Article:
Ng, K.L., Saunders, L., Collier, G. et al. (2026) Using hyperpolarised xenon MRI to explore breathlessness in long COVID. ERJ Open Research. ISSN: 2312-0541

https://doi.org/10.1183/23120541.00040-2026

ERJ open research
THE BEST IN OPEN ACCESS BASIC, TRANSLATIONAL & CLINICAL RESPIRATORY RESEARCH

Early View

Original Research Article

Using Hyperpolarised Xenon MRI to Explore Breathlessness in Long COVID

Kher Lik Ng, Laura Saunders, Guilhem Collier, James Grist, Nina Kainth, Victoria Harris, Laurie Smith, Scarlett Strickland, Lotta Gustafsson, Paul Hughes, Ionitiana Evans, Leizl Alvarez, Avianna Laws, Siby Thomas, Alberto Biancardi, Jennifer Rodgers, Samantha Jones, Demi Jakymelen, Martin Brook, Thomas Newman, Megan Plowright, Lily Dryhurst-Pearce, Amany Elbehairy, Kenneth Jacob, Anthony McIntyre, David Capener, Jonathan Bray, Marianne Durrant, Kylie Yeung, Huw Walters, Violet Matthews, Lisa Watson, Gavin Vuddamalay, Gabriele Abu-Eid, Viktorie Madhusudhan Stisova, Mark Cox, Maja Lachut, Nina Mulvey, William Hickes, Alexander Horsley, Helen Davies, Alfred A. Roger Thompson, Fergus Gleeson, Jim Wild, Emily Fraser

Please cite this article as: Ng KL, Saunders L, Collier G, *et al*. Using Hyperpolarised Xenon MRI to Explore Breathlessness in Long COVID. *ERJ Open Res* 2026; in press (https://doi.org/10.1183/23120541.00040-2026).

Using Hyperpolarised Xenon MRI to Explore Breathlessness in Long COVID

Results from the HypErpolarised Xenon Magnetic Resonance PuLmonary ImAging In PatieNts with Long-COVID (EXPLAIN study)

*Kher Lik Ng[1,2], *Laura Saunders[3], Guilhem Collier[3], James Grist[4], Nina Kainth[5], Victoria Harris[6], Laurie Smith[3], Scarlett Strickland[7], Lotta Gustafsson[7], Paul Hughes[3], Ionitiana Evans[5], Leizl Alvarez[5], Avianna Laws[5], Siby Thomas[7], Alberto Biancardi[3], Jennifer Rodgers[3], Samantha Jones[8], Demi Jakymelen[3], Martin Brook[3], Thomas Newman`, Megan Plowright[1], Lily Dryhurst-Pearce[1], Amany Elbehairy[9], Kenneth Jacob[5], Anthony McIntyre[10], David Capener[3], Jonathan Bray[3], Marianne Durrant[10], Kylie Yeung[2], Huw Walters[10], Violet Matthews[5], Lisa Watson[3], Gavin Vuddamalay[5], Gabriele Abu-Eid[5], Viktorie Madhusudhan Stisova[11], Mark Cox[11], Maja Lachut[5,11], Nina Mulvey[5], William Hickes[5], Alexander Horsley[9], Helen Davies[8], ‡Alfred A Roger Thompson[1], ‡Fergus Gleeson[2,5,10], ‡Jim Wild[3], ‡†Emily Fraser[12]

*joint first authors; ‡Senior authors
† **Corresponding author:**
Emily Fraser
Department of Respiratory Medicine, Oxford University Hospitals NHS Foundation Trust, Oxford, UK
Email: emily.fraser@ouh.nhs.uk
Phone: +447872996244,

1 Division of Clinical Medicine, University of Sheffield, Sheffield, UK
2 Department of Oncology, University of Oxford, Oxford, UK
3 POLARIS, University of Sheffield, Sheffield, UK
4 Department of Physiology, Anatomy and Genetics, University of Oxford, Oxford, UK
5 Oxford Radiology Research Unit, Oxford University Hospitals NHS Foundation Trust, Oxford, UK
6 Primary Care Clinical Trials Unit, Nuffield Department of Primary Care Health Sciences, University of Oxford, Oxford, UK
7 Sheffield Teaching Hospitals NHS Foundation Trust, Sheffield, UK
8 Department of Respiratory Medicine, Cardiff and Vale University Health Board, Cardiff, UK
9 Department of Respiratory Medicine, Manchester University NHS Foundation Trust, Manchester, UK
10 Department of Radiology, Oxford University Hospitals NHS Foundation Trust, Oxford, UK
11 Department of Radiopharmacy, Oxford University Hospitals NHS Foundation Trust, Oxford, UK
12 Department of Respiratory Medicine, Oxford University Hospitals NHS Foundation Trust, Oxford, UK

Take Home Message

Breathlessness is not caused by hidden damage affecting the lungs in the majority of Long Covid sufferers. Although subtle reductions in gas transfer were observed in 13%, overall, this study highlights the complexity of dyspnoea and importance of rehabilitative approaches.

Abstract

Background: Breathlessness is a common symptom in Long COVID (LC). Using hyperpolarised xenon MRI (^{129}Xe-MRI), we assessed whether this symptom could be attributed to abnormalities in the alveolar-capillary membrane not detected by standard investigations. We focused on never-hospitalised individuals without an identified cause for breathlessness.

Methods: In this prospective, multicentre study, we compared ^{129}Xe-MRI, lung function, exercise capacity, and symptom questionnaires in LC patients with breathlessness (BLC) to those without breathlessness (NBLC) and healthy controls. Primary outcome was whether BLC demonstrated measurable impairments in gas exchange focusing on dissolved-phase ^{129}Xe-MRI metrics: red-blood cell to membrane ratio (RBC:M) and red-blood cell to gas ratio (RBC:Gas). We also explored associations between symptoms and physiological measures.

Results: Of 269 participants recruited, 196 were included in the analysis (109 BLC, 43 NBLC, 44 controls), with age and sex well matched across groups. BLC had a significantly longer interval from infection to imaging (median 632 days; $p<0{\cdot}001$). No significant differences in global or regional RBC:M or RBC:Gas were observed across groups. BLC showed lower FEV_1, FVC, TLCO, and KCO z-scores compared to controls, though >90% of values remained within normal range. A subset of BLC participants with low TLCO (14/109) showed reduced ^{129}Xe-MRI metrics and higher breathlessness scores.

Interpretation: Most non-hospitalised LC participants exhibited no detectable pulmonary abnormalities, including those with breathlessness. However, approximately 13% of breathless individuals demonstrated minor reductions in TLCO and ^{129}Xe-MRI gas exchange suggesting a potential pulmonary contribution for symptoms in this subgroup.

Introduction

Post-COVID-19 condition, or Long COVID, is estimated to affect two million people living in the UK.[1] It is characterised by a wide range of symptoms although breathlessness, fatigue, and brain fog are most frequently described (*Office of National Statistics, April 2024*). The mechanisms driving symptoms are complex and, to date, remain poorly understood, but carry a large healthcare burden.[2] Understanding the causes of breathlessness post COVID-19 infection is an important therapeutic priority.

Long COVID (LC) is broadly defined by the persistence of symptoms after 12 weeks from the acute infection.[3] The severity of initial symptoms does not predict the development of LC and many patients managed in the community experienced a relatively mild acute COVID-19 illness. Hospital admission is often determined by the requirement for oxygen therapy and patients typically have evidence of pneumonic changes on lung imaging during the acute phase. These changes can take several months to resolve, and in some cases, result in a fibrotic 'footprint' of the acute infection causing persistent lung function impairment and exercise intolerance.[4,5] In contrast, patients who managed their infection in the community rarely have imaging undertaken during their acute illness and chest x-ray and computed tomography (CT) to investigate ongoing breathlessness after the infection are normal in most cases.[6] Despite this, up to 50% of LC sufferers report breathlessness, and in some it can be the main symptom limiting the return to functional activity.[1]

Hyperpolarised xenon MRI (^{129}Xe-MRI) is a novel imaging modality that is safe, reproducible and can provide quantitative metrics.[7,8] It does not involve radiation exposure. The technique involves inhaling xenon-129 (^{129}Xe), an inert gas that has been hyperpolarised to transiently increase signal inside the MRI scanner. The gas mirrors the flow of oxygen through the bronchial tree, across the alveolar-capillary membrane, into the pulmonary vasculature. Delays in the diffusion of gas through the membrane due to abnormalities at this interface, or abnormalities in the pulmonary vasculature, can be quantified with dissolved-phase imaging.[9] After inhalation, ^{129}Xe diffuses from the alveoli (gas phase) into the alveolar-capillary membrane. A small percentage of this then diffuses into the blood stream where it transiently binds with haemoglobin in red blood cells (RBCs). ^{129}Xe exhibits different resonance frequencies in the gas phase, tissue membrane/plasma phase and on binding within RBCs and so the ratio of signal delivered to the RBCs compared to the signal measured in the gas or membrane phase can be assessed (Figure 1). These RBC:Membrane (RBC:M), Membrane:Gas (M:Gas) and RBC:Gas ratios provide functional measurements of gas exchange at a regional level, providing detailed spatial information about membrane and vascular integrity not necessarily identifiable using global summary measures such as carbon monoxide gas transfer.[10] RBC:M provides a measure of gas transfer efficiency across the membrane and into the capillaries, M:Gas provides a measure of xenon uptake in the membrane and RBC:Gas assesses gas transfer from alveolus to the RBC and therefore provides a broader measurement of the gas exchange properties of the lung.

