# Evaluation of e-Lung automated quantitative computed tomography biomarkers in idiopathic pulmonary fibrosis

Peter M. George [1,2,9], Christian Rennison-Jones[3], Giacomo Benvenuti[3], Aliki Sifostratoudaki[3], Finja A. Ottink[3], Wassim Bou-Zeid[3], Roberta E. Ledda[4], Roqia F. Abul Kadir[5], Bushra Johari[5], Claire Fernandez[3], George W.J. Harston [3,6], Olivier Joly[3], Stephen Gerry[7] and Anand Devaraj[2,8,9]

[1]Department of Interstitial Lung Disease, Royal Brompton Hospital, London, UK. [2]National Heart and Lung Institute, Imperial College, London, UK. [3]Brainomix, Oxford, UK. [4]Department of Medicine and Surgery, University of Parma, Parma, Italy. [5]Department of Radiology, Faculty of Medicine, Universiti Teknologi MARA, Sungai Buloh, Malaysia. [6]Oxford University Hospitals NHSFT, Oxford, UK. [7]Centre for Statistics in Medicine, University of Oxford, Oxford, UK. [8]Department of Radiology, Royal Brompton Hospital, London, UK. [9]P.M. George and A. Devaraj contributed equally to this article as lead authors and supervised the work.

Corresponding author: Peter M. George (p.george@rbht.nhs.uk)

Shareable abstract (@ERSpublications)
**e-Lung is an AI-powered technology that provides prognostic information from a baseline standard CT scan and, with a defined minimally clinical important difference, is more sensitive to serial change than conventional tools in the assessment of IPF** https://bit.ly/3VEuPMI

**Cite this article as:** George PM, Rennison-Jones C, Benvenuti G, *et al*. Evaluation of e-Lung automated quantitative computed tomography biomarkers in idiopathic pulmonary fibrosis. *ERJ Open Res* 2024; 10: 00570-2024 [DOI: 10.1183/23120541.00570-2024].



Received: 4 June 2024
Accepted: 16 June 2024

## Abstract

*Background* In patients with idiopathic pulmonary fibrosis (IPF) there is a need to identify biomarkers that 1) are associated with increased risk of adverse outcome and 2) can be used to monitor treatment response or identify disease progression over time.

*Methods* Two consecutive cohorts of patients with IPF were accessed from the Open Source Imaging Consortium database. Automated computed tomography (CT) biomarkers of disease severity incorporating fibrotic and pulmonary vascular features (the reticulovascular score and weighted reticulovascular score (WRVS)) were studied. Relationships between imaging biomarkers, lung function and survival were analysed.

*Results* In separate test and validation cohorts, 168 and 176 patients with IPF respectively (median survival 2.6 years) were studied. A threshold of WRVS ⩾15% at baseline CT was most strongly associated with transplant-free survival (HR 3.00, 95% CI 1.47–6.10, p=0.002) when adjusted for baseline forced vital capacity (FVC) and age. In patients with 12-month follow-up CT and lung function tests (n=89) an increase in 3% of WRVS (the minimal clinically important difference) was also significantly associated with reduced survival independent of FVC, and outperformed visual evaluation of progressive fibrosis.

*Conclusions* WRVS is an automated CT biomarker which can identify patients with IPF at increased risk of progression and is able to reliably capture disease progression over time.

## Introduction

Idiopathic pulmonary fibrosis (IPF) is a chronic progressive interstitial lung disease (ILD) characterised by irreversible architectural destruction and lung function decline [1, 2]. While antifibrotic therapies have been shown to improve patient outcomes in both clinical trials [3, 4] and real world settings [5], studies are ongoing to identify novel therapies that may further improve outcomes for these patients [6, 7].

Although many patients with IPF will experience rapid decline, the clinical course of IPF is variable, and some patients deteriorate at a much slower rate. Recruiting stable patients to new clinical trials in IPF may potentially negatively influence the success of these trials. On the other hand, enriching trial cohorts with patients at increased risk of progression is an accepted method to enhance the prospect of detecting a clinically meaningful effect [8].

The preferred Food and Drug Administration (FDA) clinical trial end-point for therapeutic trials in ILD studies is change in forced vital capacity (FVC) over time, which is acknowledged as an accepted surrogate for mortality [9]. However, FVC is subject to substantial variability and inconsistency in part due to the effort-dependent nature of this test. This variability in FVC represents a further potential threat to the emergence of new IPF drugs in clinical trials particularly when considering the need to demonstrate improvement above standard of care. There is therefore a clear need to identify novel and robust methods by which IPF disease progression can be 1) accurately predicted and 2) assessed over time. Validating these novel methods requires demonstration of links to clinically meaningful end-points such as mortality.

Over the past decade, there has been substantial interest in the application of artificial intelligence (AI) and machine learning approaches to the high-resolution computed tomography (HRCT) assessment of IPF [10]. Many studies have shown that automated HRCT biomarkers can provide prognostic value in patients with IPF from a baseline computed tomography (CT) scan [11, 12], often focussing on the quantification of reticular abnormalities [13, 14]. Meanwhile other studies evaluating automated tools that segment the lung vasculature have shown promise by demonstrating a relationship between pulmonary vascular volume and outcomes [15–17].

The aim of this study was to evaluate the performance of an automated biomarker that quantifies a combination of both reticulation and pulmonary vascular structures on CT (reticulovascular score (RVS) and weighted reticulovascular score (WRVS), e-Lung, Brainomix, Oxford, UK) [18, 19]. WRVS thresholds have previously been shown to be associated with risk of 52-week FVC decline in a retrospective analysis of a completed IPF randomised control clinical trial [20], but the association of WRVS with long-term outcomes in IPF is not known neither is its sensitivity to serial change on CT. The performance of e-Lung was evaluated by determining: 1) links between baseline CT biomarker scores and future FVC decline and mortality; and 2) associations between change in CT biomarker scores over time and mortality. Cases were obtained from a large open source dataset, and performance was compared to that of FVC and visual radiological assessment of CT.

