Thorax

The relationship between exhaled volatile organic compounds and lung function change in idiopathic pulmonary fibrosis

Journal:	*Thorax*
Manuscript ID	thorax-2024-222321.R3
Article Type:	Short report
Date Submitted by the Author:	03-Mar-2025
Complete List of Authors:	Hayton, Conal; Manchester University NHS Foundation Trust; University of Manchester, Faculty of Biology, Medicine and Health Ahmed, Waqar; University of Manchester, Faculty of Biology, Medicine and Health Terrington, Dayle; University of East Anglia White, Iain; The University of Manchester, Laboratory for Environmental and Life Sciences; University of Nova Gorica, Division of Immunology, Immunity to infection & Respiratory Medicine, School of Biological Sciences Chaudhuri, Nazia; University of Ulster Faculty of Life and Health Sciences; Western Healthcare trust Leonard, Colm; Manchester University NHS Foundation Trust, NIHR-Manchester Biomedical Research Centre Wilson, Andrew; University of East Anglia, Norwich Medical School Fowler, Stephen; University of Manchester
Keywords:	Idiopathic pulmonary fibrosis

I, the Submitting Author has the right to grant and does grant on behalf of all authors of the Work (as defined in the below author licence), an exclusive licence and/or a non-exclusive licence for contributions from authors who are: i) UK Crown employees; ii) where BMJ has agreed a CC-BY licence shall apply, and/or iii) in accordance with the terms applicable for US Federal Government officers or employees acting as part of their official duties; on a worldwide, perpetual, irrevocable, royalty-free basis to BMJ Publishing Group Ltd ("BMJ") its licensees and where the relevant Journal is co-owned by BMJ to the co-owners of the Journal, to publish the Work in this journal and any other BMJ products and to exploit all rights, as set out in our licence.

The Submitting Author accepts and understands that any supply made under these terms is made by BMJ to the Submitting Author unless you are acting as an employee on behalf of your employer or a postgraduate student of an affiliated institution which is paying any applicable article publishing charge ("APC") for Open Access articles. Where the Submitting Author wishes to make the Work available on an Open Access basis (and intends to pay the relevant APC), the terms of reuse of such Open Access shall be governed by a Creative Commons licence – details of these licences and which Creative Commons licence will apply to this Work are set out in our licence referred to above.

Other than as permitted in any relevant BMJ Author's Self Archiving Policies, I confirm this Work has not been accepted for publication elsewhere, is not being considered for publication elsewhere and does not duplicate material already published. I confirm all authors consent to publication of this Work and authorise the granting of this licence.

Title: The relationship between exhaled volatile organic compounds and lung function change in idiopathic pulmonary fibrosis

Short title: VOCs and lung function in IPF

Authors:

Conal Hayton, MBChB, PhD,[1,2]

Waqar M. Ahmed, PhD,[1]

Dayle Terrington, MBBS,[3]

Iain R. White, PhD,[1,4]

Nazia Chaudhuri, MBChB, PhD,[5]

Colm Leonard, MBBCh, MD,[2]

Andrew M. Wilson, MBChB, MD,[3,6]

Stephen J. Fowler, MBChB, MD.[1,2]

Affiliation:

[1]Division of Immunology, Immunity to Infection and Respiratory Medicine; School of Biological Sciences; Faculty of Biology, Medicine and Health; University of Manchester, Manchester, United Kingdom

[2]NIHR-Manchester Biomedical Research Centre, Manchester University NHS Foundation Trust, Manchester, United Kingdom

[3]Norwich Medical School, University of East Anglia, Norwich, United Kingdom

[4]Laboratory for Environmental and Life Sciences, University of Nova Gorica, Nova Gorica, Slovenia

[5]School of Medicine, The University of Ulster, Magee Campus, Londonderry, United Kingdom

[6]Department of Respiratory Medicine, Norfolk and Norwich NHS Foundation Trust, Norwich, United Kingdom

Corresponding author:

Conal Hayton (conal.hayton@manchester.ac.uk)

Word Count:

Abstract: 111

Main body: 1150

Conflict of interest:

CH has received an honorarium from Boehringer Ingelheim for providing educational seminars. NC has received consultancy fees from Boehringer Ingelheim, Carrick Therapeutics, Roche, Redex, Novartis, Liminal Biosciences, Vicor Pharma, Bridge Biotherapeutics and Transcrip Ltd, speaker fees from Intermune, Roche, Boehringer Ingelheim and Astra Zeneca and educational support from Intermune, Chiesi, Boehringer Ingelheim, Roche and Ferrer. WMA, DT, IRW, CL, AMW and SJF have no relevant conflict of interests.

Funding information:

This study was funded by an investigator-initiated grant from Boehringer Ingelheim (Award number: 199.0311).

Acknowledgements: We would like the acknowledge the work of former members of the Manchester Breathomics Group, Dr Max Wilkinson and Kirti Vekaria, and the support of the Manchester Clinical Research Facility. This study has been accepted for presentation at the American Thoracic Society International Conference 2024.

Author Contributions: CH, WMA, DT, IRW, NC, CL, AMW and SJF were involved in project conceptualisation and method development. SJF and NC secured funding. CH and DT performed data collection. WMA and IRW performed sample analysis. CH and WMA performed data analysis. CH prepared the manuscript first draft. WMA, DT, IRW, NC, CL, AMW and SJF reviewed manuscript drafts. CH is the guarantor of the manuscript.

Abstract

Volatile organic compounds (VOCs) in exhaled breath have shown promise as biomarkers in idiopathic pulmonary fibrosis (IPF). We analysed breath from 57 people with IPF using thermal desorption-gas chromatography-mass spectrometry to identify VOCs related to lung function change over twelve months. A LASSO regression model selected 63 VOCs associated with relative change in FVC (eight with correlation coefficient (CC) ≥0.20 on Spearman's rank analysis), and 28 associated with relative change in D_{LCO} % predicted (12 with CC ≥0.20). Secondary analyses demonstrated correlation between VOCs and baseline lung function parameters and association with survival. This study suggests that there may be a volatile signature of prognosis in IPF that merits further validation.

INTRODUCTION

Idiopathic pulmonary fibrosis (IPF) is a progressive lung disease associated with morbidity and premature death. It is recognised that IPF displays heterogeneity in disease progression.[1] Currently few tools exist to aid prognostication and predict individual disease behaviour.

There is growing interest in the use of volatile organic compounds (VOCs) as biomarkers. VOCs are a diverse group of metabolites present in exhaled breath. It has been demonstrated using electronic nose technology that VOCs in breath can distinguish IPF from healthy controls and other interstitial lung diseases with a high level of accuracy.[2] However, this analysis method does not allow individual VOCs to be identified or links to underlying pathophysiology to be explored.

The aim of this study was to explore the relationship between individual VOCs present in exhaled breath and longitudinal changes in lung function in IPF.

METHODS

Study design and participants

Eligible participants were recruited to an observational cohort study (IPF VOC; ISRCTN1806574) at two UK centres. Demographic, clinical and lung function data, including forced vital capacity (FVC), FVC % predicted, diffusion capacity for carbon monoxide (D_{LCO}) and D_{LCO} % predicted, were collected at baseline and at three, six and twelve months.

Breath sampling

Exhaled breath was collected from participants in clinic using the ReCIVA device (Owlstone Medical, Cambridge, UK) onto two TenaxGR (Markes International, Bridgend, UK) stainless steel sorbent tubes according to a locally derived protocol.[3] The tubes were sealed and transferred to the laboratory where they underwent thermal desorption-gas chromatography-mass spectrometry (TD-GC-MS). Breath samples were collected at baseline and at each subsequent visit. Further details about participant eligibility, TD-GC-MS and data pre-processing are provided in the supplementary material.

Statistical analysis

Untargeted analysis was performed on baseline breath samples to identify VOCs associated with lung function change. This was assessed in two ways: relative change in FVC % predicted and relative change in D_{LCO} % predicted over twelve months. A LASSO regression model was used to select VOCs associated with either parameter. Relative change in FVC and DLCO % predicted were estimated using a linear mixed effects model. Correlation of relative concentration of individual VOCs selected was tested using Spearman's rank correlation with adjustment for false discovery using the Benjamini-Hochberg procedure. VOCs with a correlation coefficient ≥0.20 were reported. These VOCs were retained for secondary analyses including correlation with baseline FVC % predicted, DLCO % predicted, University San Diego Shortness of Breath Questionnaire (UCSD-SOBQ) score and Medical Research Council (MRC) dyspnoea score, association with survival and progression free-survival and change in VOC relative concentration in response to antifibrotic treatment. A Cox proportional hazards model was used to test associations between VOC relative concentration with survival and progression-free survival. Hazard ratios were adjusted for GAP stage and antifibrotic use within the model. Progression was defined as a ≥10% relative decline in FVC % predicted, or ≥15% relative decline in D_{LCO} % predicted at twelve months, or death, as previously described.[4] Further detail of statistical analysis can be found in the supplementary material.

RESULTS

Study cohort

Eighty-eight patients were recruited with 57 included in final analysis (Figure S1). Baseline demographics are shown in table 1 with further details provided in the supplementary material (Figures S2 and S3).

Table 1. Demographic and clinical data of cohort. Data presented as mean (±standard deviation) or number (percentage of total).

