

# Deep Learning–based Segmentation of Computed Tomography Scans Predicts Disease Progression and Mortality in Idiopathic Pulmonary Fibrosis

Ⓞ Muhunthan Thillai<sup>1,2</sup>, Justin M. Oldham<sup>3</sup>, Alessandro Ruggiero<sup>1,2</sup>, Fahdi Kanavati<sup>2</sup>, Tom McLellan<sup>1</sup>, Gauri Saini<sup>5</sup>, Simon R. Johnson<sup>5</sup>, Francois-Xavier Ble<sup>6</sup>, Adnan Azim<sup>6</sup>, Kristoffer Ostridge<sup>7,9</sup>, Adam Platt<sup>7</sup>, Maria Belvisi<sup>4,8</sup>, Toby M. Maher<sup>4,10</sup>, and Philip L. Molyneaux<sup>4,11</sup>

<sup>1</sup>Royal Papworth Hospital, Cambridge, United Kingdom; <sup>2</sup>Qureight Ltd., Cambridge, United Kingdom; <sup>3</sup>Division of Pulmonary and Critical Care Medicine, University of Michigan, Ann Arbor, Michigan; <sup>4</sup>National Heart and Lung Institute, Imperial College London, London, United Kingdom; <sup>5</sup>Translational Medical Sciences, National Institute for Health and Care Research Biomedical Research Centre and Biodiscovery Institute, University of Nottingham, Nottingham, United Kingdom; <sup>6</sup>Translational Science and Experimental Medicine, Research and Early Development, Respiratory and Immunology, BioPharmaceuticals R&D, AstraZeneca, Cambridge, United Kingdom; <sup>7</sup>Translational Science and Experimental Medicine, <sup>8</sup>Research and Early Development, Respiratory and Immunology, BioPharmaceuticals R&D, AstraZeneca, Gothenburg, Sweden; <sup>9</sup>Faculty of Medicine, University of Southampton, Southampton, United Kingdom; <sup>10</sup>Keck School of Medicine, University of Southern California, Los Angeles, California; and <sup>11</sup>Royal Brompton and Harefield Hospital, Guy's and St. Thomas' NHS Foundation Trust, London, United Kingdom

ORCID IDs: 0000-0003-4957-8869 (J.M.O.); 0000-0002-9837-2763 (S.R.J.); 0000-0001-7192-9149 (T.M.M.); 0000-0003-1301-8800 (P.L.M.).

## Abstract

**Rationale:** Despite evidence demonstrating a prognostic role for computed tomography (CT) scans in idiopathic pulmonary fibrosis (IPF), image-based biomarkers are not routinely used in clinical practice or trials.

**Objectives:** To develop automated imaging biomarkers using deep learning–based segmentation of CT scans.

**Methods:** We developed segmentation processes for four anatomical biomarkers, which were applied to a unique cohort of treatment-naïve patients with IPF enrolled in the PROFILE (Prospective Observation of Fibrosis in the Lung Clinical Endpoints) study and tested against a further United Kingdom cohort. The relationships among CT biomarkers, lung function, disease progression, and mortality were assessed.

**Measurements and Main Results:** Data from 446 PROFILE patients were analyzed. Median follow-up duration was 39.1 months (interquartile range, 18.1–66.4 mo), with a cumulative incidence of death of 277 (62.1%) over 5 years. Segmentation was

successful on 97.8% of all scans, across multiple imaging vendors, at slice thicknesses of 0.5–5 mm. Of four segmentations, lung volume showed the strongest correlation with FVC ( $r = 0.82$ ;  $P < 0.001$ ). Lung, vascular, and fibrosis volumes were consistently associated across cohorts with differential 5-year survival, which persisted after adjustment for baseline gender, age, and physiology score. Lower lung volume (hazard ratio [HR], 0.98 [95% confidence interval (CI), 0.96–0.99];  $P = 0.001$ ), increased vascular volume (HR, 1.30 [95% CI, 1.12–1.51];  $P = 0.001$ ), and increased fibrosis volume (HR, 1.17 [95% CI, 1.12–1.22];  $P < 0.001$ ) were associated with reduced 2-year progression-free survival in the pooled PROFILE cohort. Longitudinally, decreasing lung volume (HR, 3.41 [95% CI, 1.36–8.54];  $P = 0.009$ ) and increasing fibrosis volume (HR, 2.23 [95% CI, 1.22–4.08];  $P = 0.009$ ) were associated with differential survival.

**Conclusions:** Automated models can rapidly segment IPF CT scans, providing prognostic near and long-term information, which could be used in routine clinical practice or as key trial endpoints.

**Keywords:** IPF; machine learning

(Received in original form November 29, 2023; accepted in final form March 7, 2024)

Ⓞ This article is open access and distributed under the terms of the Creative Commons Attribution 4.0 International License (<https://creativecommons.org/licenses/by/4.0/>). For reprints, please contact Diane Gern ([dgern@thoracic.org](mailto:dgern@thoracic.org)).

Supported by Qureight Ltd. from AstraZeneca.

**Author Contributions:** M.T., A.R., F.K., G.S., S.R.J., T.M.M., and P.L.M. designed the study and were involved in data collection. T.M. collected the additional Cambridge data. J.M.O., and T.M.M. assisted in developing the statistical analysis plan. M.T., A.R., F.J., J.M.O., and P.L.M. verified the data. J.M.O. undertook statistical analyses with oversight from M.T., A.R., and P.L.M. M.T., J.M.O., and P.L.M. drafted the manuscript, which was read, revised, and approved by all authors. All authors were responsible for the final decision to submit the manuscript.

Correspondence and requests for reprints should be addressed to Philip L. Molyneaux, M.D., Ph.D., National Heart and Lung Institute, Imperial College London, London, UK. E-mail: [p.molyneaux@imperial.ac.uk](mailto:p.molyneaux@imperial.ac.uk).

This article has a related editorial.

A data supplement for this article is available via the Supplements tab at the top of the online article.