## Methods

### *Patient clinical and CT data*

HRCT scans and matched clinical data from patients with IPF were accessed from the Open Source Imaging Consortium (OSIC) in May 2022. OSIC is a non-for-profit organisation with a General Data Protection Regulation (GDPR) secure data repository which contains anonymised data from thousands of patients with ILD contributed by clinical, academic and industry partners across the world. The OSIC Data Repository has Central Institutional Review Board (IRB) and multiple institution IRB approvals.

Consecutive patients from the OSIC database who met the following inclusion criteria were selected for the study: patients with a diagnosis of IPF as submitted by the contributing site, availability of at least one non-contrast enhanced volumetric thoracic CT with slice thickness $\leqslant$2.5 mm and lung function tests within 12 weeks of the HRCT scan. Patients with CT scans which did not capture the entire thorax or with excessive motion artefact or expiratory CTs as adjudicated by study investigators were excluded. Other clinical data accessed were patient demographics and survival time from date of CT scan to time of death or to time of censoring. The final study group comprised 344 patients with baseline CT, divided randomly into a test and validation cohort of 168 and 176 subjects respectively. From the entire cohort, 107 patients had adequate serial CT scans 12 months apart. Of these, 89 patients had contemporaneous lung function tests at both CT time points (figure 1). This subset was used to evaluate serial change of e-Lung biomarkers.

### *CT visual radiological evaluation*

In those cases where serial imaging was available, three thoracic radiologists with between 3 and and 10 years' experience, blinded to clinical information, independently reviewed and compared follow-up CT with baseline CT side by side to determine changes in fibrosis extent [21]. This was scored on a 5-point scale: definite and unequivocal improvement, possible improvement, stability, possible worsening or definite and unequivocal worsening. CT reading was performed using Open Health Imaging Foundation (OHIF) Viewer plugin on an Extensible Neuroimaging Archive Toolkit (XNAT) platform. To achieve a consensus response for each CT, cases where at least two readers scored follow-up scans as showing probable or definite progression were classified as progressive fibrosis.

### *e-Lung automated CT biomarker evaluation*

e-Lung (Brainomix) uses multi-class convolutional neural network (CNN) techniques based on the 2D UNet architecture to automatically segment the lung parenchyma on CT, and then detect and quantify lung patterns and structures based on imaging features including shape and intensity. CNNs are a form of deep

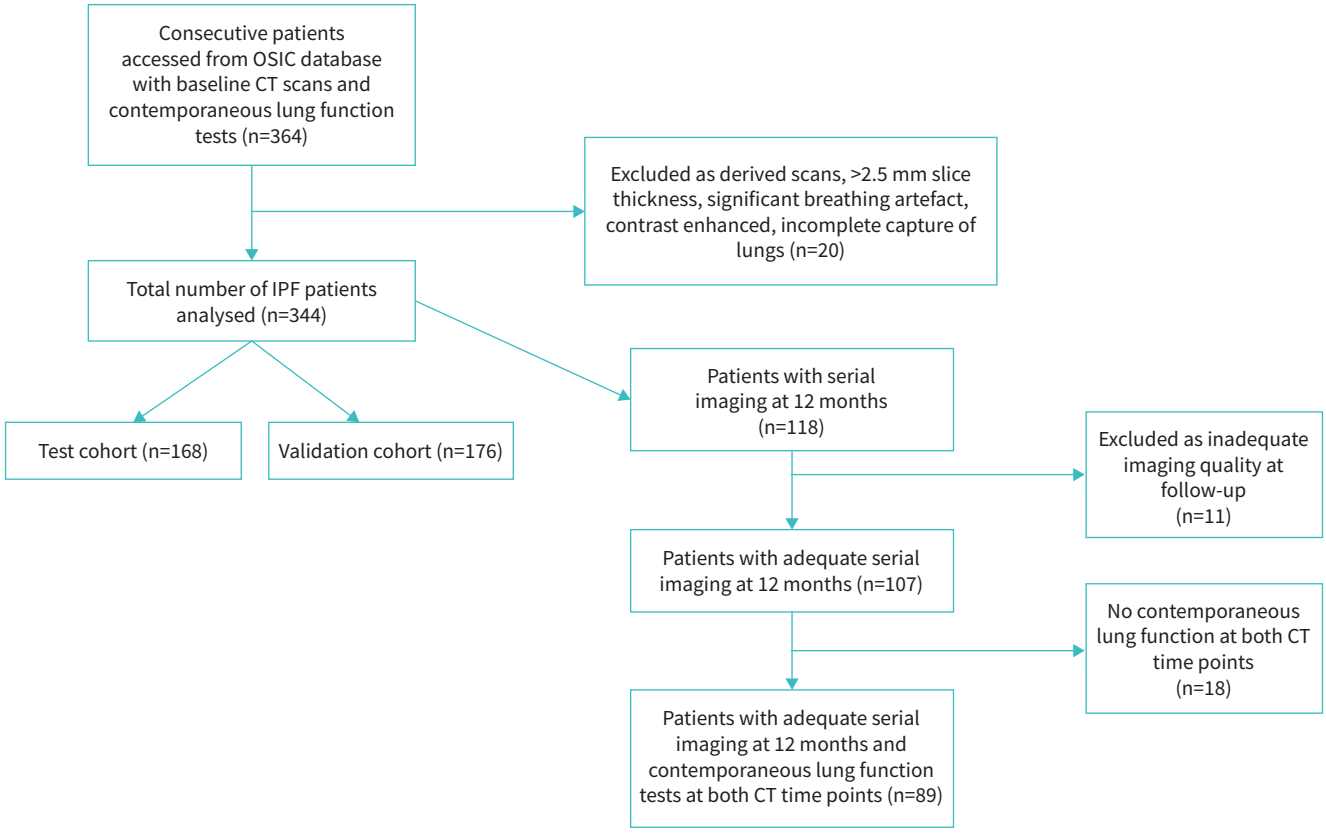


FIGURE 1 Consort diagram illustrating test and validation cohort derivation. OSIC: Open Source Imaging Consortium; CT: computed tomography; IPF: idiopathic pulmonary fibrosis.