Demographics	n=57
Age (yrs)	75.1 (±6.5)
Sex male	46 (80.7%)
BMI (kg/m^2)	28.4 (±4.9)
Ex-smoker	43 (75.4%)
Pack years	25.5 (±16.3)
Years since stopping	26.7 (±16.7)
FVC (L)	2.77 (±3.15)
FVC % Predicted	79.7 (±18.8)
D_{LCO} (mmol/min/kPa)	3.71 (±1.21)
D_{LCO} % predicted	41.6 (±19.2)
GAP stage	
Stage 1	15 (26.3%)
Stage 2	28 (49.1%)
Stage 3	14 (24.6%)
MRC 1	3 (5.3%)
MRC 2	24 (42.1%)
MRC 3	18 (31.6%)
MRC 4	7 (12.3%)
MRC 5	5 (8.8%)
UCSD-SOBQ Total	40.5 (±27.2)
Supplementary oxygen use at baseline	2 (3.5%)
Antifibrotic use at baseline	
Pirfenidone	3 (5.3%
Nintedanib	1 (1.8%)
None	53 (92.9%)

BMI=body mass index, FVC=forced vital capacity, D_{LCO}=diffusion capacity of the lung for carbon monoxide. GAP=gender, age and physiology stage, MRC=medical research council dyspnoea score, UCSD-SOBQ=University of California San Diego shortness of breath questionnaire,

Untargeted VOC analysis

A total of 180 VOCs were identified across baseline samples. LASSO regression selected 63 VOCs that were associated with relative change in FVC % predicted and 28 with relative change in D_{LCO} % predicted at twelve months (Table S1). Table 2 lists VOCs with a correlation coefficient ≥0.2 for each lung function parameter.

Secondary analysis

Significant negative correlation was observed between baseline D_{LCO} % predicted and m-cymene (R=-0.34, 95% confidence interval -0.58 to -0.10, p=0.01), 2-chloro-p-xylene (R=-0.27, -0.51 to -0.03, p=0.03) and 4,6-dimethylundecane (R=-0.26, -0.50 to -0.02 p=0.04, Table S2). 4,6-Dimethylundecane correlated with baseline UCSD-SOBQ (R=0.31, 0.07 to 0.55, p=0.01, Table S3) and MRC dyspnoea score (R=0.36, 0.13 to 0.60, p=0.003, Table S4). No significant correlation was noted between VOCs and baseline FVC % predicted or change in symptom scores at twelve months (Tables S5-S9). Increasing concentration of 4,6-dimethylundecane was associated with reduced survival after adjusting for Gender Age and Physiology (GAP) stage and antifibrotic use (hazard ratio [HR] 4.14, 1.24-13.74, p=0.02), while increasing concentration of d-limonene was associated with improved survival (HR 0.25, 0.08 to 0.80, p=0.02) (Figure 1 and Table S10). No association was observed with progression-free survival (Table S11). Three VOCs (octanal, 3-methlyfuran and m-cymene) demonstrated a significant change in relative concentration post-treatment with antifibrotics (Table S12).

Table 2. Volatile organic compounds with a correlation coefficient ≥0.20 between relative concentration and relative change in FVC % predicted or DLCO % predicted at twelve months. *p<0.05

VOC	IUPAC name	CAS	Class	Correlation Coefficient	95% CI	P value
Relative change in FVC % predicted at 12 months						
4-Cyclopentene-1,3-dione	Cyclopent-4-ene-1,3-dione	930-60-9	Unsaturated cyclic dicarbonyl	0.36	(0.10 to 0.57)	0.02*
1-Chloropentane	1-Chloropentane	543-59-9	Haloalkane	0.33	(0.08 to 0.62)	0.03*
2-Chloro-p-xylene	2-Chloro-1,4-dimethylbenzene	95-72-7	Aromatic hydrocarbon	-0.29	(-0.58 to -0.04)	0.06
trans-3,3,5-Trimethylcyclohexanol	(1S,5R)-3,3,5-Trimethylcyclohexan-1-ol	767-54-4	Secondary alcohol	0.26	(-0.002 to 0.47)	0.10
o-Cymene	1-Methyl-2-propan-2-ylbenzene	527-84-4	Monoterpenoid	-0.25	(-0.52 to 0.02)	0.12
Mesitylene	1,3,5-Trimethylbenzene	108-67-8	Aromatic hydrocarbon	0.25	(-0.02 to 0.53)	0.12
Cyclopentadiene	Cyclopenta-1,3-diene	542-92-7	Alicyclic hydrocarbon	0.23	(-0.03 to 0.46)	0.15
2-Phenyl-2-propanol	2-Phenylpropan-2-ol	617-94-7	Benzyl alcohol	-0.22	(-0.42 to 0.04)	0.18
Relative change in D_{LCO} % predicted at 12 months						
2-Methyltetrahydrofuran	2-Methyloxolane	25265-68-3	Cyclic ester	-0.51	(-0.74 to -0.28)	<0.001*
m-Cymene	1-Methyl-3-propan-2-ylbenzene	535-77-3	Monoterpenoid	-0.40	(-0.67 to -0.16)	0.002*
3-Methylfuran	3-Methylfuran	930-27-8	Heteroaromatic compound	-0.32	(-0.59 to -0.07)	0.01*
1,1,3-Trimethylcyclohexane	1,1,3-Trimethylcyclohexane	3073-66-3	Cyclic alkane	0.32	(0.05 to 0.51)	0.02*
D-Limonene	(4R)-1-Methyl-4-prop-1-en-2-ylcyclohexene	5989-27-5	Monoterpene	-0.31	(-0.59 to -0.05)	0.02*
o-Cymene	1-Methyl-2-propan-2-ylbenzene	527-84-4	Monoterpenoid	-0.30	(-0.57 to -0.04)	0.02*
4,6-Dimethylundecane	4,6-Dimethylundecane	17312-82-2	Branched alkane	-0.30	(-0.56 to -0.04)	0.03*
1,1,1-Trichloroethane	1,1,1-Trichloroethane	71-55-6	Haloalkane	-0.29	(-0.58 to -0.04)	0.03*
2-Butoxyethanol	2-Butoxyethanol	111-76-2	Glycol ether derivative	-0.29	(-0.56 to -0.03)	0.03*
1-Butanol	Butan-1-ol	71-36-3	Primary alcohol	-0.27	(-0.52 to -0.01)	0.04*
Octanal	Octanal	124-13-0	Saturated fatty aldehyde	-0.27	(-0.52 to -0.01)	0.04*

1,4-Benzoquinone	Cyclohexa-2,5-diene-1,4-dione	106-51-4	Unsaturated cyclic dicarbonyl	-0.26	(-0.52 to 0.003)	0.04*

VOC=volatile organic compound, IUPAC= International Union of Pure and Applied Chemistry, CAS=Chemical Abstracts Service, CI=confidence interval, FVC=forced vital capacity, D_{LCO}=diffusion capacity of the lung for carbon monoxide.

DISCUSSION

We identified individual VOCs in exhaled breath which may be related to lung function change in IPF. Correlations between VOC concentration and lung function change were generally weak (r 0.2-0.5) although a moderate negative correlation (r > 0.5) was observed between 2-methyltetrahydrofuran and relative change in D_{LCO} % predicted at twelve months. m-Cymene, a monoterpenoid, also demonstrated a negative correlation with relative change in D_{LCO} % predicted as well as baseline D_{LCO} % predicted. In addition, treatment with nintedanib led to a reduction in m-cymene concentration on repeated sampling. An isomer, p-cymene, has been shown to discriminate IPF from healthy controls, and correlate with measures of disease severity.[5] We previously observed an increased concentration of monoterpenes in the headspace of lung cells stimulated with the profibrotic cytokine transforming growth factor (TGF)-β.[6] Many detectable VOCs appear to be exogenous and may indicate ambient air pollution exposure, which is associated with both the incidence and progression of IPF.[7, 8] We found increased 4-6-dimethylundecane, a suspected exogenous VOC, correlated with decline in D_{LCO} % predicted, increased symptom scores and was associated with reduced survival. Exposure to air pollution is known to alter the composition of VOCs on exhaled breath,[9] but the exact interplay between the exposome and volatilome remains uncertain.

This study was limited by a small sample size, impacted by a significant batch effect as a consequence of drift in analytical performance over the timeframe of the study. This limited the number of baseline and follow-up breath samples we could use. In addition, in the absence of echocardiography data, we cannot exclude the impact of emerging pulmonary hypertension, which is known to alter VOC composition,[10] on change in D_{LCO} rather than progressive fibrosis.