Am J Respir Crit Care Med Vol 210, Iss 4, pp 465–472, Aug 15, 2024

Copyright © 2024 by the American Thoracic Society

Originally Published in Press as DOI: 10.1164/rccm.202311-2185OC on March 7, 2024

Internet address: [www.atsjournals.org](http://www.atsjournals.org)

## At a Glance Commentary

### Scientific Knowledge on the

**Subject:** A number of imaging tools have been developed to analyze computed tomography scans in idiopathic pulmonary fibrosis, but image-based biomarkers are still not routinely used in hospital practice or in clinical trials.

### What This Study Adds to the

**Field:** Deep learning was used to create anatomical segmentation models, which were applied to a unique set of treatment-naïve patients from the PROFILE (Prospective Observation of Fibrosis in the Lung Clinical Endpoints) study and further validated against 195 real-world patients with idiopathic pulmonary fibrosis. These models can rapidly segment computed tomography scans and provide prognostic near- and long-term information, which could be used in routine clinical practice or as key trial endpoints.

Idiopathic pulmonary fibrosis (IPF) is a progressive and ultimately fatal condition for which there is no cure (1). Antifibrotic medications slow, but do not prevent, disease progression (2). In clinical practice, monitoring of IPF progression is performed using serial measurements of lung physiology, including FVC and  $DL_{CO}$  (3–5). Measurement variability often makes the interpretation of short-term lung function trends difficult (6). Furthermore, many patients struggle to perform lung function because of fatigue or cough, and comparison of results obtained at different institutions is frequently challenging (7).

Thoracic computed tomography (CT) is the modality of choice for diagnosing individuals with suspected IPF and has largely supplanted lung biopsy, which carries a significant risk of morbidity and mortality in patients with lung fibrosis (8). Various CT features of fibrosis are predictive of prognosis when measured both at baseline and longitudinally (9–11). However, visual

assessment and scoring of CT patterns is time consuming, subjective, and associated with high interobserver variability (12). Manual segmentation of CT scans into individual components (e.g., airways) is also possible but requires both expert radiology training and significant time to segment each CT scan. Although computer-assisted approaches for classifying and quantifying CT patterns of fibrosis have been available for nearly three decades, none is routinely used in either clinical practice or as an endpoint in trials (13). To be used as an effective biomarker, automated assessment of CT imaging needs to be agnostic to CT equipment, tolerant of technical differences between scans, reproducible, rapid to perform, and readily interpretable. As with other biomarkers, imaging algorithms require validation in prospective cohorts so that the relationship between change in any given imaging parameter and outcomes of importance to patients (disease worsening, death, etc.) can be clearly defined.

We sought to develop CT-based imaging biomarkers by using high-throughput segmentation consisting of four different anatomical models identified using deep learning methods. These models were applied to a unique cohort of treatment-naïve patients with IPF enrolled in the PROFILE (Prospective Observation of Fibrosis in the Lung Clinical Endpoints) study, a multicenter prospective study of IPF conducted in the United Kingdom. Although a small number of patients in PROFILE were treated toward the end of the observational study, they were treatment naïve at the point of data capture (as it was before the widespread use of antifibrotics). The relationships between CT imaging biomarkers and lung function, disease progression, and mortality were assessed at baseline and longitudinally.

## Methods

Study approval was obtained from ethics committees (10/H0402/2, 10/H0720/12). Incident IPF cases were recruited into PROFILE (14) after informed consent was obtained. Diagnosis for all patients with IPF occurred in multidisciplinary meetings, with clinical input from both pulmonologists and thoracic radiologists. Patients were assessed at baseline; at 1, 3, 6, and 12 months; and then annually for 3 years. Survival status was

assessed on June 3, 2020. Thoracic CT scans were captured as part of routine care before enrollment and during follow-up. Pulmonary function testing (PFT) closest to CT acquisition (within 180 d) was used to ascertain baseline lung function relative to CT. The cohort was stratified into equally sized discovery and validation cohorts on the basis of the date of recruitment. Findings were also tested in an independent validation cohort of 195 patients with IPF from Royal Papworth Hospital (Cambridge cohort). These patients had CT scans analyzed at baseline (CT performed closest to the time of multidisciplinary diagnosis), and follow-up vital status was acquired by review of the medical record. Ethical approval was given by the hospital research and development committee (reference S02467).

Fibrosis and airway models were trained using supervised approach with CT scans as inputs and ground-truth segmentations as outputs. Ground-truth labeling for both models was bootstrapped by manual segmentation masks performed by two independent radiologists, each with more than 10 years' IPF reporting experience. Airways were segmented from the trachea up to the last visible generation of distal airways. Fibrosis (areas of reticulation and honeycombing) were segmented in both lungs. Initial segmentation masks were used to train a first model, which was then applied to a new image set manually corrected by the radiologists. Intermediate models were iteratively retrained after 20–30 manual corrections were performed. Fibrosis and airway models were based on three-dimensional convolutional neural networks with UNet architecture consisting of encoder and decoder networks with skip connections. For fibrosis, images were resampled to 1.4-mm resolution and masked with lung segmentations obtained from an open-source model (15) also used for lung volume segmentation. To focus on intrapulmonary airways (up to the seventh generation), trachea and proximal main bronchi were removed by lung mask segmentation multiplication. Intrapulmonary vessels were segmented by applying the Frangi enhancement filter (16). A full description of the training and testing CT scans, detailed model development, and segmentation methods is provided in the online supplement.

Univariable and multivariable Cox proportional-hazards regression adjusted for baseline gender, age, and physiology (GAP)

stage (13) at the time of CT was used to assess the primary endpoint of five-year overall survival, defined as the time from CT to death of any cause or censoring at 60 months or before if lost to follow-up. The PROFILE discovery and validation cohorts were then pooled and secondary analyses performed. Pearson correlation was used to estimate correlations between CT and baseline PFT measures. Two-year progression-free survival, defined as the time from CT to death,  $\geq 10\%$  relative FVC decline, or censoring was assessed using a multivariable Cox proportional-hazards regression model adjusted for baseline GAP stage.

Annual change in CT features was then assessed using linear mixed-effects regression in a subset of individuals who underwent repeat CT 6–24 months after baseline CT ( $n = 134$ ). This model included an exchangeable correlation structure and a random slope term, which produced the best model fit. Using simple linear regression, we then estimated patient-specific annualized changes in CT biomarkers and, using these values, categorized annual change in each CT biomarker as stable or decreasing versus increasing. The association between categorical change and subsequent two-year survival, defined as the time from second CT to death or censoring, was assessed using Cox proportional-hazards regression. Survival between categorical change groups was compared using a log-rank test and plotted using the Kaplan-Meier estimator.