learning that contrasts with traditional hand-crafted feature engineered automated disease quantification. The RVS biomarker (quantified as a percentage of lung) is calculated automatically by identifying all voxels within the lung volume belonging to branching vessel-like regions or linear structures with density equivalent to pulmonary vessels (figure 2a–c). The WRVS is the sum of these voxels after applying a weighting based on the relative voxel positions within the lungs, expressed as a percentage of total lung voxels. The weighting assigns more importance to peripheral regions of each lung, based on Euclidean distance, on the understanding that peripheral fibrotic disease is known to be associated with poorer prognosis [22]. The final WRVS score is expressed as a percentage of peripheral lung. Further details on the development of the e-Lung algorithm are provided in the supplementary methods.

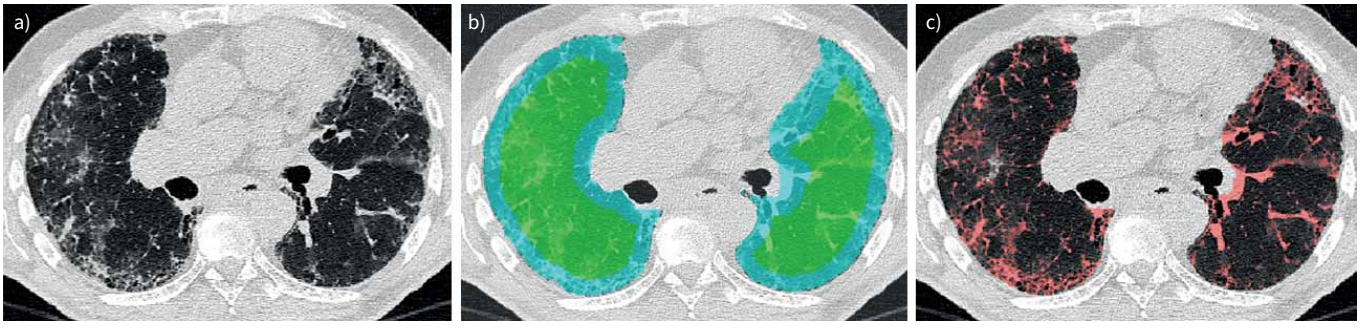


FIGURE 2 Patient with idiopathic pulmonary fibrosis: a) axial thoracic high-resolution computed tomography; b) e-Lung lung segmentation with the peripheral portion used in the weighted reticulovascular score in cyan; c) e-Lung segmentation of reticular and vascular structures producing reticulovascular score.

### *Statistical methods*

The characteristics of the cohort are presented using mean±SD or median and interquartile range (where appropriate) for continuous variables, and number and percentage for binary and categorical variables.

To model the relationships between biomarkers and transplant-free survival we used Cox regression, where we censored at the last known follow-up time point compared to time of CT acquisition. To assess the strength of any relationships we report hazard ratios (HRs) with 95% confidence intervals (CI). Where different predictors are being compared, but are quantified using different units, the HR is not directly comparable, and therefore we additionally report the Harrell's C-Index, which can be directly compared (a measure of concordance ranging between 0 and 1, where 0.5 indicates no relationship, and 1 indicates perfectly correct prediction). We plotted Kaplan–Meier curves for binary and categorical variables. Prior to reporting results from the survival analyses we assessed whether the proportion hazards assumption was valid.

To assess the relationship between biomarkers and 12-month FVC decline, we first used a mixed effects model to estimate the level of FVC at 12 months based on serial measurements. The model incorporated all measurements of FVC in a random coefficients model with a polynomial term to allow for a nonlinear trajectory. Patient number was included as a random effect. The predicted value at 12 months was used to define decline relative to the baseline measurement in litres with a clinically significant change considered to be ⩾10%. Additionally, if patients died before 12 months, we included them in the analysis as decliners. The outcome was then used in a logistic regression model to determine odds ratios (OR) for predictive biomarkers.

To assess biomarker performance, the dataset was randomly divided into test and validation cohorts. Clinically applicable binary thresholds capable of identifying patients at risk of mortality were created based on the median values [23] of RVS and WRVS at baseline, and performance evaluated in the test cohort. Median values have been utilised to identify suggested cut-points in previously published work [15]. Additionally, the optimal thresholds were identified that maximised the log-rank test statistic. Replication of performance was tested in the validation cohort also using Cox regression. Demographics of the two cohorts are presented to ensure that they were well matched.

Univariable and multivariable Cox regression was used to determine the relationship between change in automated biomarkers between baseline and 1 year follow-up CT, and outcome in those patients with follow-up CT. The minimum clinically important difference (MCID) for the imaging biomarker was calculated by using the distribution-based approach, in the absence of any "anchor" for an anchor-based approach. This approach estimates the MCID to be half of the standard deviation of the change in the biomarkers between baseline and follow-up [24]. To assess the performance of MCID at determining survival, we used univariable and multivariable Cox regression analyses, and compared this to visual assessment of radiological progression and FVC decline of 10%. Interobserver agreement for radiological progression was expressed using interclass correlation coefficient (ICC).

## Results

Across the entire dataset of 344 patients, 290 (84.4%) were male, with a mean age of 69 years, mean FVC of 2.8 L (81% predicted) and median survival of 2.6 years (table 1). The test cohort (n=168) and validation cohort (n=176) were matched for disease extent and demographics (table 1). CT scans were performed using a spectrum of acquisition protocols from a variety of scanner manufacturers (supplementary table S1). Of 107 patients with repeat CT scans available for analysis, 47 (43.9%) had differences in one or more acquisition parameter at follow-up compared to baseline (supplementary table S2).