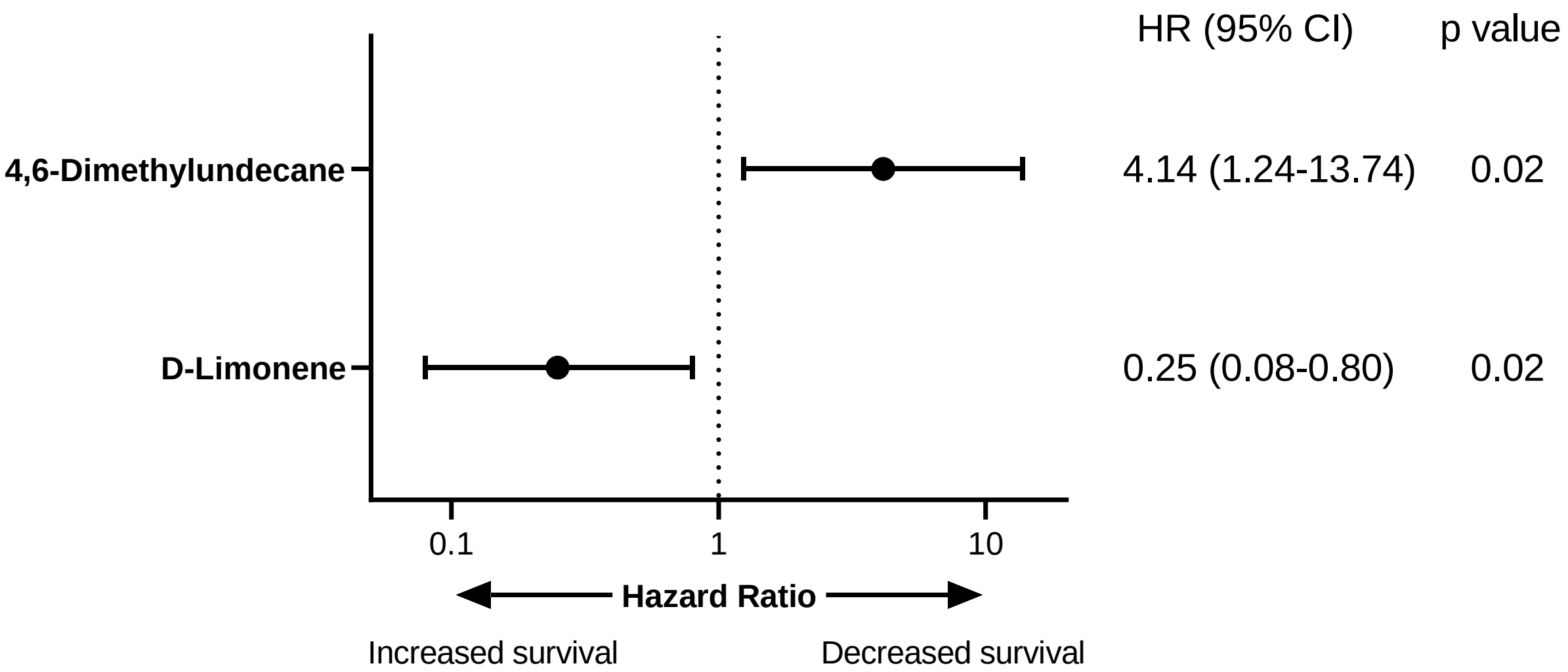
The results of this study support previous work that VOCs may be altered in lung fibrosis,[2, 5] although a targeted validation study is required to confirm this and further work needed to establish the pathophysiological link between identified VOCs and IPF. Breath analysis using GC-MS offers the opportunity to perform biomarker discovery, however the utility of this technique in clinical practice is uncertain given analysis is performed offline. Ultimately, online devices developed to target specific VOCs are likely to represent the best solution as clinically useful breath biomarkers in respiratory disease.

In conclusion, we identified VOCs in exhaled breath that may be related to measures of lung function decline in IPF. These results support further studies to investigate the possibility of a volatile signature for prognosis in IPF.

REFERENCES

1. Ley B, Collard HR, King TE, Jr. Clinical course and prediction of survival in idiopathic pulmonary fibrosis. Am J Respir Crit Care Med. 2011;183(4):431-40.
2. Moor CC, Oppenheimer JC, Nakshbandi G, et al. Exhaled breath analysis by use of eNose technology: a novel diagnostic tool for interstitial lung disease. Eur Respir J. 2021;57(1).
3. Peel A, Wang R, Ahmed W, et al. Changes in exhaled volatile organic compounds following indirect bronchial challenge in suspected asthma. Thorax. 2023;78(10):966-73.
4. Noble PW, Albera C, Bradford WZ, et al. Pirfenidone in patients with idiopathic pulmonary fibrosis (CAPACITY): two randomised trials. The Lancet. 2011;377(9779):1760-9.
5. Yamada Y, Yamada G, Otsuka M, et al. Volatile Organic Compounds in Exhaled Breath of Idiopathic Pulmonary Fibrosis for Discrimination from Healthy Subjects. Lung. 2017;195(2):247-54.
6. Hayton C, Ahmed W, Cunningham P, et al. Changes in lung epithelial cell volatile metabolite profile induced by pro-fibrotic stimulation with TGF-β1. J Breath Res. 2023;17(4).
7. Cui F, Sun Y, Xie J, et al. Air pollutants, genetic susceptibility and risk of incident idiopathic pulmonary fibrosis. Eur Respir J. 2022:2200777.
8. Goobie GC, Carlsten C, Johannson KA, et al. Association of Particulate Matter Exposure With Lung Function and Mortality Among Patients With Fibrotic Interstitial Lung Disease. JAMA Intern Med. 2022;182(12):1248-59.
9. Zhang L, Li X, Chen H, et al. Haze Air Pollution Health Impacts of Breath-Borne VOCs. Environ Sci Technol. 2022;56(12):8541-51.
10. Nakhleh MK, Haick H, Humbert M, et al. Volatolomics of breath as an emerging frontier in pulmonary arterial hypertension. Eur Respir J. 2017;49(2).

Figure 1. Forest plot showing significant associations between relative volatile organic compounds (VOCs) concentration and survival. Adjusted hazard ratios reported.


HR (95% CI)
p value
4,6-Dimethylundecane
4.14 (1.24-13.74)
0.02
D-Limonene
0.25 (0.08-0.80)
0.02
0.1
1
10
Hazard Ratio
Increased survival
Decreased survival

Supplementary material

ADDITIONAL METHODS

Study design and participants.

Participants were recruited to this observational cohort study (IPF VOC; ISRCTN1806574) at two tertiary ILD centres in the UK between July 2018 and June 2019: Manchester University NHS Foundation Trust and Norfolk and Norwich NHS Foundation Trust. Patients were eligible if they met the following criteria:

Inclusion criteria

1) Age ≥ 18
2) Multi-disciplinary team (MDT)-diagnosis of IPF, as per international consensus guidelines.[1]

Exclusion criteria

1) Significant respiratory co-morbidity (i.e. where the major respiratory diagnosis is not IPF) as determined by investigator.
2) Residual volume ≥ 90 % predicted on full lung function testing.
3) Current smoker (within 4 weeks of enrolment).
4) Received treatment for acute lower respiratory tract infection within last 4 weeks.
5) Unwilling to participate in the study.
6) Current participation in a double-blind placebo controlled pharmaceutical trial.

Demographic and clinical data were collected including age, gender, smoking status, comorbidities, and medication. Lung function data including FVC, FVC % predicted, DLCO and DLCO % predicted were collected as part of routine clinical care. Patients completed two breathlessness questionnaires: the Medical Research Council (MRC) breathlessness scale[2] and the University of California San Diego Shortness of Breath Questionnaire (UCSD-SOBQ).[3]

Patients were followed up for twelve months and provided exhaled breath samples at baseline, three, six and twelve months. Clinical data, breathlessness questionnaires and lung function data were also collected at these time points.

Breath sampling

Exhaled breath was collected in the outpatient clinic using the ReCIVA sampling device (Owlstone Medical, Cambridge, UK) according to a locally derived protocol.[4] The ReCIVA device was connected to a CASPER filtered air supply (Owlstone Medical) to reduce exogenous VOC collection. The device was connected to a reusable silicon mask (Owlstone Medical) which was baked at 180°c for 24 hours prior to use to reduce background siloxane VOCs. A disposable bacterial/viral filter (Philips Respironics, Eindhoven, Netherlands) was placed between the mask and the ReCIVA device to prevent contamination. Tenax GR (Markes International, Rhondda Cynon Taff, UK) stainless steel sorbent tubes containing a composite of a polymer-based resin and graphite were placed inside the silicon mask to trap VOCs. Two tubes were used to take duplicate samples. The device was programmed to sample end-tidal breath only and 500ml of exhaled breath was sampled at a rate of 200ml/minute. Prior to clinical sampling, background samples of the ambient air without the silicon mask were taken, along with samples from a glass head dummy to isolate VOCs produced by the mask. Once sample collection was completed, log files were analysed to confirm a sample collection volume. After collection, the sorbent tubes were sealed at both ends with caps and stored in a secure refrigerator at 4°c for a maximum of seven days.

Thermal desorption-gas chromatography-mass spectrometry (TD-GC-MS)

Sorbent tubes were dry purged with nitrogen to remove excess water at a flow rate of 50 ml min−1 for 8 min. Before primary desorption each sample tube was automatically injected with 100µl of a gaseous calibrated internal standard (1 ppmv, 4-bromofluorobenzene in nitrogen, Thames Restek, UK, High Wycombe, UK). Desorption was performed using a TD-100 system (Markes International, Bridgend, UK) at 280 ◦C for 5 min and subsequently transferred onto a cryofocusing trap held at 0 ◦C. The trap was then flash heated to 280 ◦C for 3 min and transferred to the GC column into a capillary column (DB-5 ms Ultra Inert, length 30m× internal diameter 0.25 mm × 25 µm film thickness, Agilent Technologies, Cheadle, UK) housed within a GC oven (7890, Agilent Technologies, Cheadle, UK). The following temperature ramp was used: 40 ◦C (no hold), ramp by 6 ◦C min−1 ramp to 170 ◦C, and 15 ◦C min−1 ramp to 190 ◦C (constant pressure at 69 kPa, total GC cycle time of 23 min) using a helium carrier flow of 1 mL min−1. VOCs were then ionised (EI 70 eV) in a mass spectrometer (7010, Agilent

Technologies, Cheadle, UK) in full scan mode with a scan range of m/z 40–500 and scan speed of 4 Hz.

Data processing

A reference library and peak integration method were developed by automatic peak scanning through individual GC-MS data files (Agilent Masshunter Quantitative analysis). Quantifier ion retention time windows were optimised using an external standard mixture as a reference to adjust for retention time shift across all samples. Peak areas were generated using the Agile2 integrator. Peak areas from background samples were subtracted in matched breath samples to remove any exogenous compounds from the sampling system and inhaled air. Features were rejected and excluded from further analysis if they fit the following criteria: 1) a mass spectral library (NIST version 14) match score of < 80; 2) an ICC value =< 0.6 between replicate breath samples; 3) they were within a retention index delta of 5% compared to literature values (NIST database of non-polar temperature-ramp Kovats or Alkane retention indices), or the retention index (RI) could not be estimated, and 4) they had missing data in more than 80% of samples.