Previous work has shown impaired ^{129}Xe-MRI gas transfer in both hospitalised and non-hospitalised LC patients which fails to normalise over time.[11–14] We reported pilot ^{129}Xe-

MRI findings of non-hospitalised LC patients infected at the start of the 2020 COVID-19 pandemic with persistent breathlessness and reported significant functional abnormalities.[15]

In this study, we again focused on breathlessness, this time in a larger Long COVID cohort. We used ^{129}Xe-MRI to help determine whether this key symptom is driven by abnormalities not readily detectable by conventional physiology and imaging tests. We aimed to establish whether abnormalities observed were specific to those with breathlessness or also identified in those with fatigue-predominant presentations without reported breathlessness. Finally, we investigated whether abnormalities observed correlated with lung function tests and measures of dyspnoea and exercise capacity.

Materials and Methods

Participants were diagnosed with Long COVID by a clinical specialist in a dedicated clinic and had symptoms for a minimum of three months. With the exception of first-wave participants for whom SARS-CoV2 testing was not readily available, others had a positive PCR or rapid antigen test during their index infection. The symptom of breathlessness was determined at the time of the consultation by the referring physician undertaken at the time of referral to the study. Control participants formed two groups: 1) LC with typical symptoms but without breathlessness (NBLC); 2) participants who had proven COVID-19 infection more than three months previously and had fully recovered. The second control group were recruited by word of mouth and poster advertisement. Written informed consent was obtained from all participants.

Participants who had identifiable explanations for breathlessness were excluded from the study. This included those with CT scores >5 (detailed below), uncontrolled asthma, obstructive spirometry ($FEV_1/FVC<0{\cdot}7$), current smokers, smoking history >10 pack years and clinically significant cardiorespiratory pathology were all exclusion criteria. Contraindications for ^{129}Xe-MRI followed those of conventional MRI: claustrophobia, MRI-incompatible implanted devices or metal, and prohibitive BMI.

Study design

This was a prospective, observational, multi-centre cohort study enrolling non-hospitalised LC patients from four dedicated post-COVID-19 centres in the UK (Oxford, Sheffield, Cardiff, and Manchester) between May 2022 to March 2024. All study procedures were undertaken in Oxford and Sheffield sites and within 6 weeks of ^{129}Xe-MRI. This study was designed and implemented with the assistance of a Long COVID Public and Patient Involvement (PPI) group based at the Oxford site. It was approved by the Research Ethics Committee and informed consent was provided by all participants.

Demographic details and comorbidities were collected for all participants. LC participants completed the following patient-reported outcome measures: Dyspnoea-12 (D-12), Functional Assessment of Chronic Illness Therapy–Fatigue (FACIT-F), Generalised Anxiety Disorder 7-item scale (GAD-7), and EuroQoL visual analogue score (VAS).

Pulmonary function tests (PFTs) included forced expiratory volume in one second (FEV_1), forced vital capacity (FVC), and gas transfer (transfer factor of the lung for carbon monoxide (TLCO), alveolar volume (Va) and carbon monoxide transfer coefficient (KCO)). Lung function test data are presented as both % predicted and z-scores derived from the European Respiratory Society and Global Lung Function Initiative online calculator.[16,17] z-scores below 1.64 were considered abnormal. Exertional dyspnoea was assessed by a one-minute sit-to-stand test (STST), with the number of repetitions recorded alongside modified BORG dyspnoea score (mBORG), heart rate, and oxygen saturations pre- and post-exertion.

The study was funded by the National Institute for Health and Care Research (NIHR) and approved by the South Central Research Ethics Committee (21/SC/0398). It was registered on International Standard Randomised Controlled Trial Number (ISRCTN14304264).

Imaging Protocols

^{129}Xe dissolved phase imaging was acquired at either 3T (Oxford) or 1·5T (Sheffield) following inhalation of one-litre maximum of mixture of ^{129}Xe and nitrogen from functional residual capacity. Further details are provided in the OLS.

Maps of ^{129}Xe RBC:M, M:Gas and RBC:Gas were constructed for each subject. Gas transfer ratios acquired at 1·5T and 3T were combined using correction factors based on scanning 14 healthy volunteers at 1·5 and 3T. Age and sex adjusted z-scores were calculated for each mean metric based on published data.[7] Further details on gas polarisation, image acquisition, and post-processing are provided in the OLS.

Regional metrics of gas transfer (RBC:M, RBC:Gas) were calculated by dividing whole lung masks evenly in the inferior-superior axis into three regions (lower, middle, and upper lungs). Gas transfer maps were also divided evenly in the anterior-posterior axis into two regions corresponding to anterior and posterior lung.

Research CT scanning was performed following one-litre inspiration (for potential comparative purposes with ^{129}Xe-MRI) with a slice thickness of 0·625-1·000mm. Participants who had had a clinical CT to investigate breathlessness did not have a repeat research CT due to the initial scan being unremarkable and additional radiation exposure from a repeat CT scan unjustified. CTs were scored to measure presence of interstitial abnormality, detailed in the OLS. The images were reviewed by a senior radiologist blinded to research data

Statistical analysis

Analyses were conducted using STATA v18·0. Normality was assessed visually; non-parametric models were used if assumptions were violated. Data are presented as mean (SD) [min, max] for normally distributed and median (IQR) [min, max] for non-normally distributed variables. Group comparisons used ANOVA for normally distributed continuous variables, Kruskal-Wallis for non-normal data, and Chi-square or Fisher's exact test for categorical variables.

MRI metrics were summarised by group and compared using linear regression models adjusted for age and sex, unless z-scores already accounted for these. Missing data were not imputed and arose from incomplete questionnaires, inability to perform tests, or MRI quality control (due to low image quality, incorrect acquisition of scan parameters, or image artefacts).

No multiplicity correction was applied, given the exploratory nature of the analysis. A two-sided p-value <0·05 was considered significant.

The proportion of subjects with ^{129}Xe-MRI abnormalities (z-score<-1·64) was summarised by group (breathless and non-breathless vs controls) and compared using logistic regression, adjusted for age and sex. Model outputs included group difference estimates, 95% CIs, and P-values.

Correlations were assessed between (i) D-12 and FACIT-F, GAD-7, EuroQoL-VAS, and (ii) RBC:Gas, RBC:M and M:Gas z-scores and TLCO, KCO, Va, spirometry z-scores, BMI, and D-12. Pearson or Spearman coefficients were used based on distribution normality.

Results

Recruitment

269 participants were enrolled into the study. 24 were withdrawn due to: withdrawal of consent (5), symptom resolution (2), loss of contact (10), and abnormal CT chest findings (7). Following ^{129}Xe-MRI scan acquisition, 49 were excluded from the final analysis due to: current smoker status (2), obstructive spirometry (FEV_1/FVC ratio <0·7) (26), inability to complete PFTs (1) and image acquisition/quality issues (20). 196 participants were included in the final ^{129}Xe-MRI analysis (109 BLC, 42 NBLC and 44 controls) (Figure 2).

Study group characteristics

Age, gender, BMI, or past smoking status were not significantly different between breathless Long COVID (BLC), non-breathless LC (NBLC) and control participants. There was a longer time period from infection to MRI scanning in the BLC group compared with NBLC and control participants (607 days versus 375 and 232 days, respectively). The incidence of asthma and diabetes were similar between groups.

No statistically significant differences were found between the BLC and NBLC groups using questionnaires for fatigue (FACIT-F), anxiety (GAD-7), and general health status

(EuroQoL-VAS). The Dyspnoea was measure with the D-12, which was higher in the BLC than the NBLC group although there was considerable overlap in the groups and only 13 (30·2%) NBLC participants reported a score of 0 (no breathlessness) (Table 1).

Combining all LC participants, there was a positive correlation in breathlessness measured by D-12 score with fatigue (FACIT-F), anxiety (GAD-7), and health status (EQ5D-VAS) (Figure S1).

^{129}Xe-MRI analysis

Global analysis

We analysed global measures of gas exchange using ^{129}Xe-MRI. We found no significant differences between BLC, NBLC, and control participants in RBC:Gas, RBC:M or M:Gas z-scores (Figures 3–4 and Table 2). The percentage of participants with z-scores <-1·64 ('abnormal') did not differ for either RBC:M z-scores (18%,19%, and 23% for Controls, BLC and NBLC, respectively) or RBC:Gas z-scores (16%, 23%, and 20%, respectively). Although the graphs demonstrated that there was a non-significant trend towards lower z-scores in the BLC and NBLC subgroups with a greater proportion having <-1·64 z-scores compared with controls. To assess regional heterogeneity in ^{129}Xe distribution, we also measured the coefficient of variation for both RBC:Gas, RBC:M and M:Gas and found no significant differences between groups.

There was a positive correlation between RBC:Gas z-score with laboratory measures of gas transfer, TLCO, and KCO z-score (r=0·404 and 0·246, respectively, both p<0·001). However, RBC:M z-score was positively correlated with TLCO z-score (r=0·201, p=0·001) but not KCO z-score (r=0·089, p=0·103). RBC:M z-score but not RBC:Gas z-score correlated with BMI and Va. Neither RBC:M and RBC:Gas z-scores were found to correlate with spirometry values or D-12 scores (Table S1).