### *Identifiers of outcome from baseline HRCT scan*

In the test cohort, RVS and WRVS were found to be highly correlated (Pearson correlation coefficient 0.976, p<0.001). RVS and WRVS provided comparable significant associations with transplant-free survival when adjusted for age and baseline FVC: HR 1.09 (95% CI 1.05–1.14) and HR 1.07 (95% CI 1.03–1.11), C-index=0.76 for RVS and WRVS respectively.

For clinical applicability, binary thresholds were created for RVS and WRVS, based on the median values (12.4 and 15.1 respectively) rounded to nearest 5% (10% and 15% respectively), and their association with survival was tested. A threshold of ⩾15% WRVS was more strongly associated with transplant-free survival (table 2 and figure 3a–d). The same thresholds (*i.e.* 10% and 15%) were found to be the optimal thresholds which maximised the log-rank test statistic.

**TABLE 1** Patient demographics

| | Test cohort | Validation cohort | Total |
|---|---|---|---|
| **Patients n** | 168 | 176 | 344 |
| **Male, n (%)** | 140 (84) | 150 (85) | 290 (84.4) |
| **Age years, mean±SD** | 69±8.7 | 69±8.2 | 69±8.5 |
| **FVC L, mean±SD** | 2.9±0.8 | 2.8±0.8 | 2.8±0.8 |
| **% FVC, mean±SD** | 82.1±18.5 | 79.7±18.7 | 80.9±18.6 |
| **$D_{LCO}$ mL·min$^{-1}$·kPa$^{-1}$, mean±SD** | 10.1±5.8 | 9.8±5.5 | 9.9±5.6 |
| **Smoking history, n (%)** | | | |
| Unknown | 8 (5) | 6 (3) | 14 (3.9) |
| Never-smoker | 43 (26) | 40 (22) | 83 (24.2) |
| Ex-smoker | 103 (61) | 116 (66) | 219 (63.9) |
| Active smoker | 14 (8) | 14 (8) | 28 (8.1) |
| **Survival status** | | | |
| Deceased, n (%) | 55 (32) | 61 (34) | 116 (33.1) |
| Lung transplant recipient, n (%) | 4 (2) | 8 (4) | 12 (3.3) |
| Median±SEM survival years | 2.46±0.28 | 2.69±0.23 | 2.6±0.18 |

FVC: forced vital capacity; $D_{LCO}$: diffusion capacity of the lung for carbon monoxide.

In the test and validation cohorts a WRVS threshold of ⩾15% retained significance when corrected for baseline FVC and age (HR 4.75 (95% CI 2.63–8.56), p<0.001; and HR 3.00 (95% CI 1.47–6.10, p=0.002, respectively). To understand whether the segmentation of reticulovascular structures by e-Lung was comparable to the identification of high-attenuation areas (HAAs) (a more traditional densitometry-based metric), we performed a *post hoc* analysis evaluating the association between HAA% (percentage lung between −250 and −600 Hounsfield units) [25] above median threshold and transplant-free survival. We found the performance of HAA to be inferior to WRVS and RVS with a HR of 2.60 (95% CI 1.27–5.32).

### *Identifying transplant-free survival from serial change in CT and lung function over 1 year*

There was poor to moderate agreement for visual progression of fibrosis on CT between radiologists (ICC 0.48–0.60). Rise in WRVS (C-index 0.72) and decline in FVC% (C-index 0.72) over 12 months were most strongly associated with survival, with visually assessed increase in extent of fibrosis performing less well (C-index 0.65) (table 3).

The ability of binary thresholds in change of e-Lung and lung function and radiological parameters over time to identify transplant-free survival was evaluated. The MCID in WRVS on serial CT was calculated as 3% using a distribution-based method, and was compared to visual radiological progression and relative FVC decline of at least 10%, an accepted clinical trial end-point [4].

An increase in WRVS of at least 3% was more closely associated with transplant-free survival (HR 6.30, p<0.001) (figure 4) than a relative FVC (litres) reduction of 10% (HR 2.47, p<0.001) and radiological progression on CT assessed visually (HR 5.49, p<0.001) (table 4). Sensitivity analyses demonstrated that there was no difference in the survival of patients with and without serial imaging at 12 months excluding this as a source of potential bias (supplementary figure S1a–b) nor was there a difference in WRVS % by scanner manufacturer (supplementary figure S2).

### *Identifying FVC decline or mortality at 1 year*

Using mixed effects models with deaths within the first 12 months imputed as an event, the same variables were evaluated for their ability to identify ⩾10% relative FVC decline (litres) or mortality over the

**TABLE 2** Univariate analysis of binary thresholds of weighted reticulovascular score (WRVS) and reticulovascular score (RVS) biomarkers against transplant-free survival in test and validation cohorts

| | Test cohort | | | Validation cohort | | |
|---|---|---|---|---|---|---|
| **Variable** | **Proportion above threshold** | **HR (95% CI)** | **p-value** | **Proportion above threshold** | **HR (95% CI)** | **p-value** |
| **WRVS % (<15 *versus* ⩾15)** | 85/168 | 4.82 (2.66–8.73) | <0.001 | 85/176 | 4.44 (2.53–7.77) | <0.001 |
| **RVS % (<10 *versus* ⩾10)** | 111/168 | 3.50 (1.81–6.78) | <0.001 | 119/176 | 3.41 (1.78–6.54) | <0.001 |