Statistical analysis

Untargeted analysis was performed on baseline exhaled breath samples. Two lung function parameters were used: relative change in FVC % predicted and relative change in DLCO % predicted at twelve months. A least absolute shrinkage and selection operator (LASSO) regression model was used to identify VOCs associated with either parameter. The regularisation parameter L1 in the LASSO model was set to 1 specifying a pure LASSO model and model stability and tuning the λ parameter (best R^2 value) was assessed using 1000 bootstrap samples. A linear mixed effects model was used to estimate relative change in FVC % predicted and DLCO % predicted at twelve months. This model allowed the inclusion of all available lung function data and account for missing data and the influence of GAP stage and antifibrotic use on lung function change. Antifibrotic use was a dichotomised variable based on the use of antifibrotics for 50% of the study period. The model included GAP stage and antifibrotic use as fixed effects and patient subject number as a random effect, with random slopes and intercepts. Residuals were tested for normality using the Shapiro-Wilk test. Patients who were lost to follow-up after the first visit were excluded from the analysis.

Correlation of individual VOCs identified through the LASSO model with either lung function parameter was tested using Spearman's rank correlation coefficient with correction for false discovery using the Benjamini-Hochberg procedure. Individual VOCs with a correlation coefficient ≥0.2 were reported and retained for secondary analyses.

Secondary analyses were performed on individual VOCs retained from the primary analysis. Association between relative VOC concentration and baseline FVC % predicted, DLCO % predicted, UCSD-SOBQ score and MRC dyspnoea scales were analysed using Spearman's rank correlation coefficient. Change in UCSD-SOBQ score and MRC dyspnoea scale at twelve months were analysed as both continuous and dichotomised variables (symptom progression vs no symptom progression). A five-point increase in UCSD-SOBQ score was used to define symptom progression as previously described.[5] A one point increase in MRC dyspnoea scale was used to define symptom progression. Spearman's rank correlation coefficient was used to test the association with continuous change in UCSD-SOBQ score and MRC scale and independent t test was used to compare differences in mean VOC relative concentration between dichotomised change in UCSD-SOB and MRC scores.

Associations were tested between individual VOCs and survival and progression-free survival. Progression was defined as a ≥10% relative decline in FVC % predicted, or ≥15% relative decline in D_{LCO} % predicted at twelve months, or death.[6] A Cox proportional hazards model was used to test associations between VOC relative concentration with survival and progression-free survival. Hazard ratios were adjusted for GAP stage and antifibrotic use within the model.

The impact of antifibrotic therapy on VOC concentration was tested by examining their relative change in concentration over time. A paired t-test was used to compare the concentration of selected VOCs between baseline pre-treatment samples and three-month post-treatment samples. Breath samples of untreated patients were also tested for comparison.

ADDITONAL RESULTS

Study cohort

Eighty-eight patients were recruited to the study. Ten patients were excluded from analysis: one provided baseline samples but subsequently had a change in diagnosis at a second MDT

and was therefore withdrawn; two were unable to provide adequate baseline breath samples; and eight were lost to follow-up after providing baseline samples. Samples from seventy-seven patients were included in initial analysis. However, during the recruitment period a chromatography column was replaced as part of routine GC-MS maintenance which resulted in a pronounced batch effect that could be observed on data inspection. To minimise this effect, results from 20 patients who underwent baseline breath sampling after the GC-MS maintenance, were excluded. The remaining 57 patients were included in subsequent analysis (Figure S1).

All patients were within three years of diagnosis. Figure S2 shows common comorbidities affecting at least 10% of the study cohort. Other respiratory co-morbidities included any reported condition (including COPD). Individuals with additional respiratory conditions were included in the study if this was deemed to be of secondary clinical importance (as per the selection criteria). Figure S3 shows common medication use reported in at least 10% of the study population. In addition to the four patients taking antifibrotic medication at baseline, an additional 30 patients were started on antifibrotics during follow-up (fifteen pirfenidone, fifteen nintedanib). Amongst individuals treated with antifibrotics, the median percentage time spent on antifibrotic therapy during the study was 80.3 (range 4.1-100) %. Mean change in FVC % predicted at 12 months was -9.1% (± standard deviation 13.4) and D_{LCO} % predicted was -19.6% (±11.9). Thirteen patients died during follow-up and 43 patients suffered disease progression with a median time to disease progression of 255.7 (95% confidence interval 199.6-260.4) days. Amongst those who suffered disease progression, two patients suffered a relative decline in FVC % predicted ≥10%, twelve suffered a relative decline in D_{LCO} % predicted ≥15% and sixteen suffered both.

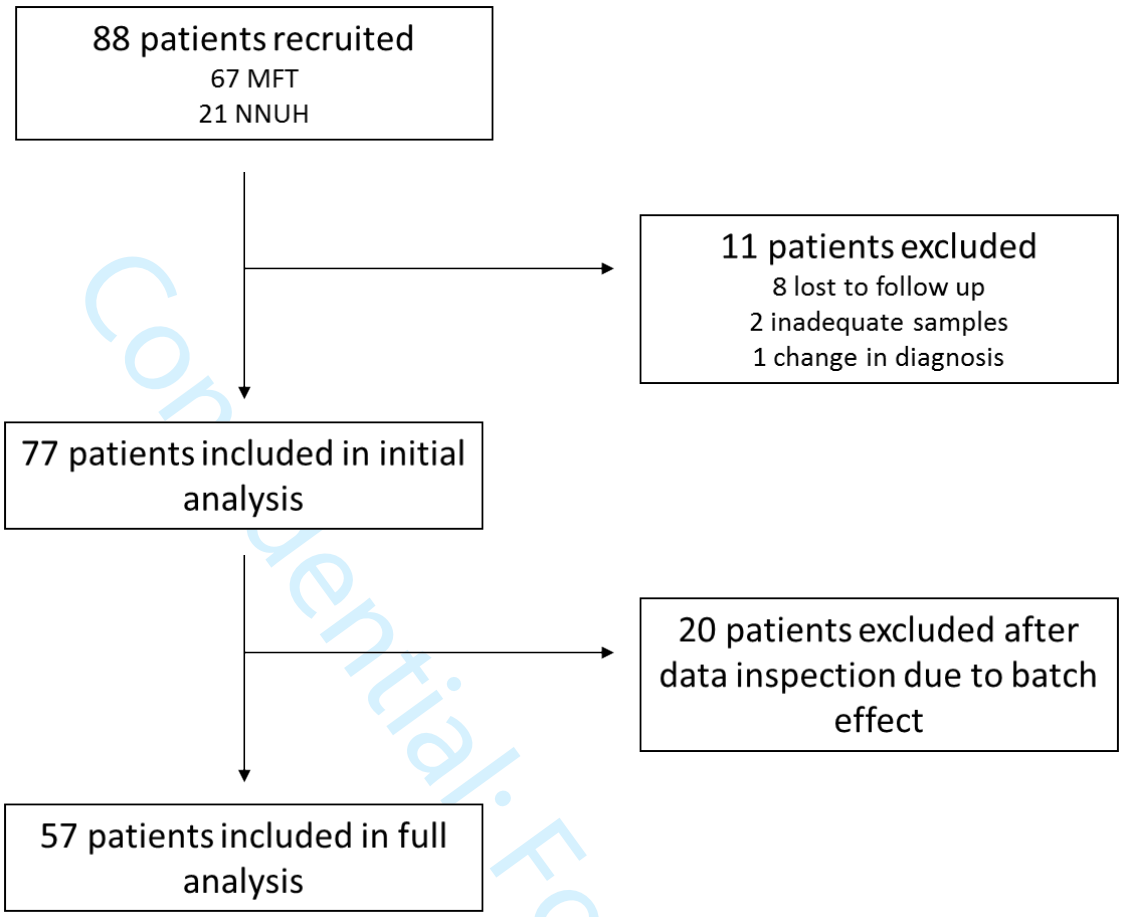


Figure S1. Study flow.

MFT=Manchester University NHS Foundation Trust; NNUH=Norfolk and Norwich NHS Foundation Trust.

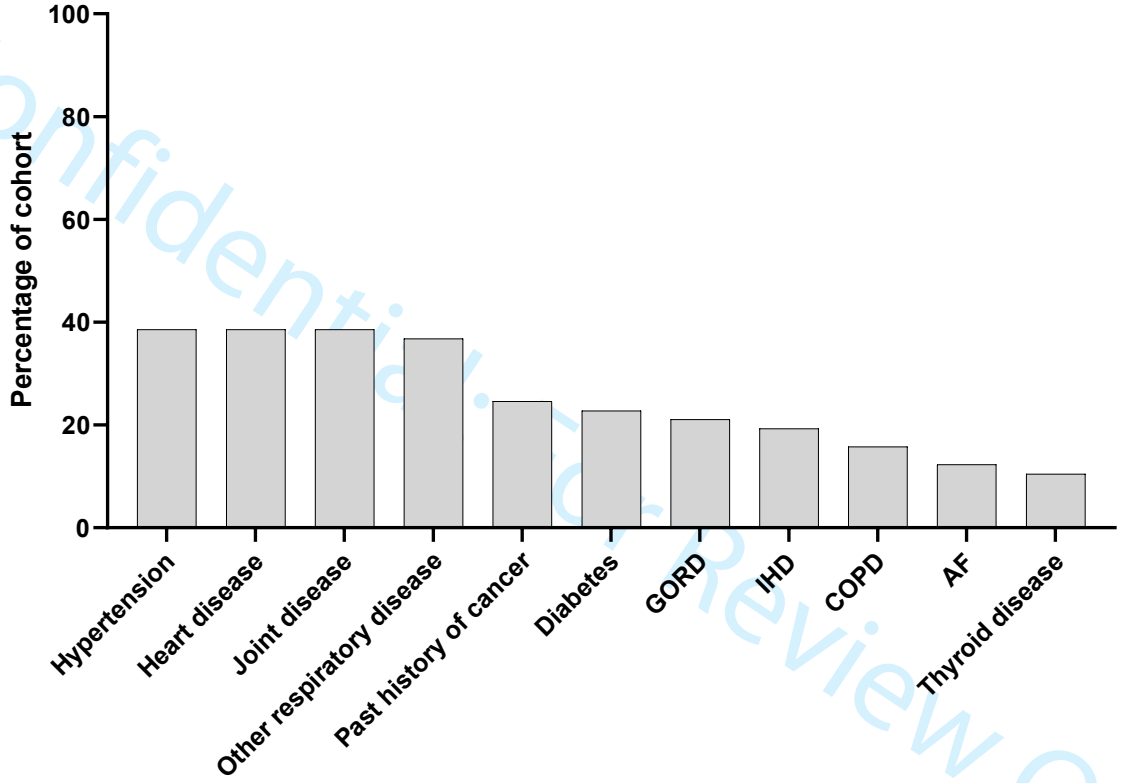


Figure S2. Comorbidities in at least 10% of the total cohort.