^{129}Xe-MRI regional analysis

To determine if there were regional abnormalities that might be masked by global scores, the lungs of each participant were divided into upper, middle, and lower zones and analysed individually (Figure S2). Separately, anterior and posterior lung divisions were also analysed. No significant differences were observed between any of these groups (Table S2).

Lung function

The majority of participants had lung function values within normal range, although the group mean z-scores for FEV_1 and FVC were both statistically lower in the BLC group than the control group (FEV_1 p=0·023 and FVC p=0·043, respectively, Figure 5). Overall, only seven (6·4%) of BLC participants had a z-score less than -1·64 (i.e. below the lower limit of normal) for FEV_1 and three (2·7%) had a z-score below -1·64 for FVC.

Group mean TLCO % predicted and z-scores were all within the normal ranges across all groups but were lower in BLC compared to NBLC and controls (BLC 91·2%, NBLC 96·8%,

100·5%; p<0·001 and BLC -0·7, NBLC -0·3, controls 0·0; p<0·001, respectively) (Table 3). Mean % predicted TLCO lies within normal range at 91% in the BLC group. The data on lung volume (Va) demonstrates a similar distribution to the gas transfer values, although in this case, the differences between groups were not statistically significant (Figure 5e). The proportion of participants with TLCO z-score <-1·64 was higher in the BLC but not significantly different from the NBLC and control groups (BLC - 14 (13%), NBLC – three (7%), controls – two (4·5%), p=0·279).

To examine whether the characteristics of BLC participants with lower TLCO z-score (<-1·64, n=14) differed from those with normal values, we compared clinical and physiological data, and found that these participants had higher D-12 scores (mean 22·3 vs 13·6), lower Va (-1·4 versus -0·8, p=0·020) and statistically lower RBC:Gas z-scores (-0·8 versus -0·7, p=0·008) and RBC:M z-scores (-1·0 versus -0·9, p=0·049). (Table S3).

Exercise tests and physiological data

Pre-exertion resting mBORG scores were higher in BLC compared to NBLC and controls and remained significantly higher in the BLC participants post-exertion (Fig S3 a-b, both p<0·001). Table S4 demonstrated that BLC participants performed fewer STS repetitions than both NBLC participants and controls (p<0·001 for both absolute number and percentile accounting for age and sex). No differences were seen in oxygen saturations and heart rate pre- and post-STS between the groups.

Discussion

This is the first large-scale multicentre study to comprehensively investigate whether lung abnormalities are associated with breathlessness in never-hospitalised Long COVID (LC) patients. Previous studies using ^{129}Xe-MRI in post-COVID-19 and pulmonary fibrosis populations have demonstrated that this modality can detect functional abnormalities at the level of the alveolar-capillary membrane not easily identified with conventional tests. Longitudinal ^{129}Xe-MRI studies in other diseases have shown the technique to be highly reproducible and capable of detecting early changes in disease state.[18,19] In this study, we used ^{129}Xe-MRI alongside established respiratory investigations to assess the lungs of breathless and non-breathless LC patients, aiming to understand the relationship between symptoms and lung function and to identify potential underlying pathophysiological processes affecting the lung parenchyma.

Our results show that in the majority of never-hospitalised LC participants, gas transfer measurements using ^{129}Xe-MRI were similar to those of fully recovered controls. In a subgroup of breathless LC participants (n=14; 12.9%) with modest reductions in TLCO, we found that TLCO correlated with lower RBC:Gas and RBC:M z-scores and higher

breathlessness scores, suggesting that a small subset of LC patients may have underlying abnormalities contributing to their symptoms.

In a pilot study, significant reductions in RBC:M scores were observed in 54% (n=6) of breathless LC participants.[15] In contrast, the EXPLAIN study found mild reductions in RBC:M z-scores (<1.64) in only 20% of breathless participants, and no significant difference in the prevalence of RBC:M or RBC:Gas abnormalities between breathless LC participants and controls. These differences may relate to sample size and timing of recruitment.[14,15] Earlier studies enrolled participants during the initial waves of the pandemic, when infection severity was higher and residual lung damage more likely to be detected on ^{129}Xe-MRI. Although EXPLAIN included participants from early waves, most were recruited during later phases, when COVID-19 generally caused milder manifestations.[20] LC participants were diagnosed following evaluation in dedicated clinics according to NICE criteria.[3,21] Questionnaires related to fatigue, breathlessness, and health status confirmed substantial symptom burden, indicating that these findings are likely representative of the broader Long COVID population.[1,6]

Lung function values were generally lower in the predefined breathless LC group compared with non-breathless LC and control participants. However, most z-scores remained within the internationally standardised 'normal range,' indicating that observed differences were unlikely to explain the breathlessness reported. While some lung function metrics showed statistically significant differences, these are unlikely to be clinically meaningful or account for symptoms in most breathless LC participants. Pre-COVID lung function was unknown, limiting causal interpretation, although groups were matched for age, gender, BMI, smoking status, and comorbidities, reducing potential confounding. Interestingly, lower lung volumes in some LC individuals have also been observed in other non-hospitalised LC cohorts.[22]

Comparing breathless participants with TLCO z-scores <-1.64 to those within the normal range revealed some notable differences. Participants with low TLCO had higher D-12 scores, indicating greater breathlessness, and lower RBC:Gas values. Given the heterogeneous pathophysiology of LC, it is plausible that those with low TLCO represent a subgroup in which lung damage contributes to symptoms. These participants also had lower KCO and alveolar volume (Va), suggesting that gas transfer abnormalities reflect changes in both ventilated airspace and efficiency of gas exchange between alveoli and pulmonary capillaries.[15] However, caution is warranted in interpreting these findings, as the absolute differences were modest and unlikely to account for the excessive dyspnoea reported. Due to the absence of longitudinal data, it cannot be determined whether impaired gas transfer recovers over time.

The sit-to-stand (STS) test demonstrated that both breathless and non-breathless LC participants performed fewer repetitions than controls, indicating reduced exercise

capacity. Deconditioning has been identified as a major factor in exercise limitation among LC populations.[23] The mBORG dyspnoea score, commonly used to quantify breathlessness after exercise, was elevated in about one-third of breathless participants pre-STS.[24] Lung function measures derived from ^{129}Xe-MRI and traditional pulmonary function tests cannot account for this degree of breathlessness. Alternative/additional explanations include breathing pattern disorder (BPD), fatigue, subclinical ventilation abnormalities, deconditioning, anxiety, weight gain, and pain.[6,25-27] Breathlessness and reduced exercise capacity may also relate to metabolic or structural abnormalities outside the lungs. Small studies have linked exercise intolerance to impaired oxygen extraction by skeletal muscle and demonstrated structural and metabolic changes in peripheral muscles of LC participants.[28,29]

This study has limitations. Participants were referred by clinical teams and defined as 'breathless' or 'non-breathless' following assessment. Breathlessness scores (D-12, mBORG) obtained on the day of research visits showed overlap between groups, with some 'non-breathless' participants reporting elevated scores. No correlation was observed between D-12 scores and ^{129}Xe-MRI metrics (Table 3), highlighting the difficulty of using such a complex and multifaceted symptom to differentiate groups. Several previous studies have also demonstrated poor correlation between breathlessness and objective respiratory pathology.[30-32] A key strength of this study is recruitment from a real-world clinical setting based on self-reported symptoms, reinforcing the finding that the majority of participants do not exhibit lung parenchymal abnormalities explaining their breathlessness.

^{129}Xe-MRI is a promising tool for detecting subclinical respiratory disease. The lack of significant differences between groups provides reassurance that lung pathology is not a primary driver of breathlessness for most LC participants. Nevertheless, its use remains largely within research, and its clinical utility and limitations are still emerging. Traditional lung function z-scores are derived from thousands of healthy participants, whereas this study's z-scores were based on 62 volunteers, limiting adjustment for demographic effects on gas transfer metrics.[7,33] Larger studies are needed to establish normative ^{129}Xe-MRI data across age, gender, and ethnicity. Additionally, correction factors used to adjust for field-strength differences may not fully account for technical variability, although statistical findings for RBC:M and RBC:Gas were consistent when analysed separately at different field strengths (Tables S5-6), supporting the robustness of the conclusions.

In summary, this study did not identify significant gas transfer abnormalities affecting the alveolar-capillary membrane in Long COVID participants. The combination of novel functional imaging and traditional investigations provides evidence that occult lung abnormalities do not account for breathlessness in the majority of LC patients.

Acknowledgement of financial support

This research is funded by the National Institute for Health and Care Research (NIHR) (Long Covid grant, Ref: COV-LT2-0049) with additional funding from Sheffield Biomedical Research Centre (NIHR203321), Oxford Biomedical Research Centre, MRC (grant MR/M008894/1) and INSIGNEO institute for in silico medicine. The views expressed are those of the author(s) and not necessarily those of the NIHR or the Department of Health and Social Care.