## Results

### PROFILE Data Clinical and Demographic Analysis

Of 628 PROFILE participants, 426 met inclusion criteria (i.e., had at least a baseline CT scan, baseline PFT, and PFT values within 180 days of CT scan). CT segmentation resulted in approximately 7,800 segmentations derived from the four different models (fibrosis, airway, vascular, and lung volumes), with a success rate of 99.3%. The discovery and validation cohorts were well balanced in terms of age, sex, smoking history, lung function, and outcomes (Table 1). Forty-six percent ( $n = 97$ ) of patients died within three years of baseline CT in the discovery cohort and 44% ( $n = 93$ ) in the validation cohort. Nearly half of each cohort experienced IPF progression within two years (i.e., 24-mo progression).

### Model Training Results and CT Correlation with Lung Function

The performance of the fibrosis and airways models was validated using the Dice score coefficient, which is a measure of overlap widely used to assess segmentation performance against the ground truth (here represented by the testing dataset consisting of segmentations performed by the two independent radiologists and excluded from the training dataset) (17). With respect to the testing dataset, the airways segmentation model achieved a Dice score  $0.93 \pm 0.05$ . The fibrosis segmentation model achieved a

Dice score of  $0.90 \pm 0.11$ . There was no correlation between Dice score and any of image attributes, such as scanner manufacturer, convolution kernel, or slice thickness.

Baseline CT scans from PROFILE were each segmented into lung, airways, vascular, and fibrosis (Figure 1). Of the four imaging segmentations, lung volume showed the strongest correlation with FVC ( $r = 0.82$ ;  $P < 0.001$ ). No strong correlations were seen between imaging outputs and any other physiological variable (Figure 2).

### Survival Analysis

In cross-sectional analysis, baseline CT scans with lower lung volume, higher vascular volume, higher airway volume, and higher fibrosis volume were all associated with reduced five-year survival in both the discovery and validation cohorts (Table 2). When adjusted for baseline disease severity, lung volume (hazard ratio [HR], 0.98 [95% confidence interval (CI), 0.97–0.99];  $P = 0.007$ ), Vascular volume (HR, 1.37 [95% CI, 1.20–1.56];  $P < 0.001$ ), fibrosis volume (HR, 1.17 [95% CI, 1.12–1.22];  $P < 0.001$ ), and airway volume (HR, 5.41 [95% CI, 1.87–15.66];  $P = 0.002$ ) all maintained this survival association in the cohort as a whole. When stratifying by tertile, each measure effectively discriminated survival in combined cohort analysis (Figures 3A–3D). In the additional Cambridge validation set, lung volume, vascular volume, and fibrosis volume were all associated with five-year

**Table 1.** Baseline Characteristics and Outcomes for the PROFILE and Cambridge Cohorts

Characteristic	PROFILE Discovery ( $n = 223$ )*	PROFILE Validation ( $n = 223$ ) <sup>†</sup>	Cambridge Validation ( $n = 195$ ) <sup>‡</sup>
Age, yr, mean (SD)	69.4 (8.2)	70.8 (8.2)	72.6 (7.7)
Male sex, $n$ (%)	179 (80.3)	170 (76.2)	142 (85.0)
Ever-smoker, $n$ (%)	163 (73.0)	142 (63.7)	153 (78.5)
Pulmonary function, mean (SD)			
FVC% predicted	75.5 (18.9)	79.3 (18.9)	77.2 (15.5)
DL <sub>CO</sub> % predicted	43.5 (15.7)	44.7 (14.6)	50.1 (14.4)
Lung volume, L, mean (SD)	4.02 (1.13)	4.00 (1.01)	3.97 (1.12)
Airway volume, L, mean (SD)	0.02 (0.01)	0.02 (0.01)	0.07 (0.03)
Vascular volume, L, mean (SD)	0.27 (0.10)	0.25 (0.09)	0.23 (0.08)
Fibrosis volume, L, mean (SD)	0.63 (0.33)	0.54 (0.26)	0.57 (0.26)
Outcomes			
24-mo progression	116 (52.0)	101 (45.3)	NA
60-mo mortality	150 (67.3)	127 (57.0)	138 (70.8)

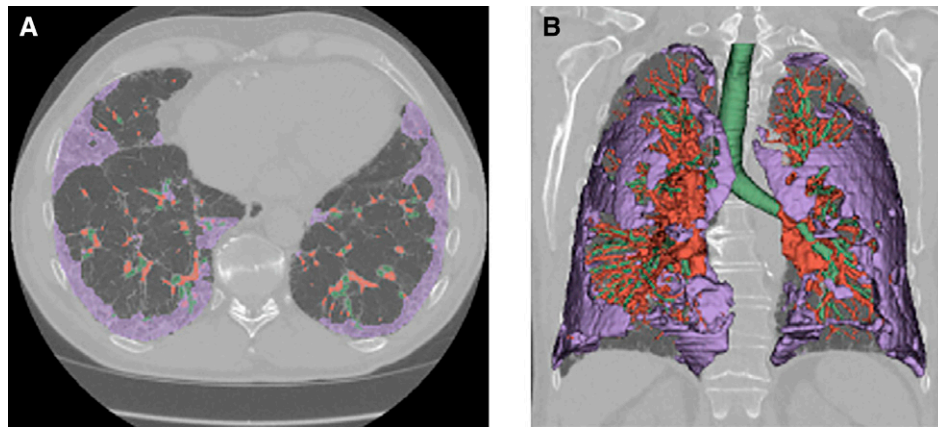
*Definition of abbreviation:* PROFILE = Prospective Observation of Fibrosis in the Lung Clinical Endpoints.

\*Number missing: FVC,  $n = 2$ ; DL<sub>CO</sub>,  $n = 11$ .

<sup>†</sup>Number missing: FVC,  $n = 1$ ; DL<sub>CO</sub>,  $n = 8$ .

<sup>‡</sup>Number missing: FVC,  $n = 28$ ; DL<sub>CO</sub>,  $n = 28$ .