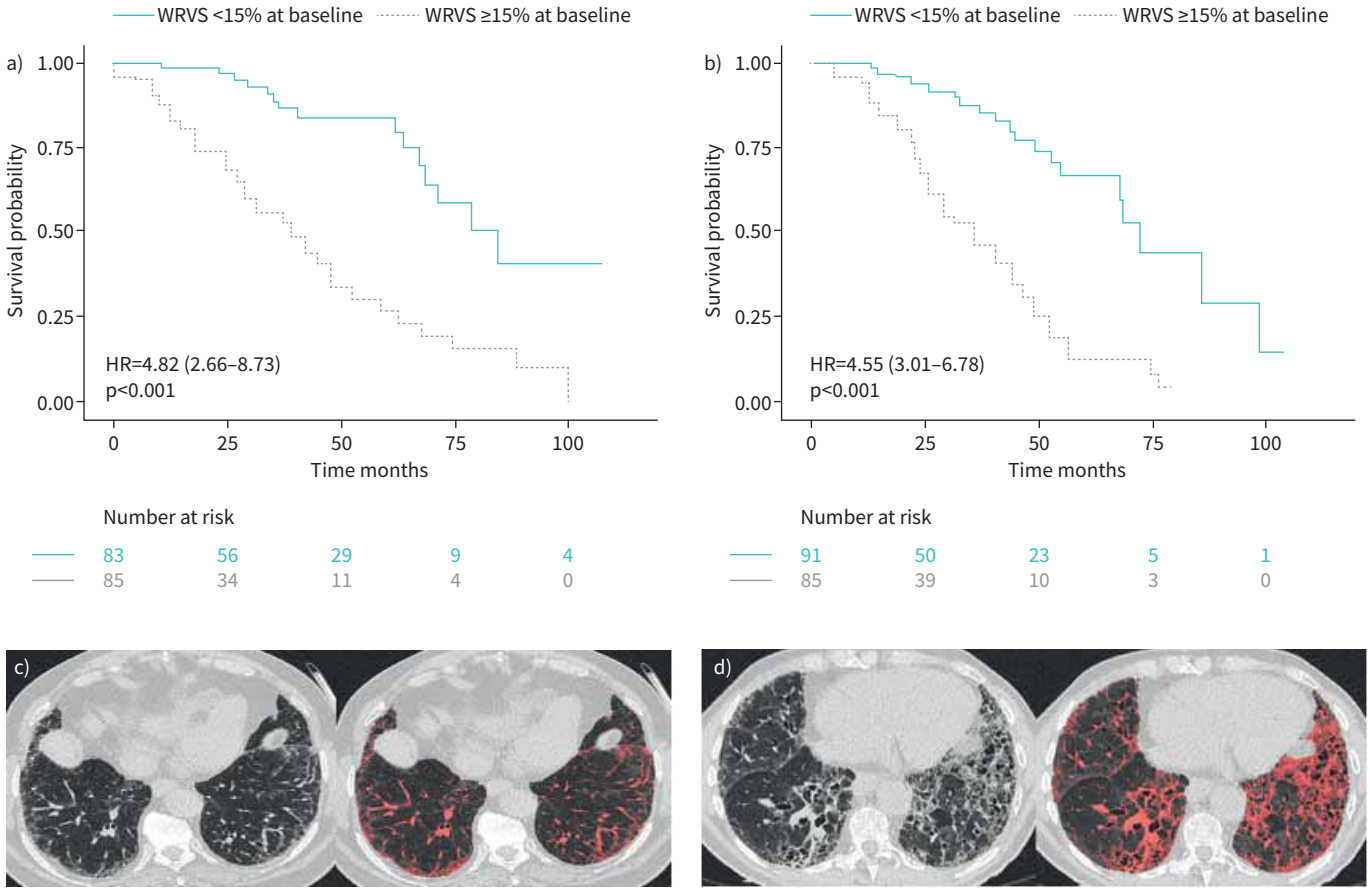


FIGURE 3 Kaplan–Meier curves showing survival separation of high- and non-high-risk groups in idiopathic pulmonary fibrosis (IPF) in a) test and b) validation cohorts using e-Lung weighted reticulovascular score (WRVS) threshold of 15% on baseline computed tomography. HRCT and e-lung images in patients with IPF with c) non-high-risk WRVS score (10%) and d) high-risk WRVS score (25%). HR: hazard ratio.

following 12 months from the time of the baseline CT scan. Adjusted for age and baseline FVC, a WRVS threshold of at least 15% was not associated with future FVC decline in the test cohort (OR 1.57, 95% CI 0.68–3.68, p>0.05), but was predictive in the validation cohort (OR 3.12, 95% CI 1.41–7.29, p=0.006).

## Discussion

In this study, we have validated the performance of an automated HRCT biomarker (the e-Lung WRVS) in a large and geographically diverse group of patients with IPF. In test and validation cohorts we have shown that the WRVS is prognostically associated with transplant-free survival from a baseline HRCT scan.

TABLE 3 Performance of change in markers from baseline to 1 year to identify subsequent transplant-free survival in patients with serial computed tomography (CT) across entire cohort

| | Univariable | | | Age-adjusted | | | Multivariable[#,¶] | | |
|---|---|---|---|---|---|---|---|---|---|
| Variable | n | HR (95% CI) | C-Index | n | HR (95% CI) | C-Index | n | HR (95% CI) | C-Index |
| WRVS % | 107[+] | 1.16 (1.10–1.23) | 0.72 | 107 | 1.16 (1.10–1.23) | 0.74 | 89 | 1.06 (0.98–1.15) | 0.72 |
| FVC L | 89[§] | 0.91 (0.88–0.94) | 0.72 | 89 | 0.91 (0.88–0.94) | 0.72 | | 0.95 (0.91–0.99) | |
| Visually assessed CT progression | 107 | 5.49 (2.75–10.97) | 0.65 | 107 | 5.51 (2.74–11.09) | 0.68 | | 2.66 (0.91–7.76) | |

WRVS: weighted reticulovascular score; FVC: forced vital capacity. [#]: from a multivariable model including change in WRVS, change in FVC, radiological change and age. [¶]: when including baseline WRVS, age and sex, HR is 1.13 (1.07–1.18), p<0.001, C-Index is 0.80. [+]: 107 patients with data on both the change in WRVS and visually assessed progression. The CT scans of 11 patients were excluded due to being considered of inadequate quality by readers. [§]: 89 patients with serial data on FVC change from baseline.