	Control subjects (N=44)	*Breathless Long COVID (N=109)*	*Non-breathless Long COVID (N=43)*	*P-value**
Age (years) (SD) [range]	45·7 (13·4) [23·3 to 72·6]	46·9 (12·8) [18·8 to 81·4]	45·9 (10·8) [22·1 to 64·1]	0·821
BMI (kg/m²)	28·4 (5·8) [19·1 to 45·5]	28·4 (6·4) [16·4 to 49·5]	27·7 (5·1) [20·8 to 40·0]	0·802
Ethnicity, n (%)				p(exact) = 0·015
White	38 (86·4%)	94 (86·2%)	43 (100·0%)	
Non-white	6 (13·6%)	15 (13·8%)	0 (0·0%)	
Sex, n (%)				p = 0·156
Female	34 (77·3%)	71 (65·1%)	25 (58·1%)	
Male	10 (22·7%)	38 (34·9%)	18 (41·9%)	
CT score‡	0 (0 to 0) [0 to 3]	0 (0 to 0) [0 to 4]	0 (0 to 0) [0 to 2]	0·195
Missing, n (%)	0 (0·0%)	1 (0·9%)	0 (0·0%)	
Time from infection to CT (days) ‡	246 (183 to 505) [110 to 1319]	440 (245 to 749) [28 to 1489]	329 (220 to 632) [25 to 1275]	0·026
Missing, n (%)	0 (0·0%)	2 (1·8%)	0 (0·0%)	
Time from infection to MRI (days) ‡	234 (180 to 512) [109 to 1319]	632 (349 to 893) [112 to 1489]	352 (212 to 756) [120 to 1275]	<0·001
Missing, n (%)	0 (0·0%)	1 (0·9%)	0 (0·0%)	
Vaccinated, n (%)				p(exact) = 0·045
No	0 (0·0%)	8 (7·3%)	0 (0·0%)	
Yes	44 (100·0%)	101 (92·7%)	43 (100·0%)	
Smoking, n (%)				p = 0·729
Former	10 (22·7%)	29 (26·6%)	9 (20·9%)	
Diabetes, n (%)				p(exact) = 0·797
No	43 (97·7%)	104 (95·4%)	41 (95·3%)	
Yes	1 (2·3%)	5 (4·6%)	2 (4·7%)	
Asthma, n (%)				p(exact)=0·386
Childhood only	4 (9·1%)	9 (8·3%)	2 (4·7%)	
No	33 (75·0%)	86 (78·9%)	39 (90·7%)	
Dyspnoea-12‡	NA	13 (9 to 19) [1 to 36]	3 (0 to 6) [0 to 27]	<0·001
FACIT F	NA	21 (11) [2 to 52]	23 (10) [9 to 49]	0·413
GAD-7‡	NA	6 (1 to 11) [0 to 21]	5 (3 to 7) [0 to 20]	0·591
EuroQoL VAS	NA	53·2 (19·1) [0·0 to 95·0]	55·5 (17·4) [20·0 to 90·0]	0·520
Missing, n (%) for questionnaires	NA	27 (24·8%)	5 (11·6%)	

‡Median, interquartile range and minimum and maximum presented for non-normally distributed continuous variables

Table 1. The demographics and health characteristics for control, breathless Long COVID, and non-breathless Long COVID participants included within the study. Symptom-based questionnaires were evaluated in

breathless and non-breathless LC participants only. Data are presented as mean (SD) [min, max] unless otherwise stated. Group comparisons based on ANOVA for normally distributed continuous variables, Kruskal-Wallis for non-normally distributed continuous variables, and chi-square or Fisher's exact test for categorical variables. BMI – Body mass index; Dyspnoea-12 – breathlessness questionnaire; FACIT-F – Functional Assessment of Chronic Illness Therapy – Fatigue; GAD-7 – Generalised Anxiety Disorder-7; and EQ5D VAS – EuroQol Visual Analogue Scale 'Health Today'.

	Control subjects (N=44)	*Breathless Long COVID (N=109)*	*Non-breathless Long COVID (N=43)*	*P-value*
RBC:Gas mean	0·0028 (0·0010) [0·0013 to 0·0053]	0·0027 (0·0010) [0·0007 to 0·0058]	0·0030 (0·0012) [0·0012 to 0·0064]	0·366
RBC:Gas standard deviation	0·0010 (0·0004) [0·0005 to 0·0026]	0·0009 (0·0003) [0·0003 to 0·0025]	0·0010 (0·0005) [0·0004 to 0·0025]	0·362
RBC:Gas coefficient of variation	0·36 (0·09) [0·21 to 0·64]	0·36 (0·12) [0·22 to 1·07]	0·33 (0·06) [0·22 to 0·46]	0·170
RBC:Gas z-score	-0·63 (1·00) [-2·41 to 1·52]	-0·80 (1·05) [-3·28 to 2·37]	-0·55 (1·30) [-2·53 to 2·68]	0·403
RBC:M mean	0·35 (0·08) [0·23 to 0·53]	0·35 (0·09) [0·16 to 0·62]	0·36 (0·10) [0·19 to 0·72]	0·824
RBC:M standard deviation	0·09 (0·03) [0·05 to 0·18]	0·08 (0·02) [0·05 to 0·17]	0·08 (0·02) [0·05 to 0·17]	0·434
RBC:M coefficient of variation	0·27 (0·08) [0·12 to 0·43]	0·26 (0·07) [0·14 to 0·46]	0·24 (0·06) [0·13 to 0·37]	0·153
RBC:M z-score	-0·51 (1·00) [-2·01 to 2·15]	-0·62 (1·17) [-3·58 to 2·63]	-0·66 (1·30) [-3·47 to 2·88]	0·821
M:Gas mean	0·0085 (0·0019) [0·0058 to 0·0140]	0·0083 (0·0023) [0·0035 to 0·0166]	0·0091 (0·0028) [0·0046 to 0·0184]	0·195
M:Gas coefficient of variation	0·2349 (0·0635) [0·1464 to 0·4732]	0·2367 (0·0707) [0·1543 to 0·6640]	0·2224 (0·0403) [0·1650 to 0·3587]	0·264
M:Gas Z-score	-0·2607 (0·9033) [-1·5244 to 2·3799]	-0·3396 (1·0756) [-2·6099 to 3·6530]	0·0463 (1·3399) [-2·1109 to 4·4811]	0·154
Missing, n (%)	1 (2·3%)	0 (0·0%)	0 (0·0%)	

Table 2: Global ^{129}Xe-MRI measurements in control subjects, breathless, and non-breathless Long COVID participants. RBC:Gas, RBC:M and M:Gas z-score and coefficient of variation measurements displayed. Summaries presented as mean, mean standard deviation, and minimum and maximum. P-values derived from ANOVA, adjusted for age and sex unless already adjusted via z-score.

Measure	*Control subjects (N=44)*	*Breathless Long COVID (N=109)*	*Non-Breathless Long COVID (N=43)*	*P-value**
FVC z-score	0·4 (1·0) [-1·5 to 2·9]	0·0 (1·0) [-2·1 to 2·6]	0·4 (0·9) [-1·2 to 3·3]	**0·043**
FVC % predicted (%)	105·4 (11·9) [80·4 to 131·8]	100·8 (12·9) [73·7 to 132·6]	105·1 (12·7) [84·8 to 145·2]	0·053
FEV_1 z-score	0·3 (0·8) [-1·2 to 2·1]	-0·0 (0·9) [-2·7 to 1·9]	0·3 (1·0) [-1·3 to 4·0]	**0·023**
FEV_1 % predicted (%)	103·7 (10·2) [82·7 to 126·5]	99·6 (12·0) [65·8 to 125·3]	104·6 (12·8) [84·1 to 150·6]	**0·0128**
TLCO z-score	-0·0 (1·0) [-2·9 to 1·7]	-0·7 (0·9) [-3·3 to 1·5]	-0·3 (0·9) [-2·1 to 1·8]	**<0·001**
TLCO % predicted (%)	100·5 (13·5) [66·0 to 125·2]	91·2 (12·8) [58·1 to 124·4]	96·8 (13·3) [73·3 to 132·9]	**<0·001**
KCO z-score	0·4 (1·0) [-1·2 to 3·3]	0·0 (1·0) [-3·7 to 2·3]	0·2 (0·9) [-1·5 to 2·5]	0·059
KCO % predicted (%)	106·8 (15·8) [84·9 to 149·6]	100·7 (13·5) [53·8 to 131·4]	102·9 (13·0) [78·7 to 136·2]	0·050
VA z-score	-0·5 (1·0) [-2·8 to 1·3]	-0·8 (1·0) [-3·5 to 1·7]	-0·6 (0·9) [-3·3 to 1·5]	0·103
Va % predicted (%)	94·3 (11·4) [69·1 to 115·4]	90·9 (11·1) [62·3 to 122·4]	94·0 (10·0) [67·6 to 118·0]	0·111
Missing, n (%)	0 (2·3%)	1 (0·9%)	0 (0%)	

Table 3: Lung function measures by group with tests of group differences. Both z-scores and % predicted values are displayed. Summaries presented as mean, mean standard deviation, and minimum and maximum. *P-values derived from ANOVA. FVC - Forced Vital Capacity; FEV_1 - Forced Expired Volume in one second; TLCO - Transfer factor for Lung carbon monoxide; KCO - carbon monoxide transfer coefficient; Va - Alveolar Volume.

References

1 Prevalence of ongoing symptoms following coronavirus (COVID-19) infection in the UK: 30 March 2023. *Office for National Statistics* 2023. [Available from: https://www.ons.gov.uk/peoplepopulationandcommunity/healthandsocialcare/conditionsanddiseases/bulletins/prevalenceofongoingsymptomsfollowingcoronaviruscovid19infectionintheuk/30march2023.]

2 Lin LY, Henderson AD, Carlile O, et al. Healthcare utilisation in people with long COVID: an OpenSAFELY cohort study. *BMC Med* 2024; 22(1): 255.