GORD=gastro-oesophageal reflux disease, IHD=ischaemic heart disease, COPD=chronic obstructive pulmonary disease, AF=atrial fibrillation.

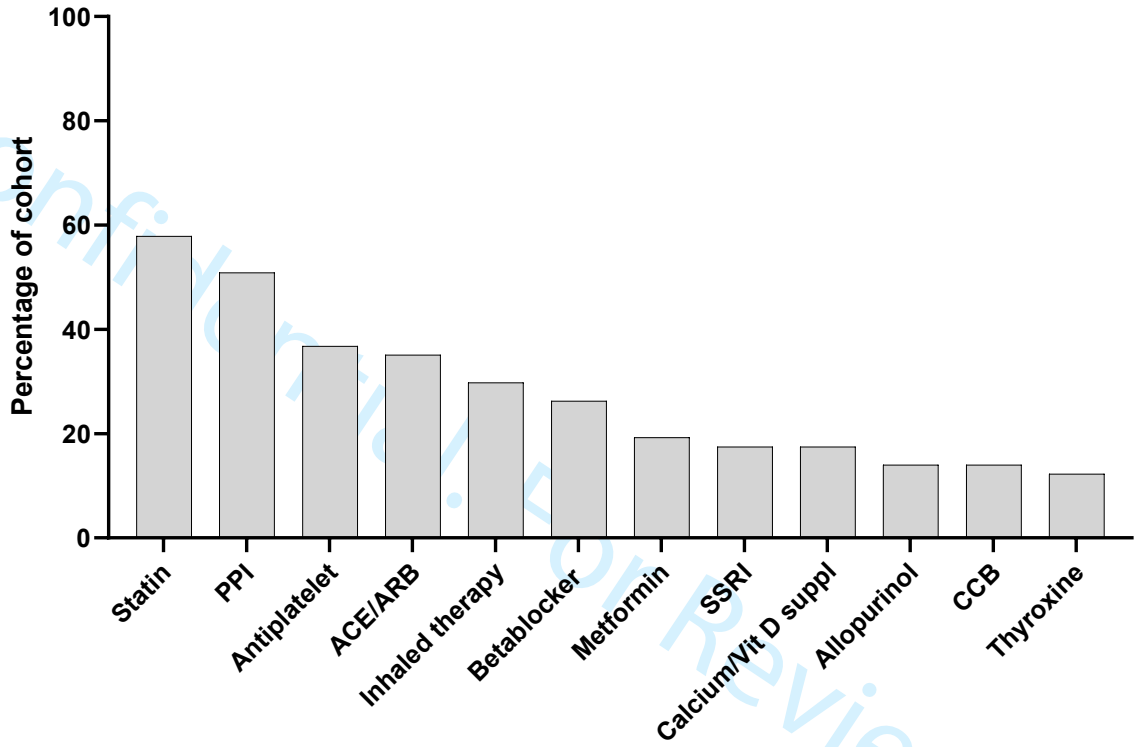


Figure S3. Medications used by at least 10% of the total cohort.

PPI=proton pump inhibitor, ACE=angiotensin converting enzyme inhibitor, ARB=angiotensin receptor blocker, CCB=calcium channel blocker, SSRI=selective serotonin reuptake inhibitor, Vit D suppl=vitamin D supplement.

Table S1. Volatile organic compounds associated with relative change in FVC or D_{LCO} % predicted at 12 months using least absolute shrinkage and selection operator (LASSO) regression.

VOC	IUPAC name	CAS number	Compound formula
Relative change in FVC % predicted at 12 months			
1,1,1-Trichloroethane	1,1,1-Trichloroethane	71-55-6	C2H3Cl3
1,1,3-Trimethylcyclohexane	1,1,3-Trimethylcyclohexane	3073-66-3	C9H18
1,2,3-Trimethylcyclohexane	1,2,3-Trimethylcyclohexane	1678-97-3	C9H18
1,2,4-Trimethylbenzene	1,2,4-Trimethylbenzene	95-63-6	C9H12
1,2,4-Trimethylcyclohexane	1,2,4-Trimethylcyclohexane	2234-75-5	C9H18
1,3,3-Trimethyl-2-oxabicyclo[2.2.2]oct-5-ene	1,3,3-Trimethyl-2-oxabicyclo[2.2.2]oct-5-ene	92760-25-3	C10H16O
1-Butanol	Butan-1-ol	71-36-3	C4H10O
1-Chloro-2-methylpropane	1-Chloro-2-methylpropane	513-36-0	C4H9Cl
1-Chloropentane	1-Chloropentane	543-59-9	C5H11Cl
1-Ethyl-2-methylbenzene	1-Ethyl-2-methylbenzene	611-14-3	C9H12
1-Ethyl-3-methylcyclohexane	1-Ethyl-3-methylcyclohexane	3728-55-0	C9H18
1-Propene, 1-(methylthio)-, (Z)-	(Z)-1-Methylsulfanylprop-1-ene	52195-40-1	C4H8S
2,2,4,4,6,8,8-Heptamethylnonane	2,2,4,4,6,8,8-Heptamethylnonane	4390-04-09	C16H34
2,3,4-Trimethylhexane	2,3,4-Trimethylhexane	921-47-1	C9H20
2,3-Dimethyloctane	2,3-Dimethyloctane	7146-60-3	C10H22
2,3-Epoxybutane	2,3-Dimethyloxirane	3266-23-7	C4H8O
2,3-Heptanedione	Heptane-2,3-dione	96-04-8	C7H12O2
2,4-Dimethyl-3-pentanone	2,4-Dimethylpentan-3-one	565-80-0	C7H14O
2,4-Dimethylheptane	2,4-Dimethylheptane	2213-23-2	C9H20
2-Chloro-p-xylene	2-Chloro-1,4-dimethylbenzene	95-72-7	C8H9Cl
2-Ethylhexan-1-ol	2-Ethylhexan-1-ol	104-76-7	C8H18O
2-Ethylhexan-1-ol	2-Ethylhexan-1-ol	104-76-7	C8H18O
2-Hydroxycyclopent-2-en-1-one	2-Hydroxycyclopent-2-en-1-one	10493-98-8	C5H6O2
2-Methyltetrahydrofuran	2-Methyloxolane	25265-68-3	C5H10O
2-Phenyl-2-propanol	2-Phenylpropan-2-ol	617-94-7	C9H12O
3,4,5-Trimethyl-2-cyclopenten-1-one	3,4,5-Trimethylcyclopent-2-en-1-one	55683-21-1	C8H12O
3-Carene	3,7,7-Trimethylbicyclo[4.1.0]hept-3-ene	13466-78-9	C10H16
3-Ethylpentane	3-Ethylpentane	617-78-7	C7H16
3-Methylhexane	3-Methylhexane	589-34-4	C7H16
4-Cyclopentene-1,3-dione	Cyclopent-4-ene-1,3-dione	930-60-9	C5H4O2
4-Octen-3-one	(E)-Oct-4-en-3-one	69065-31-2	C8H14O
Acetoin	3-Hydroxybutan-2-one	513-86-0	C4H8O2
Benzene	Benzene	71-43-2	C6H6
Benzofuran	1-Benzofuran	271-89-6	C8H6O
Benzoic acid	Benzoic acid	65-85-0	C7H6O2
Benzyl alcohol	Phenylmethanol	100-51-6	C7H8O
Benzyl bromide	Bromomethylbenzene	100-39-0	C7H7Br