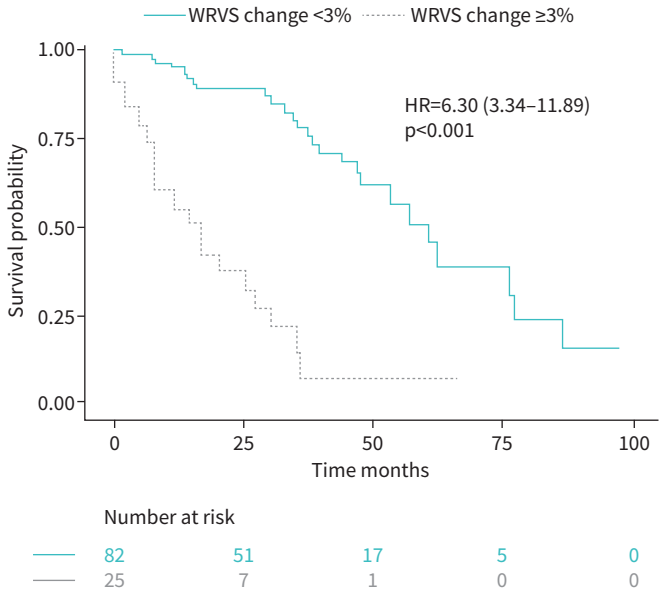


FIGURE 4 Kaplan–Meier curves showing survival separation in patients with idiopathic pulmonary fibrosis according to change in e-Lung weighted reticulovascular score (WRVS) of ⩾3% or <3% on serial computed tomography over 12 months from baseline. HR: hazard ratio.

This study has also shown the ability of e-lung to detect changes in the WRVS biomarker on serial CTs over 12 months that were associated with mortality, independent of FVC change and visually assessed fibrosis changes on multivariable analysis.

One step towards translating these findings into use in clinical practice or therapeutic trials is validation of numerical thresholds of severity or change over time that can predict outcome. We tested two thresholds, namely 15% WRVS at baseline and 3% change of WRVS on serial CT. These thresholds were also found to be associated with transplant-free survival independent of lung function or visually assessed CT parameters. The ability of WRVS to identify patients above a certain threshold (15% WRVS) at increased risk of adverse outcome may have relevance for clinical trial design in patients with IPF. These results confirm findings from a previous publication where the same WRVS threshold of 15% was associated with high risk of FVC decline in an IPF clinical trial [20]. Using e-Lung as an enrichment tool could potentially increase the likelihood of identifying real treatment effects in trials where the control arm comprises patients on standard of care antifibrotic therapy. Furthermore, while further research is needed to validate these findings in other cohorts, including in populations where information on antifibrotic therapy is known, the results show potential for change in WRVS over time to be used as an end-point in clinical trials as a secondary measure alongside FVC. Finally, ensuring that patients in control and treatment arms are well matched for WRVS may be an important clinical trial design consideration to improve the

TABLE 4 Performance of binary thresholds of change in markers from baseline to 1 year and association with transplant-free survival

| | Univariable | | | Age-adjusted | | | Multivariable[#] | | |
|---|---|---|---|---|---|---|---|---|---|
| Variable | n | HR (95% CI) | p-value | n | HR (95% CI) | p-value | n | HR (95% CI) | p-value |
| **WRVS MCID (<3% decline *versus* ⩾3% decline)** | 107[¶] | 6.30 (3.34–11.89) | <0.001 | 107 | 6.38 (3.38–12.04) | <0.001 | 89 | 2.99 (1.08–8.29) | 0.035 |
| **FVC (<10% decline *versus* ⩾10% decline)** | 89[+] | 2.47 (2.13–2.86) | <0.001 | 89 | 6.02 (3.03–11.98) | <0.001 | | 3.26 (1.44–7.37) | 0.005 |
| **Visually assessed CT progression** | 107 | 5.49 (2.75–10.97) | <0.001 | 107 | 5.51 (2.74–11.09) | <0.001 | | 2.46 (0.83–7.28) | 0.103 |

WRVS: weighted reticulovascular score; MCID: minimal clinically important difference; FVC: forced vital capacity; CT: computed tomography. [#]: from a multivariable model including change in WRVS, binary change in FVC, radiological change and age. [¶]: 107 patients with data on both the change in WRVS and visually assessed progression. The CT scans of 11 patients were excluded due to being considered of inadequate quality by readers. [+]: 89 patients with serial data on FVC change from baseline.

likelihood of capturing a true treatment effect, by ensuring equivalent disease severity and future disease behaviour in both groups.

WRVS may also be of relevance in clinical practice. In this study, radiologist assessment of disease progression was variable (with ICC 0.48–0.60 for observer agreement) and less powerful than the automated biomarker. e-Lung may be able to identify patients with visually occult progressive fibrosis who may benefit from a change in therapeutic approach at an earlier time point than can be appreciated with current tools. Automated tools also potentially mitigate issues of observer variation when fibrosis progression is visually assessed.

These data build on a field which is increasingly aware of the opportunities which AI techniques provide to quantify disease severity in lung fibrosis [10]. Previous studies have highlighted the importance of automated pulmonary vasculature assessment as a measure of disease severity in patients with IPF [15, 26]. However, it is acknowledged that separating pulmonary vessels from peripheral reticulation on HRCT using automated techniques is highly challenging [27]. Applications that segment the pulmonary vasculature may inevitably also include areas of reticulation and vice versa [14, 15]. In this study we aimed to capture the prognostic signal of both the pulmonary vasculature and reticulation by deliberately segmenting both components into a combined biomarker using a CNN-derived algorithm that focussed on segmenting tubular structures. CNNs and other deep learning algorithms have advantages over traditional hand-crafted feature engineered machine learning techniques in that they can be more easily adapted to a range of CT acquisition conditions and patient states (*e.g.* patient size, sex, scanner positioning, inspiration status), and are capable of learning to automatically detect, represent and combine complex imaging features, such as tubular structures, without requiring an expert to hand-tune an algorithm for each possible imaging protocol or disease presentation.