3 National Institute for Health and Care Excellence (NICE). COVID-19 rapid guideline: managing the long-term effects of COVID-19: *National Institute for Health and Care Excellence: Clinical Guidelines* 2024. [Available from: https://www.nice.org.uk/guidance/ng188/chapter/1-Identification.

4 Stewart I, Jacob J, George PM, et al. Residual Lung Abnormalities after COVID-19 Hospitalization: Interim Analysis of the UKILD Post-COVID-19 Study. *Am J Respir Crit Care Med* 2023; 207(6): 693–703.

5 Evans RA, McAuley HJC, Harrison EM, et al. Physical, cognitive, and mental health impacts of COVID-19 after hospitalisation (PHOSP-COVID): a UK multicentre, prospective cohort study. *Lancet Respir Med* 2021; 9(11): 1275–1287.

6 Heightman M, Prashar J, Hillman TE, et al. Post-COVID-19 assessment in a specialist clinical service: a 12-month, single-centre, prospective study in 1325 individuals. *BMJ Open Respir Res* 2021;8(1): e001041.

7 Collier GJ, Smith LJ, Saunders LC, et al. Age, sex, and lung volume dependence of dissolved xenon-129 MRI gas exchange metrics. *Magn Reson Med* 2024; 92(4): 1471–1483.

8 Marshall H, Stewart NJ, Chan HF, et al. In vivo methods and applications of xenon-129 magnetic resonance. *Prog Nucl Magn Reson Spectrosc* 2021; 122: 42–62.

9 Roos JE, McAdams HP, Kaushik SS, et al. Hyperpolarized Gas MR Imaging: Technique and Applications. *Magn Reson Imaging Clin N Am* 2015; 23(2): 217–229.

10 Ng KL, Grist J, Gleeson F, et al. Clinical Applications of Hyperpolarised Xenon-129 MRI. *EMJ* 2024; 9(4): 48–61.

11 Grist JT, Chen M, Collier GJ, et al. Hyperpolarized (129)Xe MRI Abnormalities in Dyspneic Patients 3 Months after COVID-19 Pneumonia: Preliminary Results. *Radiology* 2021; 301(1): E353–E360.

12 Saunders LC, Collier GJ, Chan HF, et al. Longitudinal Lung Function Assessment of Patients Hospitalized With COVID-19 Using (1)H and (129)Xe Lung MRI. *Chest* 2023; 164(3): 700–716.

13 Li H, Zhao X, Wang Y, et al. Damaged lung gas exchange function of discharged COVID-19 patients detected by hyperpolarized ^{129}Xe MRI. *Sci Adv* 2021; 7: eabc8180.

14 Matheson AM, McIntosh MJ, Kooner HK, et al. Longitudinal follow-up of postacute COVID-19 syndrome: DL_{CO}, quality-of-life and MRI pulmonary gas-exchange abnormalities. *Thorax* 2023; 78: 418–421.

15 Grist JT, Collier GJ, Walters H, et al. Lung Abnormalities Depicted with Hyperpolarized Xenon MRI in Patients with Long COVID. *Radiology* 2022; (305)3: 709–717.

16 Stanojevic S, Graham BL, Cooper BG, et al. Official ERS technical standards: Global Lung Function Initiative reference values for the carbon monoxide transfer factor for Caucasians. *Eur Respir J* 2017; 50(3): 1700010.
17 Quanjer PH, Stanojevic S, Cole TJ, et al. Multi-ethnic reference values for spirometry for the 3-95-yr age range: the Global Lung Function 2012 equations. *Eur Respir J* 2012; 40(6):1324–1343.
18 Hahn AD, Carey KJ, Barton GP, et al. Hyperpolarized (129)Xe MR Spectroscopy in the Lung Shows 1-year Reduced Function in Idiopathic Pulmonary Fibrosis. *Radiology* 2022; 305(3): 688–696.
19 Ruppert K, Qing K, Patrie JT, et al. Using Hyperpolarized Xenon-129 MRI to Quantify Early-Stage Lung Disease in Smokers. *Acad Radiol* 2018; 26(3): 355–366.
20 Kirwan PD, Charlett A, Birrell P, et al. Trends in COVID-19 hospital outcomes in England before and after vaccine introduction, a cohort study. *Nat Commun* 2022; 13(1): 4834.
21 Soriano JB, Murthy S, Marshall JC, et al. A clinical case definition of post-COVID-19 condition by a Delphi consensus. *Lancet Infect Dis* 2022; 22(4): e102–e107.
22 Fagevik Olsén M, Louise L, Malin NB, et al. Long COVID – respiratory symptoms in non-hospitalised subjects – a cross-sectional study. *Eur J Physiother* 2023; 25(5): 283–290.
23 Chamley RR, Holland JL, Collins J, et al. Exercise capacity following SARS-CoV-2 infection is related to changes in cardiovascular and lung function in military personnel. *Int J Cardiol* 2024; 395: 131594.
24 Mahler DA, Horowitz MB. Perception of breathlessness during exercise in patients with respiratory disease. *Med Sci Sports Exerc* 1994; 26(9): 1078–1081.
25 Frésard I, Genecand L, Altarelli M, et al. Dysfunctional breathing diagnosed by cardiopulmonary exercise testing in 'long COVID' patients with persistent dyspnoea. *BMJ Open Respir Res* 2022; 9(1): e001126.
26 Evans R, Pick A, Lardner R, et al. Breathing difficulties after Covid-19: a guide for primary care. *BMJ* 2023; 381: e074937.
27 Sandberg J, Olsson M, Ekström M. Underlying conditions contributing to breathlessness in the population. *Curr Opin Support Palliat Care* 2021; 15(4): 219–225.
28 Norweg A, Yao L, Barbuto S, et al. Exercise intolerance associated with impaired oxygen extraction in patients with long COVID. *Respir Physiol Neurobiol* 2023; 313: 104062.
29 Appelman B, Charlton BT, Goulding RP, et al. Muscle abnormalities worsen after post-exertional malaise in long COVID. *Nat Commun* 2024; 15(1): 17.
30 Müllerová H, Lu C, Li H, et al. Prevalence and burden of breathlessness in patients with chronic obstructive pulmonary disease managed in primary care. *PLoS One* 2014; 9(1): e85540.
31 Grønseth R, Vollmer WM, Hardie JA, et al. Predictors of dyspnoea prevalence: results from the BOLD study. *Eur Respir J* 2014; 43(6): 1610–1620.
32 Bell M, Fotheringham I, Punekar YS, et al. Systematic Review of the Association Between Laboratory- and Field-Based Exercise Tests and Lung Function in Patients with Chronic Obstructive Pulmonary Disease. *Chronic Obstr Pulm Dis* 2015; 2(4): 321–342.
33 Cooper BG, Stocks J, Hall GL, et al. The Global Lung Function Initiative (GLI) Network: bringing the world's respiratory reference values together. *Breathe (Sheff)* 2017; 13(3): e56–e64.

Figure Captions

Figure 1. Diagram showing a dissolved phase ^{129}Xe spectra.
^{129}Xe exhibits a different resonant frequency in the membrane and red blood cells (RBCs) allowing 3D volumetric images of ^{129}Xe in the gas, membrane, and RBC compartments to be reconstructed. Voxel-wise maps of RBC:Gas, RBC:M and M:Gas are then calculated from the fraction of signal seen in the RBC phase compared to the gas and membrane phase respectively.

Figure 2. Study profile illustrating participant recruitment, reasons for exclusion and numbers included in analyses.
Patients recruited from four centers but all study procedures undertaken in Oxford or Sheffield (Cardiff and Manchester participants, respectively). Patients with obstructive spirometry (FEV_1/FVC ratio of <0.7 were excluded as this could be an explanation for symptoms. PFT – pulmonary function tests; ^{129}Xe-MRI – Hyperpolarised xenon MRI; FEV_1 - Forced Expired Volume in one second; FVC - Forced Vital Capacity; BLC – breathless Long COVID; NBLC – non-breathless Long COVID.

Figure 3. Pulmonary ^{129}Xe-MRI image maps
a. Figure to demonstrate gas, membrane and red blood cell (RBC) ^{129}Xe-MRI image maps and how the ratios of RBC:Gas, RBC:M and M:Gas are calculated from gas and RBC measurements and membrane and RBC uptake, accordingly.
b. Representative image maps of the RBC:M, RBC:Gas and M:Gas mean ratio maps of a control (top), breathless Long COVID (middle), and non-breathless Long COVID (bottom) participant.

Figure 4. ^{129}Xe-MRI gas transfer parameters
a. RBC:M, b. RBC:Gas and c. M:Gas z-scores in control subjects, breathless, and non-breathless Long COVID groups.
ANOVA for overall comparisons between groups, and linear regression for pairwise comparisons. Horizontal dashed lines represent +/- 1·64 (i.e. the 90% normal range, values above 1·64 represent the top 5% and below -1·64 the bottom 5%) and +/-1·96 (the 95% normal range, the top and bottom 2·5%).

Figure 5. Lung function z-scores presented by group.
a. FEV_1, b. FVC, c. TLCO, d. KCO, and e. Va.
Horizontal dashed lines represent +/- 1·64 (i.e. the 90% normal range, values above 1·64 represent the top 5% and below -1·64 the bottom 5%) and +/-1·96 (the 95% normal range, the top and bottom 2·5%). P-values are given from an ANOVA model.
FEV_1 - Forced Expired Volume in one second; FVC - Forced Vital Capacity; TLCO - Transfer factor for Lung carbon monoxide; KCO - carbon monoxide transfer coefficient; Va – Alveolar Volume.