**Figure 1.** (A and B) Axial (A) and three-dimensional rendering (B) of an example high-resolution computed tomography scan that has been segmented to show areas of fibrosis (purple), blood vessels (red), and airways (green) within a lung mask.

survival. When stratifying by category, lung volume and fibrosis volume both still discriminated survival (see Figure E1 in the online supplement).

#### Progression-Free Survival Analysis

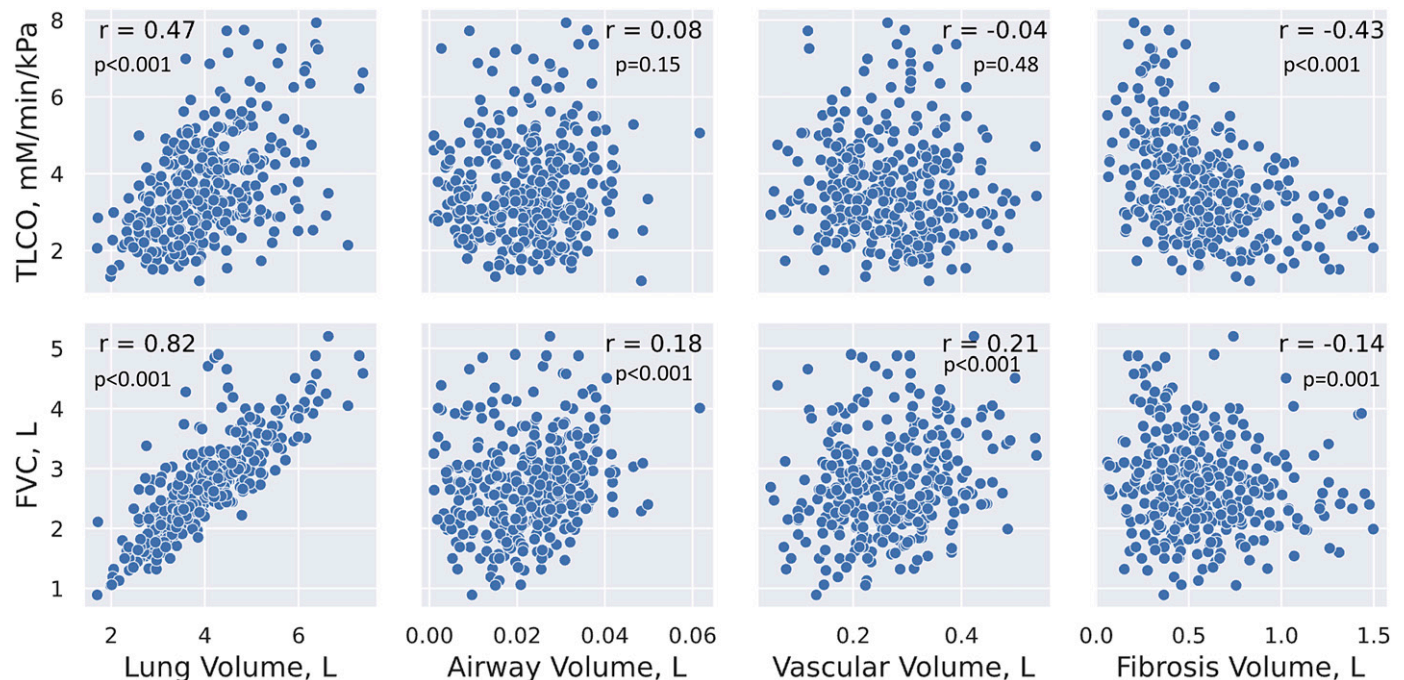
The protocolized, prospective nature of PROFILE allowed us to assess near-term outcomes. Lower lung volume, increased vascular volume, and increased fibrosis volume were again associated with reduced progression-free survival in the discovery and validation cohorts. Lung volume

(HR, 0.98 [95% CI, 0.96–0.99];  $P = 0.001$ ), vascular volume (HR, 1.30 [95% CI, 1.12–1.51];  $P = 0.001$ ), and fibrosis volume (HR, 1.17 [95% CI, 1.12–1.22];  $P < 0.001$ ) maintained this survival association in the cohort as a whole when the model was adjusted for baseline GAP index (Table 3) as well as through a sensitivity analysis iteratively adding each GAP component (see Table E1).

#### Longitudinal Analysis

In an exploratory longitudinal analysis of 134 patients who underwent repeat CT 6–24

months after baseline CT, lung volume decreased by 8.9% annually (95% CI, –14.38% to –3.43%;  $P = 0.001$ ), and fibrosis volume increased by 13.2% (95% CI, 3.59% to 22.85%;  $P = 0.007$ ). Vascular volume did not significantly change over time, and airway volume values could not be estimated, as model assumptions were not satisfied. When assessing survival after the last CT scan obtained, those with decreasing lung volume had a greater than threefold increased risk of death (HR, 3.41 [95% CI, 1.36–8.54];  $P = 0.009$ ) compared with those



**Figure 2.** Scatterplots displaying relationships between pulmonary function testing (PFT) measures and high-resolution computed tomography (HRCT) segmentations. All PFT measurements were taken within 180 days of HRCT scans. TLCO = transfer capacity of the lung for carbon monoxide.

**Table 2.** Unadjusted Associations between Baseline Quantitative CT Features and Five-Year Survival in the PROFILE and Cambridge Cohorts

CT Measure	PROFILE Discovery (n = 223)			PROFILE Validation (n = 223)			Cambridge Validation (n = 195)		
	HR	95% CI	P Value	HR	95% CI	P Value	HR	95% CI	P Value
Lung volume	0.98	0.96–0.99	0.001	0.97	0.96–0.99	0.003	0.98	0.96–0.99	0.005
Airway volume	1.90	0.46–7.82	0.374	10.09	2.06–49.35	0.004	1.82	0.96–3.45	0.068
Vascular volume	1.43	1.20–1.69	<0.001	1.31	1.09–1.59	0.005	1.23	1.01–1.50	0.042
Fibrosis volume	1.22	1.17–1.28	<0.001	1.22	1.15–1.30	<0.001	1.11	1.05–1.18	0.001