Another approach to automated evaluation in lung fibrosis has been to use CT densitometry-based metrics such as quantifying the percentage of lung within a certain density range (HAAs), which has previously been shown to be associated with adverse outcomes in IPF [25, 28]. Our results suggest that our approach provides additional value over HAA quantification.

The concept that disease location in IPF has prognostic importance has been raised previously. The SOFIA algorithm is a deep learning tool that generates a usual interstitial pneumonia (UIP) likelihood score that has been shown to be prognostic in patients with IPF, and where regions of peripheral disease on CT contribute most to the score [12, 22]. Similarly in our study, we find that the RVS weighted towards the periphery of the lung provides improved prognostic information. Further studies will be required to determine the prognostic values of WRVS in patients with non-IPF fibrosis.

The association identified in this study between WRVS change over time and mortality provides support for its use as a clinically meaningful end-point. However, we also aimed to determine the link between baseline WRVS and future FVC decline to identify whether WRVS-related mortality was likely mechanistically linked to progression of respiratory disease. We found no significant relationship between WRVS $\geqslant$15% and FVC decline of 10% in the test cohort, but a strong relationship between these variables in the validation cohort. As the two cohorts were well matched for disease severity, understanding the reasons for this discordance may relate to the small event rate but does require further investigation. A major strength of the study is the generalisability of the findings with patient scans and clinical data contributed from all around the globe, with a variety of acquisition protocols, as part of the OSIC database. The generalisability of data analysed in this study is important when considering how e-Lung may be utilised in prospective clinical trials and routine healthcare.

We showed in this study the prognostic abilities of WRVS in patients with IPF independent of FVC. However, a limitation in this study is that while we did have access to diffusion capacity of the lung for carbon monoxide ($D_{LCO}$) raw values, the lack of availability of percentage predicted data $D_{LCO}$ in the OSIC database meant we were unable to benchmark against this metric. We were therefore also unable to compare e-Lung to the GAP score, an established metric of prognosis [29]. Additionally, although we performed a sensitivity analysis and found that the group of patients with serial imaging at ~12 months were representative of the whole cohort, it is possible that the retrospectively collected nature of the OSIC database is a source of bias. Patients will have undergone follow-up CTs for clinically indicated reasons, and this cohort therefore may be skewed towards patients more likely to have shown symptomatic progression. The analysis of serial CT changes in our study was performed across the entire cohort in a smaller number of patients, and the findings will require validation in other cohorts. It is also acknowledged that we did not include quantification of ground glass opacity in our analyses, and work is

ongoing to understand whether this further improves biomarker performance. Finally, we also acknowledge that further work is required to understand the impact of patient and related imaging acquisition factors such as prone positioning or expiration on the quantification of imaging biomarkers.

In summary, we have shown that e-Lung WRVS is a prognostic automated CT biomarker in patients with IPF and is associated with long-term survival when applied to both baseline and serial CT scans.

Provenance: Submitted article, peer reviewed.

Acknowledgements: We would like to acknowledge the support of the Open Source Imaging Consortium, their members, contributors and patient groups.

Ethics statement: The OSIC Data Repository has Central Institutional Review Board (IRB) and multiple institution IRB approvals.

Author contributions: The study was conceived by P.M. George, C. Rennison-Jones, G. Benvenuti, G.W.J. Harston, O. Joly and A. Devaraj. Data management was led by A. Sifostratoudaki and C. Fernandez. S. Gerry provided expert statistical support. R.E. Ledda, R.F. Abul Kadir and B. Johari provided radiology visual assessment. All authors were involved in drafting the manuscript and approved the submitted manuscript.

Conflict of interest: P.M. George reports support for the present study from Brainomix, grants from Boehringer Ingelheim, payment or honoraria for lectures, presentations, manuscript writing or educational events from Boehringer Ingelheim, Roche, Teva, Cipla, Brainomix, AstraZeneca and Daiichi-Sankyo, support for attending meetings from Boehringer Ingelheim and Roche, participation on a data safety monitoring board or advisory board with GSK, stock options with Brainomix, and the following financial (or non-financial) interests: is the Senior Medical Director of Brainomix. C. Rennison-Jones reports support for the present study from Brainomix, and stock (or stock options) with Brainomix. G. Benvenuti reports support for the present study from Brainomix, and stock (or stock options) with Brainomix. A. Sifostratoudaki reports support for the present study from Brainomix, and stock (or stock options) with Brainomix. F.A. Ottink reports support for the present study from Brainomix, and stock (or stock options) with Brainomix. W. Bou-Zeid reports support for the present study from Brainomix, and stock (or stock options) with Brainomix. R.E. Ledda reports consultancy fees from Brainomix. R.F. Abul Kadir reports consultancy fees from Brainomix. B. Johari reports consultancy fees from Brainomix. C. Fernandez reports support for the present study from Brainomix, and stock (or stock options) with Brainomix. G.W.J. Harston reports support for the present study from Brainomix, leadership role with Open Source Imaging Consortium (Board Member), and stock options with Brainomix. O. Joly reports support for the present study from Brainomix, and stock (or stock options) with Brainomix. S. Gerry reports consultancy fees from Brainomix. A. Devaraj reports consultancy fees from Boehringer Ingelheim, Roche, Vicore, Brainomix, Daewoong, Galapagos and Takeda, payment or honoraria for lectures, presentations, manuscript writing or educational events from Boehringer Ingelheim and AstraZeneca, and stock (or stock options) with Brainomix.