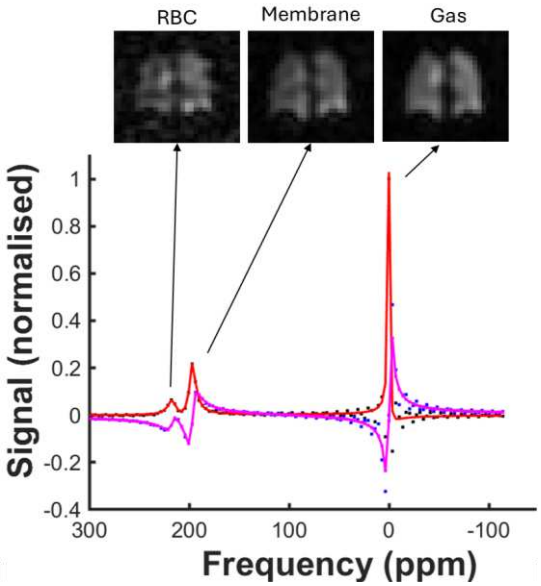


Figure 1. Diagram showing a dissolved phase ^{129}Xe spectra. ^{129}Xe exhibits a different resonant frequency in the membrane and red blood cells (RBCs) allowing 3D volumetric images of ^{129}Xe in the gas, membrane, and RBC compartments to be reconstructed. Voxel-wise maps of RBC:Gas, RBC:M and M:Gas are then calculated from the fraction of signal seen in the RBC phase compared to the gas and membrane phase respectively.

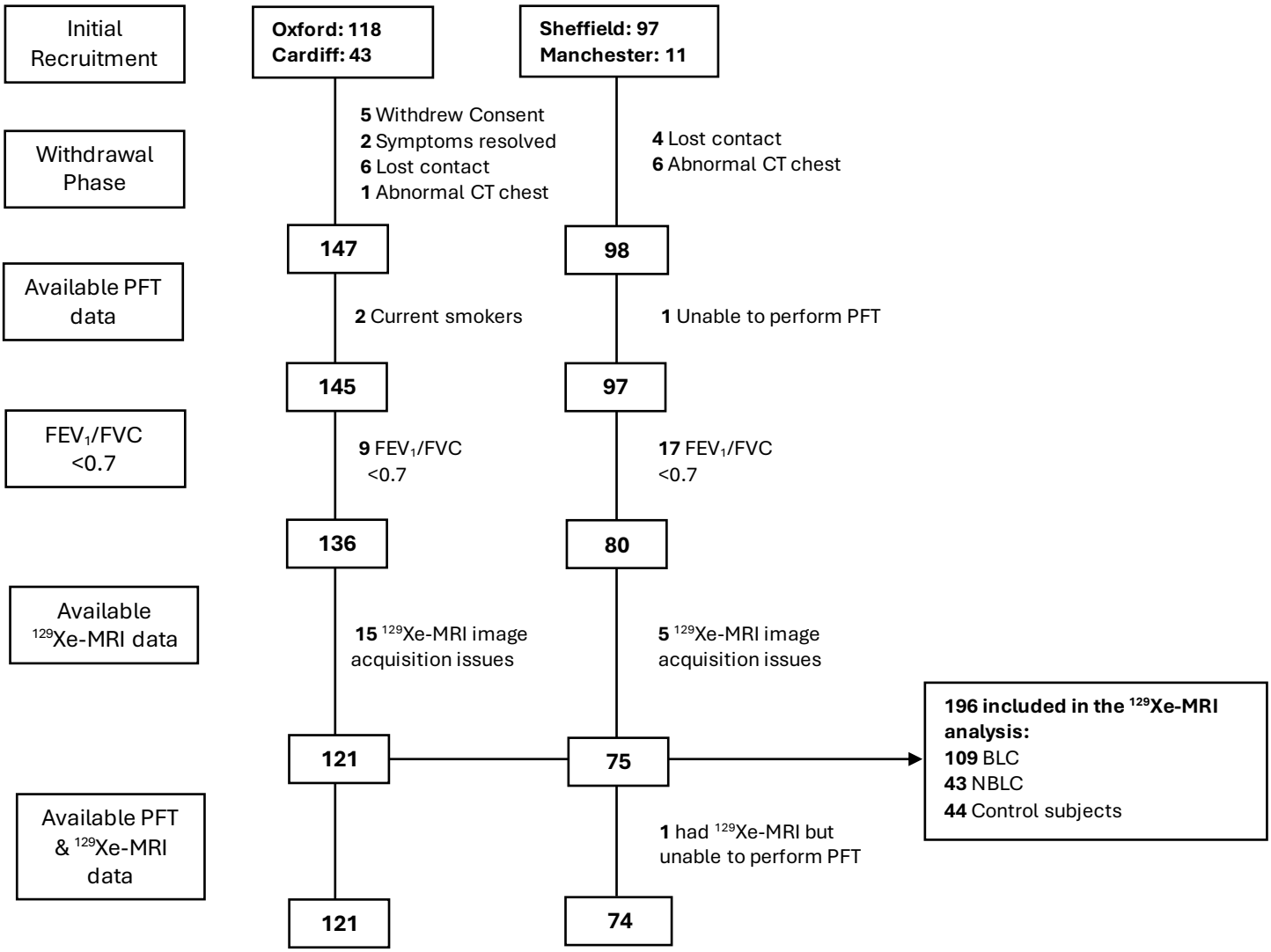
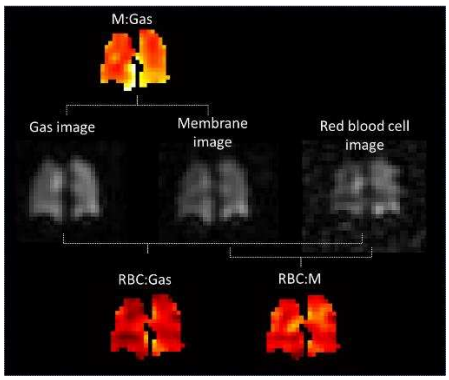


Figure 2. Study profile illustrating participant recruitment, reasons for exclusion and numbers included in analyses. Patients recruited from four centres but all study procedures undertaken in Oxford or Sheffield (Cardiff and Manchester participants, respectively). Patients with obstructive spirometry (FEV_1/FVC ratio of <0.7 were excluded as this could be an explanation for symptoms. PFT – pulmonary function tests; ^{129}Xe-MRI – Hyperpolarised xenon MRI; FEV_1 - Forced Expired Volume in one second; FVC - Forced Vital Capacity; BLC – breathless Long COVID; NBLC – non-breathless Long COVID.

a.



b.

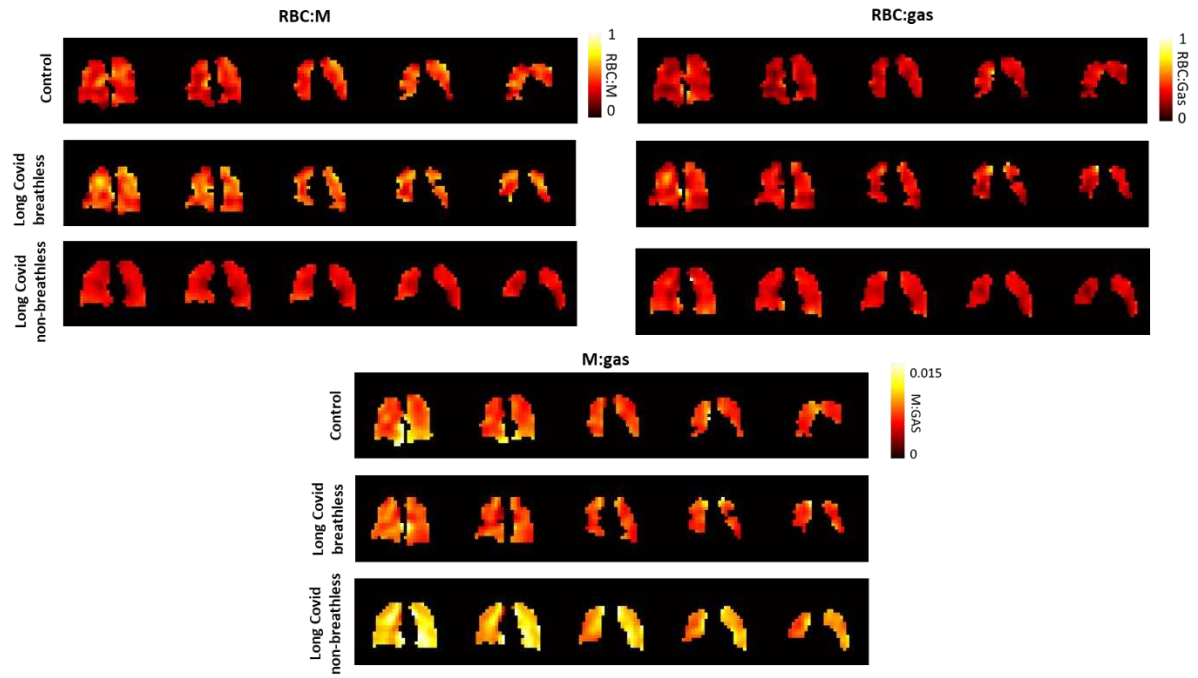


Figure 3: a) Figure to demonstrate gas, membrane and red blood cell (RBC) ^{129}Xe-MRI image maps and how the ratios of RBC:Gas, RBC:M and M:Gas are calculated from gas and RBC measurements and membrane and RBC uptake, accordingly. b) Representative image maps of the RBC:M, RBC:Gas and M:Gas mean ratio maps of a control (top), BLC (middle), and NBLC (bottom) participant.