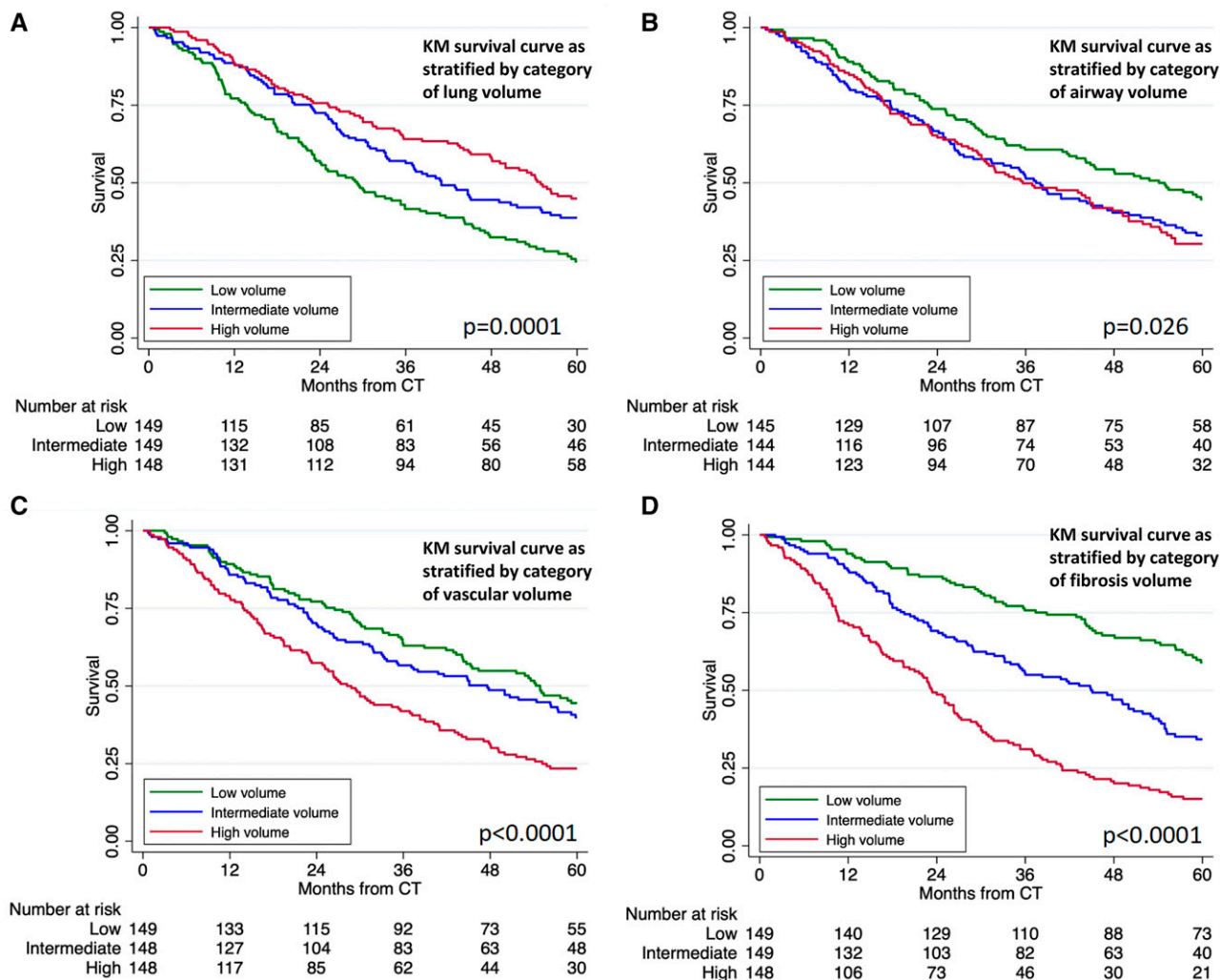
Definition of abbreviations: CI = confidence interval; CT = computed tomography; HR = hazard ratio; PROFILE = Prospective Observation of Fibrosis in the Lung Clinical Endpoints.

with stable or increasing lung volume (Figure 4A). Those with increasing fibrosis volume had more than twofold increased risk of death (HR, 2.23 [95% CI, 1.22–4.08];  $P=0.009$ ) compared with those with stable or decreasing fibrosis volume (Figure 4B).

### Discussion

We demonstrate that CT scans from patients with IPF can be used to train and develop models that can rapidly segment CT scans at scale to produce data on fibrosis, vessel,

airway, and lung volumes and that these can predict both progressive disease and mortality. All four of our CT imaging models at baseline are predictive of mortality. Fibrosis extent, vascular volume, and lung volume are also predictive of near-term



**Figure 3.** (A–D) KM survival curve for combined PROFILE cohort stratified by category (tertile) of lung volume (A), airway volume (B), vascular volume (C), and fibrosis volume (D). CT = computed tomography; KM = Kaplan-Meier; PROFILE = Prospective Observation of Fibrosis in the Lung Clinical Endpoints.

**Table 3.** Adjusted Association between Baseline Quantitative CT Features and Clinically Relevant Outcomes in the Pooled Prospective Observation of Fibrosis in the Lung Clinical Endpoints Cohort ( $n = 426$ )

CT Measure	<i>n</i>	Two-Year Progression-Free Survival			Five-Year Survival		
		HR	95% CI	<i>P</i> Value	HR	95% CI	<i>P</i> Value
Lung volume	427	0.98	0.96–0.99	0.001	0.98	0.97–0.99	0.007
Airway volume	415	1.45	0.43–4.85	0.545	3.74	1.28–10.92	0.016
Vascular volume	426	1.30	1.12–1.51	0.001	1.37	1.20–1.57	<0.001
Fibrosis volume	427	1.17	1.12–1.22	<0.001	1.17	1.12–1.22	<0.001

*Definition of abbreviations:* CI = confidence interval; CT = computed tomography; HR = hazard ratio.

