fibrosis and emphysema *versus* idiopathic pulmonary fibrosis alone. *Respirology* 2014; 19: 239–245.

35   Mejía M, Carrillo G, Rojas-Serrano J, *et al.* Idiopathic pulmonary fibrosis and emphysema: decreased survival associated with severe pulmonary arterial hypertension. *Chest* 2009; 136: 10–15.

36   Ryerson CJ, Hartman T, Elicker BM, *et al.* Clinical features and outcomes in combined pulmonary fibrosis and emphysema in idiopathic pulmonary fibrosis. *Chest* 2013; 144: 234–240.

37   Jankowich MD, Rounds S. Combined pulmonary fibrosis and emphysema alters physiology but has similar mortality to pulmonary fibrosis without emphysema. *Lung* 2010; 188: 365–373.

38   Bodlet A, Maury G, Jamart J, *et al.* Influence of radiological emphysema on lung function test in idiopathic pulmonary fibrosis. *Respir Med* 2013; 107: 1781–1788.

39   Kurashima K, Takayanagi N, Tsuchiya N, *et al.* The effect of emphysema on lung function and survival in patients with idiopathic pulmonary fibrosis. *Respirology* 2010; 15: 843–848.

40   Todd NW, Jeudy J, Lavania S, *et al.* Centrilobular emphysema combined with pulmonary fibrosis results in improved survival. *Fibrogenesis Tissue Repair* 2011; 4: 16.

41   Regan EA, Hokanson JE, Murphy JR, *et al.* Genetic epidemiology of COPD (COPDGene) study design. *COPD* 2010; 7: 32–43.