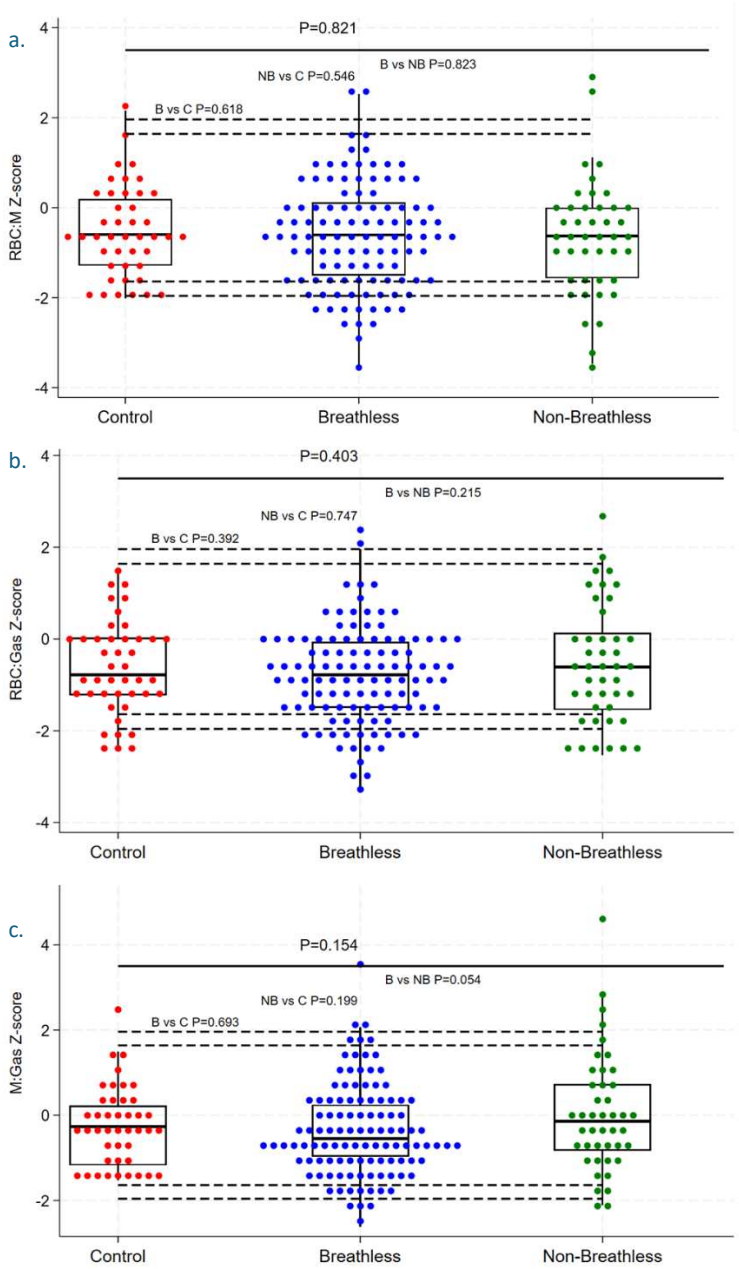


Figure 4: a) RBC:M, b) RBC:Gas and c) M:Gas z-scores in controls, breathless, and non-breathless Long COVID groups. ANOVA for overall comparisons between groups, and linear regression for pairwise comparisons. Horizontal dashed lines represent +/- 1·64 (i.e. the 90% normal range, values above 1·64 represent the top 5% and below -1·64 the bottom 5%) and +/-1·96 (the 95% normal range, the top and bottom 2·5%).

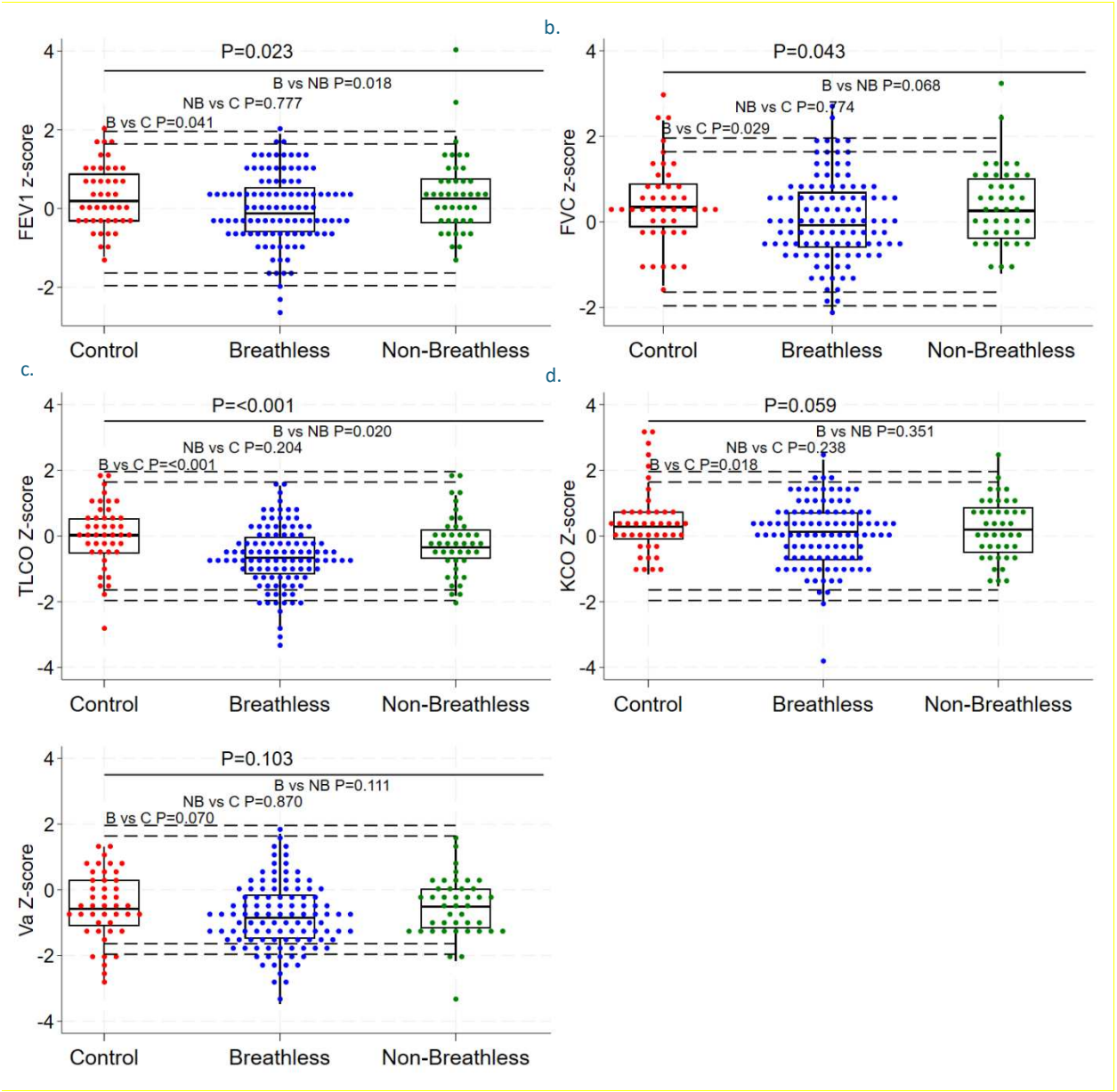


Figure 5. Lung function z-scores presented by group. a) FEV_1, b) FVC, c) TLCO, d) KCO, and e) Va. Horizontal dashed lines represent +/- 1·64 (i.e. the 90% normal range, values above 1·64 represent the top 5% and below -1·64 the bottom 5%) and +/-1·96 (the 95% normal range, the top and bottom 2·5%). P-values are given from an ANOVA model. FEV_1 - Forced Expired Volume in one second; FVC - Forced Vital Capacity; TLCO - Transfer factor for Lung carbon monoxide; KCO - carbon monoxide transfer coefficient; Va – Alveolar Volume.

Supplementary material

Methods

CT scoring

The following approach was used to visually score CT scans. The presence of ground glass opacification, reticulation, traction bronchial dilatation and honeycombing was assessed and CT images were scored as per the clinical standard of CT COVID-19 severity scoring system.[1] Each lung lobe was assessed individually, and the percentage involvement of the lobe was noted based on visual assessment, scoring 1 if < 5% area involved, 2 if 5–25% area involved, 3 if 25–50% area involved, 4 if 50–75% area involved, and 5 if > 75% area involved. Maximum CT severity score was 25 provided by the sum of the individual lobar scores.

Sit-to-stand test

The number of sit-to-stand repetitions was also reported in percentile (2.5th, 25th, 50th, 75th, 97.5th) using the reference values by age group published by Strassmann et al.[2]

^{129}Xe-MRI

Polarisation of xenon

^{129}Xe gas (86% 129Xe, Spectra Gases Inc.) was polarised as previously described.[3,4] Briefly, ^{129}Xe gas was polarised by rubidium vapour spin exchange optical pumping (SEOP), and cryogenically accumulated in 1-L doses using an in-house polariser (POLARIS, Sheffield) in Sheffield and in Oxford a commercial polariser (Model 9300, Polarean).