## References

1 Lederer DJ, Martinez FJ. Idiopathic pulmonary fibrosis. *N Engl J Med* 2018; 378: 1811–1823.
2 George PM, Patterson CM, Reed AK, *et al.* Lung transplantation for idiopathic pulmonary fibrosis. *Lancet Respir Med* 2019; 7: 271–282.
3 King TE, Bradford WZ, Castro-Bernardini S, *et al.* A phase 3 trial of pirfenidone in patients with idiopathic pulmonary fibrosis. *N Engl J Med* 2014; 370: 2083–2092.
4 Richeldi L, du Bois RM, Raghu G, *et al.* Efficacy and safety of nintedanib in idiopathic pulmonary fibrosis. *N Engl J Med* 2014; 370: 2071–2082.
5 de Andrade JA, Neely ML, Hellkamp AS, *et al.* Effect of antifibrotic therapy on survival in patients with idiopathic pulmonary fibrosis. *Clin Ther* 2023; 45: 306–315.
6 Nathan SD, Behr J, Cottin V, *et al.* Study design and rationale for the TETON phase 3, randomised, controlled clinical trials of inhaled treprostinil in the treatment of idiopathic pulmonary fibrosis. *BMJ Open Respir Res* 2022; 9: e001310.
7 Richeldi L, Azuma A, Cottin V, *et al.* Design of a phase III, double-blind, randomised, placebo-controlled trial of BI 1015550 in patients with idiopathic pulmonary fibrosis (FIBRONEER-IPF). *BMJ Open Respir Res* 2023; 10: e001563.
8 Flaherty KR, Brown KK, Wells AU, *et al.* Design of the PF-ILD trial: a double-blind, randomised, placebo-controlled phase III trial of nintedanib in patients with progressive fibrosing interstitial lung disease. *BMJ Open Respir Res* 2017; 4: e000212.

9 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.
10 Barnes H, Humphries SM, George PM, *et al.* Machine learning in radiology: the new frontier in interstitial lung diseases. *Lancet Digit Health* 2023; 5: e41–e50.
11 Humphries SM, Mackintosh JA, Jo HE, *et al.* Quantitative computed tomography predicts outcomes in idiopathic pulmonary fibrosis. *Respirology* 2022; 27: 1045–1053.
12 Walsh SLF, Calandriello L, Silva M, *et al.* Deep learning for classifying fibrotic lung disease on high-resolution computed tomography: a case-cohort study. *Lancet Respir Med* 2018; 6: 837–845.
13 Humphries SM, Yagihashi K, Huckleberry J, *et al.* Idiopathic pulmonary fibrosis: data-driven textural analysis of extent of fibrosis at baseline and 15-month follow-up. *Radiology* 2017; 285: 270–278.
14 Humphries SM, Swigris JJ, Brown KK, *et al.* Quantitative high-resolution computed tomography fibrosis score: performance characteristics in idiopathic pulmonary fibrosis. *Eur Respir J* 2018; 52: 1801384.
15 Jacob J, Bartholmai BJ, Rajagopalan S, *et al.* Predicting outcomes in idiopathic pulmonary fibrosis using automated computed tomographic analysis. *Am J Respir Crit Care Med* 2018; 198: 767–776.
16 Guiot J, Engelberts J, Ernst B, *et al.* AI-based model for the automatized quantification and prognosis of IPF patients. *Am J Respir Crit Care Med* 2023; 207: A65382023.
17 Rea G, Sverzellati N, Bocchino M, *et al.* Beyond visual interpretation: quantitative analysis and artificial intelligence in interstitial lung disease diagnosis "expanding horizons in radiology". *Diagnostics (Basel)* 2023; 13: 2333.
18 Rennison-Jones C, Gerry S, Gupta G, *et al.* The Brainomix automated e-ILD CT algorithm outperforms forced vital capacity in predicting outcomes for patients with idiopathic pulmonary fibrosis. *Eur Respir J* 2022; 60: 918.
19 George P, Rennison-Jones C, Benvenuti G, *et al.* In the serial assessment of patients with idiopathic pulmonary fibrosis, the automated e-ILD CT algorithm outperforms lung validation: a validation study. *Am J Respir Crit Care Med* 2023; 207: A6532.
20 Devaraj A, Ottink F, Rennison-Jones C, *et al.* e-Lung computed tomography biomarker stratifies patients at risk of idiopathic pulmonary fibrosis progression in a 52-week clinical trial. *Am J Respir Crit Care Med* 2024; 209: 1168–1169.
21 Devaraj A, Milanese G, Sverzellati N. Thoracic computed tomography in the progressive fibrotic phenotype. *Curr Opin Pulm Med* 2021; 27: 350–354.
22 Walsh SLF, Mackintosh JA, Calandriello L, *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.
23 Hothorn T, Lausen B. On the exact distribution of maximally selected rank statistics. *Comput Stat Data Anal* 2003; 43: 121–137.
24 Watt JA, Veroniki AA, Tricco AC, *et al.* Using a distribution-based approach and systematic review methods to derive minimum clinically important differences. *BMC Med Res Methodol* 2021; 21: 41.
25 Ash SY, Harmouche R, Vallejo DL, *et al.* Densitometric and local histogram based analysis of computed tomography images in patients with idiopathic pulmonary fibrosis. *Respir Res* 2017; 18: 45.
26 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.
27 Chen A, Karwoski RA, Gierada DS, *et al.* Quantitative CT analysis of diffuse lung disease. *Radiographics* 2020; 40: 28–43.
28 Tanizawa K, Handa T, Nagai S, *et al.* Clinical impact of high-attenuation and cystic areas on computed tomography in fibrotic idiopathic interstitial pneumonias. *BMC Pulm Med* 2015; 15: 74.
29 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.