Model adjusted for gender, age, and physiology stage at the time of CT. HRs are per 0.1-L change in high-resolution CT biomarker volume.

progressive disease. The PROFILE dataset is unique in that it consists of patients who were treatment naive in that the data were collected before the advent of widespread antifibrotics. Nevertheless, to validate the models further, we applied them to another dataset of real-world patients from Cambridge. Although those patients were from a smaller dataset and had the confounding factor that many were treated with antifibrotics, baseline lung volume, vascular volume, and fibrosis volume in that additional cohort were all still associated with 5-year survival.

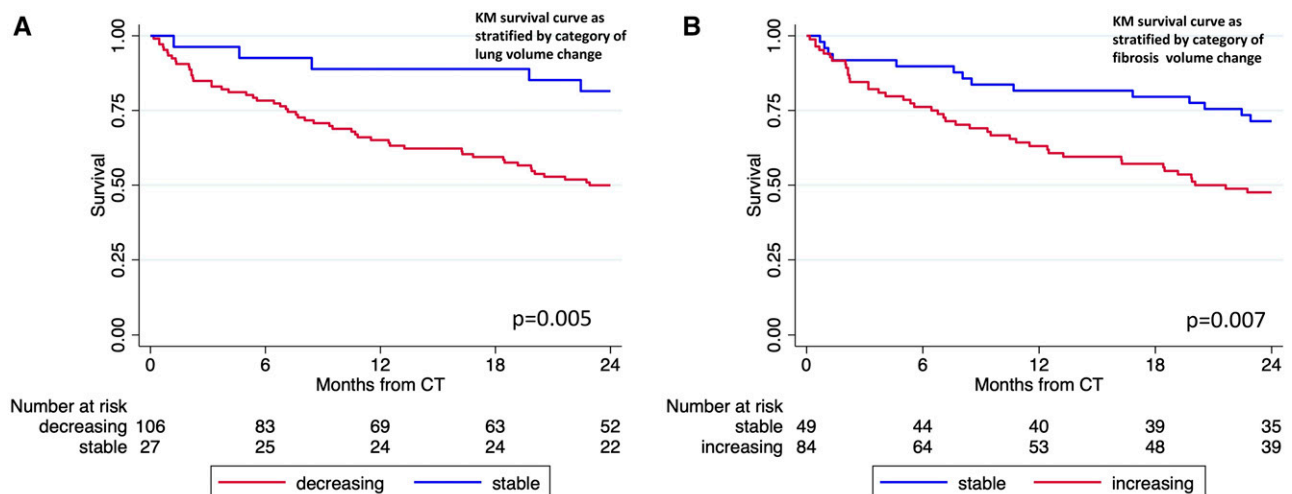
A longitudinal analysis of CT scans in patients with multiple CT scans showed that over time, there were significant reductions in lung volume and increases in fibrosis volume. This is consistent with our understanding of IPF disease progression, with progressive loss of volume and fibrosis extent. Although baseline vascular volume predicted mortality, it is interesting that we observed no significant change over time in longitudinal imaging. Our findings echo the

data linking vessels (or vessel-related structures as defined by Computer-Aided Lung Informatics for Pathology Evaluation and Rating [CALIPER]) with mortality in IPF (18, 19). It may be that although other measures (such as fibrosis volume) change as IPF progresses, the vascular volume in the lung (while being predictive of mortality) is slower to change and is therefore a less useful marker longitudinally.

Exploring associations with the segmentation models and conventional markers of disease severity in IPF, the lung volume model showed a positive correlation with FVC. We previously used this model to show a correlation over time in an independent dataset of 93 patients with CT scans and lung function data over 7 years, in which we showed a significant difference in CT lung volume decline between progressive and stable cohorts, with annual progression defined by five variables: survival time and annualized changes in FVC, FEV<sub>1</sub>, DL<sub>CO</sub>, and V<sub>A</sub> (20). It is interesting that none of the other models here were associated with

either FVC or DL<sub>CO</sub>, despite their being associated with progression (fibrosis and vessel models), mortality (fibrosis and vessel models), and change over time (fibrosis and lung volume models). This suggests that the fibrosis, vessel, and airway volume models may be measuring IPF disease severity signals that are independent of pulmonary function and may help identify subtypes of patients to study in more detail or may help explain differing treatment responses across large cohorts.

A number of visual scoring systems have been used to assess the severity of fibrotic lung disease on the basis of anatomical findings on CT (21, 22). These methods are labor intensive, subjective, and prone to poor interobserver variability. Although these visual scoring systems can be performed by radiologists, manual segmentation of individual CT scans cannot be performed at scale manually, as significant radiology expertise and time are required. The use of deep learning to segment CT scans from patients with IPF at scale is a

**Figure 4.** (A and B) KM survival curves for PROFILE participants who underwent serial CT stratified by categorical change in (A) lung volume and (B) fibrosis volume. For definition of abbreviations, see Figure 3.



novel approach that has a number of applications for the development of imaging-based biomarkers in this disease.

The advantages of this method (compared with both manual segmentation and a visual scoring system) include the speed of segmentation and the production of quantifiable data (e.g., volume of airways) compared with a scoring system that produces an aggregate of visual scores. Computer-assisted approaches to classify and quantify CT patterns have been available for nearly four decades (23, 24). CALIPER offers a computer-vision approach based on volumetric local histogram and morphologic analysis, which labels automatically the typical high-resolution CT (HRCT) parenchymal features (25). It also allows the quantification of pulmonary vascular volume, which has been shown to be a strong predictor of FVC decline in patients with IPF (26).

Another technique, quantitative lung fibrosis (QLF), uses a support vector machine to quantitatively measure honeycombing, ground-glass opacity, and a composite quantitative interstitial lung disease score (27). QLF and quantitative interstitial lung disease scores have been used as surrogate endpoints in seven IPF clinical trials to date, showing correlations with FVC change and treatment efficacy (28). An alternative method is functional respiratory imaging (FRI), in flow simulations are performed within the airways and pulmonary vessels are extracted from low-dose HRCT scans taken during inspiration and expiration (29). Changes in airway caliber using FRI have been shown to correlate with IPF disease progression (30).

The Systematic Objective Fibrotic Imaging Analysis Algorithm, a deep learning-based approach to classify usual interstitial pneumonia (UIP) on HRCT, has been shown to provide enhanced outcome prediction in patients with progressive fibrotic lung disease compared with expert radiologist evaluation or guideline-based histologic pattern (31). More recently, a larger study of more than 2,000 patients with IPF showed the development of a machine learning-based UIP classifier that classified patients as “UIP positive” if they had a significantly greater annual decline in FVC than those classified as “UIP negative” (32).