Chloroform	Chloroform	67-66-3	CHCl3
Cyclohexane	Cyclohexane	110-82-7	C6H12
Cyclohexanone	Cyclohexanone	108-94-1	C6H10O
Cyclooctatetraene	Cyclooctatetraene	629-20-9	C8H8
Cyclopentadiene	Cyclopenta-1,3-diene	542-92-7	C5H6
Dioxane	1,4-Dioxane	28552-22-9	C4H8O2
D-Limonene	(4R)-1-Methyl-4-prop-1-en-2-ylcyclohexene	5989-27-5	C10H16
Ethyl Acetate	Ethyl Acetate	141-78-6	C4H8O2
Ethylcyclopentane	Ethylcyclopentane	1640-89-7	C7H14
Eucalyptol	1,3,3-Trimethyl-2-oxabicyclo[2.2.2]octane	470-82-6	C10H18O
Isocyanatotrimethylsilane	Isocyanato(trimethyl)silane	1118-02-1	C4H9NOSi
Mesitylene	1,3,5-Trimethylbenzene	108-67-8	C9H12
Methenamine	1,3,5,7-Tetrazatricyclo[3.3.1.13,7]decane	100-97-0	C6H12N4
Methyl acetate	Methyl acetate	79-20-9	C3H6O2
Methyl Ethyl Ketone	Butan-2-one	78-93-3	C4H8O
Methyl propionate	Methyl propanoate	554-12-1	C4H8O2
m-Xylene	1,3-Xylene	108-38-3	C8H10
Myrcene	7-Methyl-3-methylideneocta-1,6-diene	123-35-3	C10H16
Nonanal	Nonanal	124-19-6	C9H18O
Nonane	Nonane	111-84-2	C9H20
o-Cymene	1-Methyl-2-propan-2-ylbenzene	527-84-4	C10H14
Trans-3,3,5-trimethylcyclohexanol	(1S,5R)-3,3,5-Trimethylcyclohexan-1-ol	767-54-4	C9H18O
Trichloroethylene	1,1,2-Trichloroethene	79-01-6	C2HCl3
Tridecane	Tridecane	629-50-5	C13H28
Valeraldehyde	Pentanal	110-62-3	C5H10O
Relative change in D_{LCO} % predicted at 12 months			
1,1,1-Trichloroethane	1,1,1-Trichloroethane	71-55-6	C2H3Cl3
1,1,3-Trimethylcyclohexane	1,1,3-Trimethylcyclohexane	3073-66-3	C9H18
1,3,3-Trimethyl-2-oxabicyclo[2.2.2]oct-5-ene	1,3,3-Trimethyl-2-oxabicyclo[2.2.2]oct-5-ene	92760-25-3	C10H16O
1-Butanol	Butan-1-ol	71-36-3	C4H10O
1-Chloro-2-methylpropane	1-Chloro-2-methylpropane	513-36-0	C4H9Cl
1-Propene, 1-(methylthio)-, (Z)-	(Z)-1-Methylsulfanylprop-1-ene	52195-40-1	C4H8S
2-Butoxyethanol	2-Butoxyethanol	111-76-2	C6H14O2
2-Chloro-1,3-dimethylbenzene	2-Chloro-1,3-dimethylbenzene	6781-98-2	C8H9Cl
2-Methyltetrahydrofuran	2-Methyloxolane	25265-68-3	C5H10O
3,4,5-Trimethyl-2-cyclopenten-1-one	3,4,5-Trimethylcyclopent-2-en-1-one	55683-21-1	C8H12O
3-Carene	3,7,7-Trimethylbicyclo[4.1.0]hept-3-ene	13466-78-9	C10H16
3-Methylfuran	3-Methylfuran	930-27-8	C5H6O
4,6-Dimethylundecane	4,6-Dimethylundecane	17312-82-2	C13H28
4-Cyclopentene-1,3-dione	Cyclopent-4-ene-1,3-dione	930-60-9	C5H4O2

4-Octen-3-one	(E)-Oct-4-en-3-one	69065-31-2	C8H14O
Cyclohexanone	Cyclohexanone	108-94-1	C6H10O
Ethylcyclopentane	Ethylcyclopentane	1640-89-7	C7H14
Geraniol	(2E)-3,7-Dimethylocta-2,6-dien-1-o	106-24-1	C10H18O
m-Cymene	1-Methyl-3-propan-2-ylbenzene	535-77-3	C10H14
Menthol	5-Methyl-2-propan-2-ylcyclohexan-1-ol	1490-04-6	C10H20O
Methenamine	1,3,5,7-Tetrazatricyclo[3.3.1.13,7]decane	100-97-0	C6H12N4
Methyl 2-methylbutyrate	Methyl 2-methylbutanoate	868-57-5	C6H12O2
Methyl isobutyl ketone	4-Methylpentan-2-one	108-10-1	C6H12O
Myrcene	7-Methyl-3-methylideneocta-1,6-diene	123-35-3	C10H16
Octanal	Octanal	124-13-0	C8H16O
Phenylethyl Alcohol	2-Phenylethanol	60-12-8	C8H10O
1,4-Benzoquinone	Cyclohexa-2,5-diene-1,4-dione	106-51-4	C6H4O2
Tetradecane	Tetradecane	629-59-4	C14H30

Table S2. Correlation between volatile organic compound relative concentration and baseline DLCO % predicted.

VOC	Correlation coefficient (95% CI)	P-value
m-Cymene	-0.34 (-0.58 to -0.10)	0.01*
2-Chloro-p-xylene	-0.27 (-0.51 to -0.03)	0.03*
4,6-Dimethylundecane	-0.26 (-0.50 to -0.02)	0.04*
4-Cyclopentene-1,3-dione	0.19 (-0.06 to 0.44)	0.13
2-Methyltetrahydrofuran	0.14 (-0.11 to 0.39)	0.28
2-Phenyl-2-propanol	-0.13 (-0.38 to 0.12)	0.31
1,4-Benzoquinone	-0.13 (-0.38 to 0.12)	0.31
1-Butanol	-0.11 (-0.36 to 0.14)	0.37
2-Butoxyethanol	-0.11 (-0.37 to 0.13)	0.37
Octanal	-0.11 (-0.36 to 0.14)	0.38
1,1,1-Trichloroethane	0.11 (-0.14 to 0.36)	0.39
o-Cymene	-0.10 (-0.35 to 0.15)	0.44
Mesitylene	0.10 (-0.15 to 0.35)	0.44
Cyclopentadiene	0.09 (-0.17 to 0.34)	0.48
1,1,3-Trimethylcyclohexane	0.08 (-0.17 to 0.34)	0.50
D-Limonene	-0.08 (-0.33 to 0.17)	0.52
1-Chloropentane	0.04 (-0.20 to 0.30)	0.72
trans-3,3,5-Trimethylcyclohexanol	0.04 (-0.20 to 0.30)	0.73
3-Methylfuran	-0.04 (-0.29 to 0.22)	0.76

Table S3. Correlation between volatile organic compound relative concentration and baseline UCSD-SOBQ score.

VOC	Correlation Coefficient (95 % CI)	P-value
4,6-Dimethylundecane	0.31 (0.07 to 0.55)	0.01*
2-Chloro-p-xylene	0.27 (0.03 to 0.52)	0.03*
2-Phenyl-2-propanol	0.22 (-0.02 to 0.47)	0.08
4-Cyclopentene-1,3-dione	-0.15 (-0.40 to 0.10)	0.24
2-Methyltetrahydrofuran	-0.11 (-0.36 to 0.15)	0.39
1,4-Benzoquinone	0.10 (-0.15 to 0.36)	0.42
1,1,1-Trichloroethane	-0.08 (-0.33 to 0.18)	0.54
D-Limonene	-0.07 (-0.33 to 0.18)	0.57
Octanal	0.07 (-0.27 to 0.24)	0.61
m-Cymene	0.06 (-0.20 to 0.31)	0.66
2-Butoxyethanol	0.04 (-0.21 to 0.30)	0.75
1-Chloropentane	0.04 (-0.22 to 0.29)	0.76
Cyclopentadiene	-0.03 (-0.29 to 0.22)	0.80
o-Cymene	-0.03 (-0.28 to 0.22)	0.81
1,1,3-Trimethylcyclohexane	-0.03 (-0.28 to 0.23)	0.83
trans-3,3,5-Trimethylcyclohexanol	0.02 (-0.24 to 0.28)	0.88
Mesitylene	0.01 (-0.23 to 0.25)	0.93
3-Methylfuran	0.01 (-0.23 to 0.25)	0.94
1-Butanol	0.0003 (-0.26 to 0.26)	1.00

Table S4. Correlation between volatile organic compound relative concentration and baseline MRC scale.

VOC	Correlation Coefficient (95 % CI)	P-value
4,6-Dimethylundecane	0.36 (0.13 to 0.60)	0.003*
1,4-Benzoquinone	0.26 (0.02 to 0.50)	0.04*
2-Chloro-p-xylene	0.22 (-0.03 to 0.46)	0.08
4-Cyclopentene-1,3-dione	-0.19 (-0.43 to 0.06)	0.14
2-Phenyl-2-propanol	0.18 (-0.07 to 0.43)	0.15
Mesitylene	-0.17 (-0.41 to 0.08)	0.19
Cyclopentadiene	-0.13 (-0.38 to 0.12)	0.30
1,1,1-Trichloroethane	-0.12 (-0.37 to 0.13)	0.33
m-Cymene	0.11 (-0.14 to 0.36)	0.40
1,1,3-Trimethylcyclohexane	-0.09 (-0.34 to 0.16)	0.47
Octanal	0.09 (-0.17 to 0.36)	0.50
trans-3,3,5-Trimethylcyclohexanol	0.08 (-0.17 to 0.33)	0.54
1-Butanol	0.06 (-0.19 to 0.31)	0.63
3-Methylfuran	0.06 (-0.19 to 0.31)	0.64
2-Methyltetrahydrofuran	-0.05 (-0.30 to 0.20)	0.70
1-Chloropentane	0.05 (-0.20 to 0.30)	0.71
2-Butoxyethanol	0.04 (-0.21 to 0.29)	0.77
D-Limonene	-0.03 (-0.29 to 0.22)	0.79
o-Cymene	-0.01 (-0.26 to 0.24)	0.95

Table S5. Correlation between volatile organic compound relative concentration and baseline FVC % predicted.