The segmentation approaches adopted in our models have a number of significant advantages and benefits over the previously described approaches. First, rather than

arbitrary scoring systems, we produce quantifiable volume data for lung, airway, vessel, and fibrosis volume. The fibrosis and airway models have been trained and segmented from a variety of real-world CT scans from multiple ( $n = 5$ ) imaging vendors and at various slice thicknesses between 0.5 and 5 mm. These models are therefore able to segment images from a wide range of CT scans in patients with IPF, unlike other methods that have rigid criteria for image suitability, such as CALIPER, which requires scans with <1-mm slice thickness with a reconstruction algorithm that is not too sharp (e.g.,  $\leq$ B70 on a Siemens scanner but not B80) (33). Our segmentation models are therefore likely to be more broadly clinically applicable.

The overall segmentation success rate across all four models on our platform was high, with almost all (97.8%) CT scans (which were all collected as part of routine clinical care) successfully segmented into the four models. This is higher than with other described computer-assisted classification approaches, which can be affected by a number of factors, including reconstruction algorithm, pixel size, and CT scanner used (34). We showed that in addition to a high rate of segmentation success, the Dice scores for our models did not differ according to any of the image attributes, such as scanner manufacturer, convolution kernel, and slice thickness. Furthermore, in comparison with techniques such as FRI that require specific protocols for patient inspiration to capture CT images (35), our models were applied to CT scans that were collected during the course of routine clinical care, with no prespecified patient breathing cycle or image specification protocol. Three of our models (fibrosis, vascular, and airway) were also created using three-dimensional convolutional neural networks (compared with two-dimensional slice models used in computer techniques such as QLF), which allows a detailed volume analysis from each CT scan.

Our work has several limitations. One is that the PROFILE CT data were collected as part of routine clinical care rather than as part of a protocolized clinical trial. However, this is also a strength, as the segmentations were performed on CT scans from multiple scanners, with multiple reconstructions and slice thicknesses, but still maintained a high segmentation success rate. The scans were obtained during inspiration but were not spirometry controlled; had they been, the

correlation between lung volume and FVC might have been stronger. Finally because the data were collected as part of routine clinical care, there was limited synchronization between CT and PFT time, and for our analysis of CT model against FVC and  $DL_{CO}$ , we chose data that were within 180 days of each other. We did show a strong correlation between FVC and lung volume, if we had had data that were more contemporaneous, we might have seen other correlations between CT models and pulmonary function.

Segmentation models have a number of uses in the management of patients with IPF. In the clinic, they might be used to provide information on disease progression to help both patients and clinicians make decisions on the timing of antifibrotic initiation or transplantation. As multiple IPF therapies become available, these models could also be used as relatively short term prognostic biomarkers to inform early change of therapy because of nonresponse to medication. The prediction of mortality from baseline CT scans can also help patients with decisions regarding quality of life versus end-of-life care. In the clinical trial setting, apart from their obvious potential as novel endpoints, these models could be used to help stratify trial participants into studies and both allow researchers to select patients with disease that is more likely to progress (and therefore increase the possibility of finding a treatment effect between study arms) and ensure that study groups are as carefully matched as possible to ensure that any difference between arms is real.

## Conclusions

We have shown the ability to automate the segmentation and interpretation of CT scans, providing prognostic information on the basis of a baseline scan on both progression-free survival and mortality in patients with IPF. Longitudinal changes in radiological lung volumes, airways, and fibrosis extent can also then identify progressive disease. These segmentation algorithms could easily be integrated into routine clinical practice to help provide prognostic information to patients with this complicated and progressive lung disease, and they show promise as future key endpoints or imaging biomarkers in clinical trials. ■

**Author disclosures** are available with the text of this article at [www.atsjournals.org](http://www.atsjournals.org).

## References

- 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.
- Karimi-Shah BA, Chowdhury BA. Forced vital capacity in idiopathic pulmonary fibrosis—FDA review of pirfenidone and nintedanib. *N Engl J Med* 2015;372:1189–1191.
- Oldham JM, Lee CT, Wu Z, Bowman WS, Pugashetti JV, Dao N, *et al.* Lung function trajectory in progressive fibrosing interstitial lung disease. *Eur Respir J* 2022;59:2101396.
- Pugashetti JV, Adegunsoye A, Wu Z, Lee CT, Srikrishnan A, Ghodrati S, *et al.* Validation of proposed criteria for progressive pulmonary fibrosis. *Am J Respir Crit Care Med* 2023;207:69–76.
- Israël-Biet D, Juvin K, Dang Tran K, Badia A, Cazes A, Delclaux C. Idiopathic pulmonary fibrosis: diagnosis and treatment in 2013 [in French]. *Rev Pneumol Clin* 2014;70:108–117.
- Nathan SD, Wanger J, Zibrak JD, Wencel ML, Burg C, Stauffer JL. Using forced vital capacity (FVC) in the clinic to monitor patients with idiopathic pulmonary fibrosis (IPF): pros and cons. *Expert Rev Respir Med* 2021;15:175–181.
- Santermans E, Ford P, Kreuter M, Verbruggen N, Meyvisch P, Wuyts WA, *et al.* Modelling forced vital capacity in idiopathic pulmonary fibrosis: optimising trial design. *Adv Ther* 2019;36:3059–3070.
- Hutchinson JP, Fogarty AW, McKeever TM, Hubbard RB. In-hospital mortality after surgical lung biopsy for interstitial lung disease in the United States: 2000 to 2011. *Am J Respir Crit Care Med* 2016;193:1161–1167.
- Oh AS, Lynch DA, Swigris JJ, Baraghoshi D, Dyer DS, Hale VA, *et al.* Deep learning–based fibrosis extent on CT predicts outcome of fibrosing interstitial lung disease independent of visually assessed CT pattern. *Ann Am Thorac Soc* 2024;21:218–227.
- Kunihiro Y, Matsumoto T, Murakami T, Shimokawa M, Kamei H, Tanaka N, *et al.* A quantitative analysis of long-term follow-up computed tomography of idiopathic pulmonary fibrosis: the correlation with the progression and prognosis. *Acta Radiol* 2023;64:2409–2415.
- Du K, Zhu Y, Mao R, Qu Y, Cui B, Ma Y, *et al.* Medium-long term prognosis prediction for idiopathic pulmonary fibrosis patients based on quantitative analysis of fibrotic lung volume. *Respir Res* 2022;23:372.
- Robbie H, Daccord C, Chua F, Devaraj A. Evaluating disease severity in idiopathic pulmonary fibrosis. *Eur Respir Rev* 2017;26:170051.
- Ley B, Ryerson CJ, Vittinghoff E, Ryu JH, Tomassetti S, Lee JS, *et al.* A multidimensional index and staging system for idiopathic pulmonary fibrosis. *Ann Intern Med* 2012;156:684–691.
- Maher TM. PROFILEing idiopathic pulmonary fibrosis: rethinking biomarker discovery. *Eur Respir Rev* 2013;22:148–152.
- Hofmanninger J, Prayer F, Pan J, Röhrich S, Prosch H, Langs G. Automatic lung segmentation in routine imaging is primarily a data diversity problem, not a methodology problem. *Eur Radiol Exp* 2020;4:50.
- Frangi RF, Niessen WJ, Vincken K, Viergever MA. Multiscale vessel enhancement filtering. In: Wells WM, Colchester A, Delp S, editors. *Medical Image Computing and Computer-Assisted Intervention—MICCAI '98*. New York: Springer; 1998. pp. 130–137.
- Hepzibah R, Anandharaj HC, Kowsalya G, Jayanthi R, Chandy DA. Review on deep learning methodologies in medical image restoration and segmentation. *Curr Med Imaging* 2023;19:844–854.
- Jiao XY, Song H, Liu WW, Yang JL, Wang ZW, Yang D, *et al.* The effect of CALIPER-derived parameters for idiopathic pulmonary fibrosis in predicting prognosis, progression, and mortality: a systematic review. *Eur Radiol* 2023;33:7262–7273.
- Jacob J, Bartholmai BJ, 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.
- Manners H, McLellan T, Kirov K, Roberts M, Kanavati F, McKenzie G, *et al.* Data-driven multivariate cohort selection and automated CT scan lung volume measurement identifies disease progression in idiopathic pulmonary fibrosis (IPF). *Eur Respir J* 2022;60:4502.
- Jacob J, Aksman L, Mogulkoc N, Procter AJ, Gholipour B, Cross G, *et al.* Serial CT analysis in idiopathic pulmonary fibrosis: comparison of visual features that determine patient outcome. *Thorax* 2020;75:648–654.
- 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.
- Sakai N, Mishima M, Nishimura K, Itoh H, Kuno K. An automated method to assess the distribution of low attenuation areas on chest CT scans in chronic pulmonary emphysema patients. *Chest* 1994;106:1319–1325.
- Rodriguez LH, Vargas PF, Raff U, Lynch DA, Rojas GM, Moxley DM, *et al.* Automated discrimination and quantification of idiopathic pulmonary fibrosis from normal lung parenchyma using generalized fractal dimensions in high-resolution computed tomography images. *Acad Radiol* 1995;2:10–18.
- Zavaletta VA, Bartholmai BJ, Robb RA. High resolution multidetector CT-aided tissue analysis and quantification of lung fibrosis. *Acad Radiol* 2007;14:772–787.
- Jacob J, Pienn M, Payer C, Urschler M, Kokosi M, Devaraj A, *et al.* Quantitative CT-derived vessel metrics in idiopathic pulmonary fibrosis: a structure-function study. *Respirology* 2019;24:445–452.
- Chong DY, Kim HJ, Lo P, Young S, McNitt-Gray MF, Abtin F, *et al.* Robustness-driven feature selection in classification of fibrotic interstitial lung disease patterns in computed tomography using 3D texture features. *IEEE Trans Med Imaging* 2016;35:144–157.
- Wu X, Kim GH, Salisbury ML, Barber D, Bartholmai BJ, Brown KK, *et al.* Computed tomographic biomarkers in idiopathic pulmonary fibrosis: the future of quantitative analysis. *Am J Respir Crit Care Med* 2019;199:12–21.
- De Backer J, Van Holsbeke C, Vos W, Vinchurkar S, Dorinsky P, Rebello J, *et al.* Assessment of lung deposition and analysis of the effect of fluticasone/salmeterol hydrofluoroalkane (HFA) pressurized metered dose inhaler (pMDI) in stable persistent asthma patients using functional respiratory imaging. *Expert Rev Respir Med* 2016;10:927–933.
- McLellan T, George PM, Ford P, De Backer J, Van Holsbeke C, Mignot B, *et al.* Idiopathic pulmonary fibrosis: airway volume measurement identifies progressive disease on computed tomography scans. *ERJ Open Res* 2020;6:00290-2019.
- Walsh SLF, Mackintosh JA, Calandriello L, Silva M, Sverzellati N, Larici AR, *et al.* Deep learning-based outcome prediction in progressive fibrotic lung disease using high-resolution computed tomography. *Am J Respir Crit Care Med* 2022;206:883–891.
- Humphries SM, Thieke D, Baraghoshi D, Strand MJ, Swigris JJ, Chae KJ, *et al.* Deep learning classification of usual interstitial pneumonia predicts outcomes. *Am J Respir Crit Care Med* 2024;209:1121–1131.
- Jankharia BG, Angirish BA. Computer-aided quantitative analysis in interstitial lung diseases—a pictorial review using CALIPER. *Lung India* 2021;38:161–167.
- Kim H, Goo JM, Ohno Y, Kauczor HU, Hoffman EA, Gee JC, *et al.* Effect of reconstruction parameters on the quantitative analysis of chest computed tomography. *J Thorac Imaging* 2019;34:92–102.
- Leemans G, Belmans D, Van Holsbeke C, Kushnarev V, Sugget J, Ides K, *et al.* A functional respiratory imaging approach to the effect of an oscillating positive expiratory pressure device in chronic obstructive pulmonary disease. *Int J Chron Obstruct Pulmon Dis* 2020;15:1261–1268.