VOC	Correlation Coefficient (95% CI)	P-value
trans-3,3,5-Trimethylcyclohexanol	-0.22 (-0.47 to 0.03)	0.08
2-Chloro-p-xylene	-0.19 (-0.37 to 0.13)	0.14
1,1,3-Trimethylcyclohexane	0.17 (-0.08 to 0.42)	0.17
2-Phenyl-2-propanol	-0.14 (-0.39 to 0.11)	0.26
Cyclopentadiene	0.12 (-0.13 to 0.37)	0.35
3-Methylfuran	-0.11 (-0.36 to 0.14)	0.38
Mesitylene	-0.11 (-0.36 to 0.14)	0.39
Octanal	-0.08 (-0.33 to 0.18)	0.51
4,6-Dimethylundecane	-0.08 (-0.33 to 0.18)	0.54
1-Chloropentane	-0.08 (-0.33 to 0.18)	0.54
4-Cyclopentene-1,3-dione	0.08 (-0.18 to 0.33)	0.55
o-Cymene	0.08 (-0.18 to 0.33)	0.55
2-Methyltetrahydrofuran	-0.06 (-0.31 to 0.20)	0.65
m-Cymene	-0.05 (-0.31 to 0.20)	0.67
1,1,1-Trichloroethane	0.04 (-0.21 to 0.30)	0.73
1-Butanol	0.03 (-0.23 to 0.23)	0.84
D-Limonene	0.02 (-0.23 to 0.27)	0.87
2-Butoxyethanol	-0.02 (-0.27 to 0.23)	0.87
1,4-Benzoquinone	0.005 (-0.25 to 0.26)	0.97

Table S6. Correlation between volatile organic compound relative concentration change in UCSD-SOBQ score at twelve months (n=25).

VOC	Correlation Coefficient (95 % CI)	P-value
Mesitylene	-0.40 (-0.76 to 0.001)	0.05
2-Phenyl-2-propanol	-0.26 (-0.58 to 0.13)	0.20
1,1,1-Trichloroethane	-0.24 (-0.63 to 0.18)	0.26
2-Methyltetrahydrofuran	-0.24 (-0.49 to 0.14)	0.26
1-Chloropentane	-0.23 (-0.76 to 0.23)	0.28
1,4-Benzoquinone	0.19 (-0.23 to 0.58)	0.37
1,1,3-Trimethylcyclohexane	-0.17 (-0.52 to 0.22)	0.42
D-Limonene	-0.13 (-0.49 to 0.27)	0.55
2-Butoxyethanol	0.09 (-0.39 to 0.60)	0.66
Octanal	0.08 (-0.35 to 0.51)	0.70
m-Cymene	0.08 (-0.38 to 0.55)	0.71
3-Methylfuran	0.05 (-0.40 to 0.50)	0.82
4,6-Dimethylundecane	0.05 (-0.37 to 0.46)	0.83
trans-3,3,5-Trimethylcyclohexanol	-0.04 (-0.44 to 0.36)	0.84
4-Cyclopentene-1,3-dione	-0.04 (-0.36 to 0.30)	0.85
2-Chloro-p-xylene	-0.02 (-0.49 to 0.45)	0.93
1-Butanol	0.02 (-0.35 to 0.38)	0.94
Cyclopentadiene	-0.01 (-0.37 to 0.35)	0.96
o-Cymene	0.0003 (-0.43 to 0.43)	1.00

Table S7. Difference in volatile organic compound relative concentration change between categorical change in UCSD-SOBQ score at twelve months (n=25).

VOC	Mean Difference (95 % CI) Progression vs No progression	P-value
Mesitylene	-0.49 (-1.01 to 0.04)	0.07
m-Cymene	0.33 (-0.09 to 0.74)	0.12
1,4-Benzoquinone	0.16 (-0.23 to 0.42)	0.17
Octanal	0.20 (-0.30 to 0.69)	0.42
1-Butanol	0.38 (-0.12 to 0.88)	0.44
2-Phenyl-2-propanol	-0.16 (-0.60 to 0.28)	0.46
1-Chloropentane	-0.23 (-0.90 to 0.43)	0.48
2-Chloro-p-xylene	-0.16 (-0.65 to 0.32)	0.49
2-Butoxyethanol	0.20 (-0.44 to 0.84)	0.53
trans-3,3,5-Trimethylcyclohexanol	-0.15 (-0.63 to 0.34)	0.54
Cyclopentadiene	0.05 (-0.49 to 0.60)	0.62
3-Methylfuran	0.18 (-0.60 to 0.93)	0.63
o-Cymene	0.53 (-0.19 to 1.25)	0.71
1,1,3-Trimethylcyclohexane	-0.09 (-0.67 to 0.48)	0.74
1,1,1-Trichloroethane	0.10 (-0.54 to 0.74)	0.75
4-Cyclopentene-1,3-dione	0.05 (-0.28 to 0.37)	0.77
2-Methyltetrahydrofuran	0.04 (-0.38 to 0.46)	0.84
4,6-Dimethylundecane	0.01 (-0.48 to 0.50)	0.97
D-Limonene	-0.01 (-0.49 to 0.48)	0.98

Table S8. Correlation between volatile organic compound relative concentration change in MRC scale at twelve months (n=30).

VOC	Correlation Coefficient (95 % CI)	P-value
2-Methyltetrahydrofuran	-0.30 (-0.55 to 0.06)	0.11
2-Phenyl-2-propanol	-0.24 (-0.56 to 0.13)	0.20
trans-3,3,5-Trimethylcyclohexanol	-0.23 (-0.54 to 0.13)	0.22
2-Butoxyethanol	0.20 (-0.20 to 0.65)	0.28
m-Cymene	0.17 (-0.23 to 0.62)	0.36
2-Chloro-p-xylene	-0.17 (-0.56 to 0.22)	0.37
1,4-Benzoquinone	0.13 (-0.23 to 0.47)	0.49
1-Chloropentane	-0.13 (-0.58 to 0.29)	0.50
D-Limonene	-0.12 (-0.46 to 0.25)	0.54
Cyclopentadiene	-0.12 (-0.40 to 0.21)	0.54
Octanal	-0.11 (-0.47 to 0.26)	0.56
1,1,3-Trimethylcyclohexane	-0.11 (-0.43 to 0.24)	0.57
4-Cyclopentene-1,3-dione	-0.09 (-0.36 to 0.22)	0.63
1-Butanol	-0.07 (-0.41 to 0.28)	0.70
3-Methylfuran	0.07 (-0.39 to 0.56)	0.72
4,6-Dimethylundecane	-0.06 (-0.45 to 0.34)	0.76
Mesitylene	-0.05 (-0.46 to 0.36)	0.80
o-Cymene	0.04 (-0.35 to 0.44)	0.82
1,1,1-Trichloroethane	0.01 (-0.38 to 0.40)	0.97

Table S9. Difference in volatile organic compound relative concentration change between categorical change in MRC scale at twelve months (n=30).

VOC	Mean Difference (95 % CI) Progression vs No progression	P-value
2-Methyltetrahydrofuran	-0.32 (-0.72 to 0.08)	0.11
2-Phenyl-2-propanol	-0.26 (-0.67 to 0.15)	0.20
trans-3,3,5-Trimethylcyclohexanol	-0.25 (-0.66 to 0.16)	0.22
2-Butoxyethanol	0.27 (-0.26 to 0.85)	0.28
m-Cymene	0.18 (-0.22 to 0.58)	0.36
2-Chloro-p-xylene	-0.18 (-0.59 to 0.22)	0.37
1,4-Benzoquinone	0.09 (-0.18 to 0.37)	0.49
1,1,1-Trichloroethane	-0.19 (-0.76 to 0.38)	0.50
1-Chloropentane	-0.19 (-0.76 to 0.38)	0.50
D-Limonene	-0.14 (-0.58 to 0.32	0.54
Cyclopentadiene	-0.14 (-0.61 to 0.33)	0.54
Octanal	-0.12 (-0.55 to 0.30)	0.56
1,1,3-Trimethylcyclohexane	-0.14 (-0.65 to 0.37)	0.57
4-Cyclopentene-1,3-dione	-0.07 (-0.36 to 0.22)	0.63
1-Butanol	-0.09 (-0.60 to 0.41)	0.70
3-Methylfuran	0.14 (-0.66 to 0.95)	0.72
4,6-Dimethylundecane	-0.07 (-0.54 to 0.40)	0.76
Mesitylene	-0.07 (-0.63 to 0.49)	0.80
o-Cymene	0.08 (-0.62 to 0.77)	0.82

Table S10. Association between volatile organic compound relative concentration survival at 12 months. Values adjusted for GAP score and antifibrotic use.

VOC	Raw Hazard Ratio (95% CI)	Raw P-value	Adjusted Hazard Ratio (95% CI)	Adjusted P-value
4,6-Dimethylundecane	6.14 (1.91 to 19.71)	0.002	4.14 (1.24 to 13.74)	0.02*
D-Limonene	0.45 (0.18 to 1.16)	0.10	0.25 (0.08 to 0.80)	0.02*
1-Butanol	2.08 (1.00 to 4.29)	0.05	1.95 (0.97 to 3.91)	0.06
Octanal	3.20 (1.05 to 9.73)	0.04	3.27 (0.91 to 11.80)	0.07
2-Methyltetrahydrofuran	1.71 (0.67 to 4.38)	0.27	2.22 (0.88 to 5.56)	0.09
1,4-Benzoquinone	2.37 (0.66 to 8.53)	0.19	2.97 (0.60 to 14.78)	0.17
o-Cymene	0.77 (0.41 to 1.47)	0.43	0.68 (0.35 to 1.32)	0.25
1,1,3-Trimethylcyclohexane	1.15 (0.55 to 2.39)	0.72	1.53 (0.64 to 3.67)	0.34
2-Butoxyethanol	1.48 (0.63 to 3.45)	0.37	1.37 (0.63 to 2.94)	0.43
m-Cymene	0.95 (0.29 to 3.07)	0.93	0.49 (0.13 to 1.86)	0.49
2-Chloro-p-xylene	1.77 (0.73 to 4.30)	0.21	1.36 (0.42 to 4.47)	0.61
Cyclopentadiene	1.13 (0.49 to 2.60)	0.77	1.16 (0.54 to 2.50)	0.71
Mesitylene	0.74 (0.33 to 1.63)	0.45	0.85 (0.34 to 2.11)	0.72
1,1,1-Trichloroethane	0.89 (0.43 to 1.85)	0.76	1.16 (0.49 to 2.74)	0.72
2-Phenyl-2-propanol	1.02 (0.40 to 2.64)	0.95	0.78 (0.29 to 2.06)	0.78
4-Cyclopentene-1,3-dione	0.78 (0.26 to 2.37)	0.26	1.08 (0.41 to 2.85)	0.87
1-Chloropentane	0.88 (0.38 to 2.05)	0.77	0.95 (0.39 to 2.35)	0.91
3-Methylfuran	1.15 (0.61 to 2.17)	0.66	1.03 (0.55 to 1.93)	0.93
trans-3,3,5-Trimethylcyclohexanol	0.87 (0.34 to 2.21)	0.77	0.99 (0.44 to 2.20)	0.97

Table S11. Association between volatile organic compound relative concentration progression-free survival at 12 months. Values adjusted for GAP score and antifibrotic use.

VOC	Raw Hazard Ratio	Raw p-value	Adjusted Hazard Ratio	Adjusted p-value
1-Chloropentane	0.52 (0.30 to 0.90)	0.02	0.53 (0.28 to 1.03)	0.06
1,1,3-Trimethylcyclohexane	0.55 (0.32 to 0.93)	0.03	0.55 (0.29 to 1.04)	0.07
Cyclopentadiene	0.65 (0.37 to 1.16)	0.14	0.55 (0.29 to 1.05)	0.07
Mesitylene	0.69 (0.43 to 1.12)	0.13	0.67 (0.39 to 1.13)	0.13
trans-3,3,5-Trimethylcyclohexanol	0.47 (0.23 to 0.97)	0.04	0.56 (0.26 to 1.23)	0.15
Octanal	0.87 (0.46 to 1.65)	0.67	0.60 (0.29 to 1.24)	0.17
D-Limonene	0.93 (0.52 to 1.67)	0.81	0.67 (0.33 to 1.34)	0.25
2-Methyltetrahydrofuran	0.97 (0.55 to 1.73)	0.92	1.39 (0.76 to 2.54)	0.28
m-Cymene	1.28 (0.63 to 2.57)	0.49	0.68 (0.31 to 1.47)	0.33
2-Phenyl-2-propanol	1.02 (0.49 to 2.10)	0.96	0.71 (0.35 to 1.45)	0.34
4-Cyclopentene-1,3-dione	0.41 (0.17 to 0.97)	0.04	0.68 (0.26 to 1.79)	0.43
1-Butanol	1.33 (0.81 to 2.19)	0.26	1.24 (0.73 to 2.09)	0.43
2-Chloro-p-xylene	1.26 (0.71 to 2.23)	0.44	0.74 (0.33 to 1.67)	0.47
2-Butoxyethanol	1.10 (0.65 to 1.86)	0.73	1.20 (0.68 to 2.12)	0.54
3-Methylfuran	0.99 (0.65 to 1.49)	0.95	1.10 (0.73 to 1.65)	0.66
o-Cymene	1.21 (0.80 to 1.83)	0.36	0.91 (0.57 to 1.45)	0.69
1,1,1-Trichloroethane	0.84 (0.55 to 1.30)	0.44	0.93 (0.57 to 1.51)	0.77
4,6-Dimethylundecane	1.39 (0.74 to 2.63)	0.31	1.02 (0.56 to 1.82)	0.96
1,4-Benzoquinone	1.05 (0.43 to 2.57)	0.92	1.00 (0.34 to 2.92)	1.00

Table S12. Difference in volatile organic compound relative concentration between baseline and three-month breath samples after initiation of antifibrotic treatment. The antifibrotic group includes individuals who received either antifibrotic medication, with separate analysis for pirfenidone and nintedanib A control group of individuals who did not receive either antifibrotic is included.

VOC	Antifibrotics (n=14)		Nintedanib (n=7)		Pirfenidone (n=7)		No antifibrotics (n=6)	
	Mean difference (95 % CI) Baseline vs 3-months	p-value	Mean difference (95% CI) Baseline vs 3 months	p-value	Mean difference (95% CI) Baseline vs 3-months	p-value	Mean difference (95% CI) Baseline vs 3 months	p-value
Octanal	0.42 (0.07 to 0.76)	0.02*	0.60 (-0.08 to 1.29)	0.07	0.22 (-0.32 to 0.76)	0.35	-0.18 (-0.90 to 0.55)	0.55
3-Methylfuran	1.01 (0.13 to 1.89)	0.03*	0.22 (-0.63 to 1.08)	0.55	1.79 (0.58 to 3.01)	0.01*	-0.38 (-1.39 to 0.64)	0.39
o-Cymene	0.37 (-0.18 to 0.93)	0.17	0.53 (-0.45 to 1.51)	0.24	0.22 (-0.85 to 1.29)	0.63	0.76 (-0.23 to 1.74)	0.11
4-Cyclopentene-1,3-dione	0.26 (-0.13 to 0.64)	0.17	0.27 (-0.43 to 0.98)	0.38	0.24 (-0.09 to 0.57)	0.12	0.27 (-0.23 to 0.77)	0.22
Mesitylene	0.25 (-0.13 to 0.62)	0.18	(0.26 (-0.37 to 0.88)	0.35	0.24 (-0.49 to 0.97)	0.46	-0.09 (-0.59 to 0.42)	0.68
1,4-Benzoquinone	0.14 (-0.14 to 0.42)	0.29	-0.14 (-0.47 to 0.19)	0.33	0.42 (-0.21 to 1.05)	0.15	-0.16 (-0.51 to 0.18)	0.27
2-Butoxyethanol	0.28 (-0.31 to 0.86)	0.33	-0.004 (-0.91 to 0.90)	0.99	0.55 (-0.79 to 1.90)	0.35	0.16 (-0.65 to 0.96)	0.64
1,1,1-Trichloroethane	0.25 (-0.31 to 0.80)	0.35	0.43 (-0.59 to 1.45)	0.34	0.06 (-0.69 to 0.81)	0.85	0.53 (-0.65 to 1.72)	0.30
Cyclopentadiene	-0.22 (-0.73 to 0.29)	0.36	-0.15 (-1.16 to 0.85)	0.72	-0.29 (-0.91 to 0.33)	0.29	-0.32 (-0.96 to 0.32)	0.25
4,6-Dimethylundecane	0.26 (-0.34 to 0.86)	0.36	0.47 (-0.44 to 1.38)	0.25	0.05 (-0.81 to 0.91)	0.89	-0.41 (-1.05 to 0.24)	0.16
2-Phenyl-2-propanol	-0.24 (-0.83 to 0.36)	0.41	-0.51 (-1.41 to 0.39)	0.22	0.04 (-0.75 to 0.83)	0.91	-0.47 (-1.21 ro 0.27)	0.16

1-Chloropentane	-0.11 (-0.42 to 0.19)	0.44	0.27 (-0.02 to 0.57)	0.07	-0.50 (-1.21 to 0.21)	0.14	0.24 (-0.82 to 1.29)	0.59
2-Chloro-p-xylene	0.12 (-0.23 to 0.47)	0.48	0.07 (-0.54 to 0.69)	0.78	0.17 (-0.52 to 0.85)	0.58	-0.01 (-0.47 to 0.45)	0.95
trans-3,3,5-Trimethylcyclohexanol	-0.19 (-0.79 to 0.41)	0.51	-0.18 (-1.29 to 0.93)	0.70	-0.19 (-0.74 to 0.36)	0.43	-0.46 (-1.12 to 0.21)	0.14
2-Methyltetrahydrofuran	0.16 (-0.45 to 0.77)	0.58	0.06 (-0.93 to 1.06)	0.88	0.26 (-0.81 to 1.32)	0.58	-0.34 (-1.31 to 0.64)	0.42
D-Limonene	-0.13 (-0.67 to 0.40)	0.60	-0.52 (-1.37 to 0.33)	0.19	0.25 (-0.58 to 1.08)	0.49	0.34 (-0.25 to 0.93)	0.20
1-Butanol	-0.06 (-0.59 to 0.48)	0.83	0.18 (-0.68 to 1.04)	0.63	-0.29 (-1.58 to 1.00)	0.60	-0.65 (-1.69 to 0.39)	0.17
1,1,3-Trimethylcyclohexane	-0.05 (-0.73 to 0.62)	0.86	0.05 (-1.34 to 1.45)	0.93	-0.16 (-0.62 to 0.29)	0.42	-0.18 (-0.80 to 0.43)	0.48
m-Cymene	0.01 (-0.34 to 0.36)	0.96	-0.39 (-0.73 to -0.05)	0.03*	0.41 (-0.39 to 1.21)	0.26	0.27 (-0.33 to 0.87)	0.31

ADDITONAL REFERENCES

1. Raghu G, Remy-Jardin M, Myers JL, et al. Diagnosis of Idiopathic Pulmonary Fibrosis. An Official ATS/ERS/JRS/ALAT Clinical Practice Guideline. Am J Respir Crit Care Med. 2018;198(5):e44-e68.
2. Fletcher CM. The clinical diagnosis of pulmonary emphysema; an experimental study. Proc R Soc Med. 1952;45(9):577-84.
3. Eakin EG, Resnikoff PM, Prewitt LM, et al. Validation of a new dyspnea measure: the UCSD Shortness of Breath Questionnaire. University of California, San Diego. Chest. 1998;113(3):619-24.
4. Peel A, Wang R, Ahmed W, et al. Changes in exhaled volatile organic compounds following indirect bronchial challenge in suspected asthma. Thorax. 2023;78(10):966-73.
5. Chen T, Tsai APY, Hur SA, et al. Validation and minimum important difference of the UCSD Shortness of Breath Questionnaire in fibrotic interstitial lung disease. Respir Res. 2021;22(1):202.
6. Noble PW, Albera C, Bradford WZ, et al. Pirfenidone in patients with idiopathic pulmonary fibrosis (CAPACITY): two randomised trials. The Lancet. 2011;377(9779):1760-9.