A 1L maximum of mixture of ^{129}Xe and nitrogen was inhaled from functional residual capacity.[5] For subjects scanned in Sheffield, the ^{129}Xe dose was titrated for subjects with height <160cm dose.[6] For subjects scanned in Oxford, 1L dose was used for all subjects.

Hyperpolarised xenon MRI Oxford and Sheffield centres

MRI was performed at 1·5T (General Electric Healthcare, WI, SIGNA Artist) in Sheffield and 3T (SIGNA Premier) in Oxford with imaging performed on a dedicated ^{129}Xe thorax transmit receive coil (CMRS in Sheffield and Pulseteq in Oxford) and the ^{1}H body coil (GE Healthcare).

Dissolved phase ^{129}Xe spectroscopic images were acquired using a 3D 4-echo flyback radial acquisition as previously described (image resolution=2·0cm in all dimensions, reconstructed isotropic resolution of 1·25cm, repetition time per spoke = 15ms, flip angle per excitation on dissolved and gas phases = 22 and 0·22 degrees, respectively, total scan time = 14s).[7]

Gas, Membrane (M), and Red Blood Cell (RBC) images were reconstructed, and ratiometric maps (RBC:M, RBC:Gas and M:Gas) were constructed. The mean and coefficient of variation of each map was calculated. Z-scores were calculated for each mean metric based on published data from 62 healthy participants scanned at 1·5 T (including 24 healthy participants from this study).[8] Gas transfer ratios acquired at 1·5T and 3T were combined using correction factors based on scanning 14 healthy volunteers at 1·5 and 3T.

Regional metrics of gas transfer (RBC:M, RBC:Gas, M:Gas) were calculated by dividing whole lung masks evenly in the inferior-superior axis into three regions (lower, middle, and upper lungs). Gas transfer maps were also divided evenly in the anterior-posterior axis into two regions corresponding to anterior and posterior lung.

Results

Correlation between Dyspnoea-12 and patient-reported outcome measures

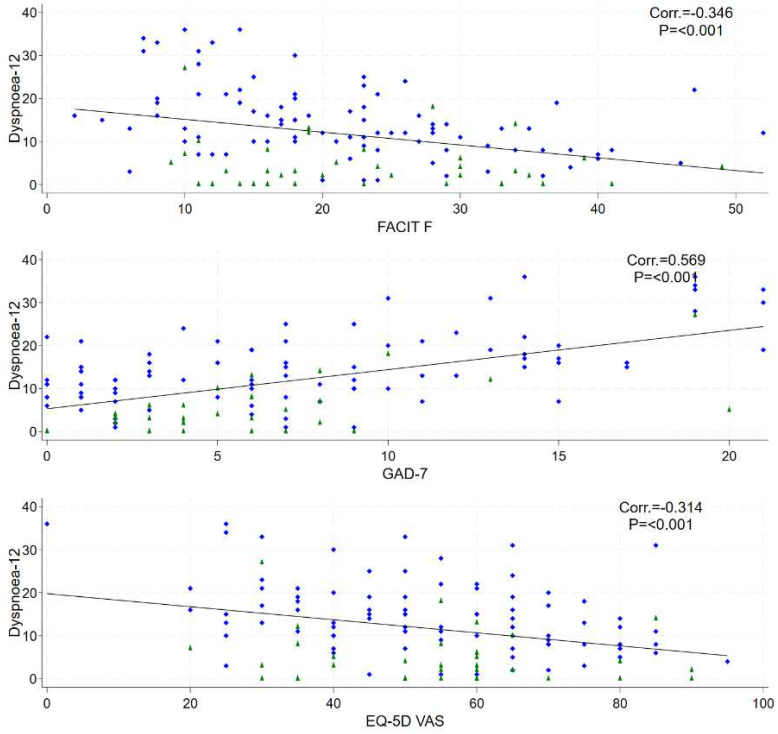


Figure S 1. Graphs showing the correlations between a) FACIT-F b) GAD-7 and c) EQ5D VAS scores and Dyspnoea-12 score in all long COVID (LC) participants. Higher D-12 values indicate higher breathlessness scores, lower FACIT-F scores indicate worse fatigue ratings, increased GAD-7 scores indicate higher anxiety scores, higher EQ-5D VAS scores indicate better perceived health. Blue markers represent the breathless and green the non-breathless LC groups.

FACIT-F – Functional Assessment of Chronic Illness Therapy – Fatigue; GAD-7 – Generalised Anxiety Disorder-7; and EQ-5D VAS – EuroQol Visual Analogue Scale 'Health Today'.

Regional metrics of RBC:Gas z-score

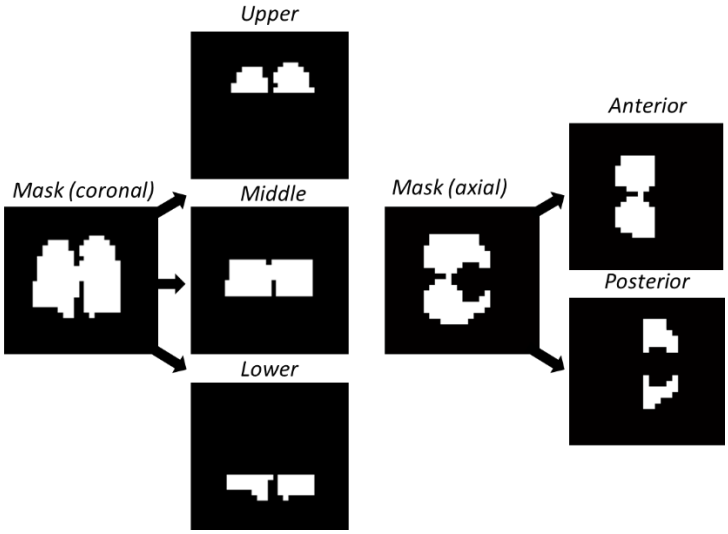


Figure S 2. The lungs were divided into upper, middle and lower areas, as well as anterior and posterior regions. Regional metrics of gas transfer (RBC:M, RBC:Gas and M:Gas z-scores) were calculated within each area to assess for evidence of regional abnormalities.

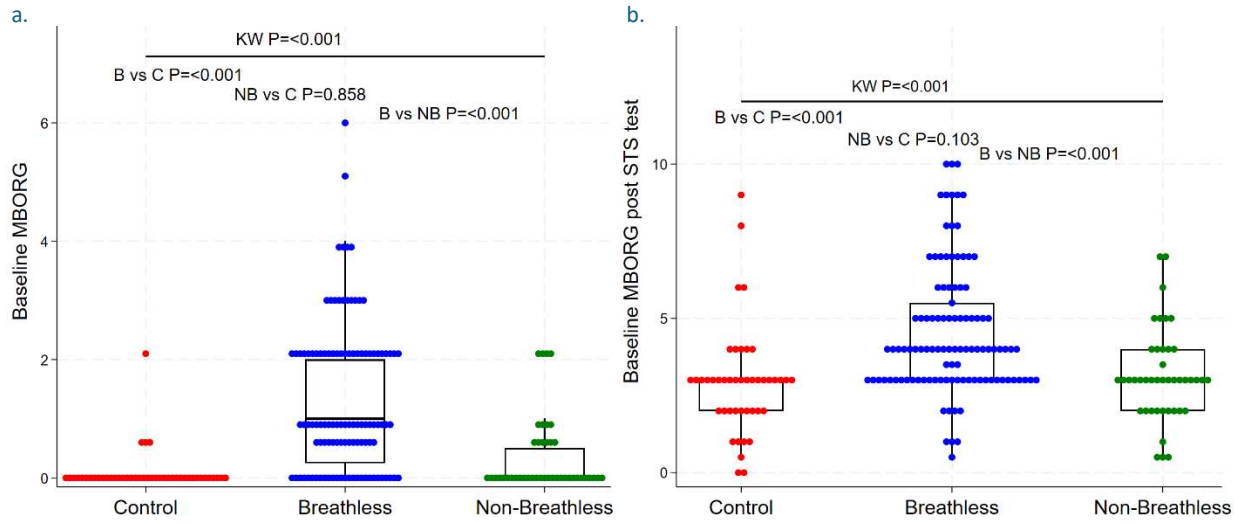


Figure S 3. Comparison of mBORG scores pre and post sit-to-stand test. a) Baseline and b) post STS change in modified BORG Dyspnoea score (mBORG) between controls, breathless (BLC), and non-breathless Long COVID (NBLC) participants.

Variable	Age and sex adjusted Correlation, P-value with RBC:Gas	Age and sex adjusted Correlation, P-value with RBC:M	Age and sex adjusted Correlation, P-value with M:Gas
FEV_1 z-score	0·100, p=0·151	0·040, P=0·467	0·102, P=0·161
FVC z-score	0·062, p=0·373	0·055, P=0·315	0·035, P=0·624
TLCO z-score	**0·304, p<0·001**	**0·185, P=0·001**	0·119, P=0·100
Va z-score	0·080, p=0·249	**0·121, P=0·025**	-0·050, P=0·487
KCO z-score	**0·247, p<0·001**	0·089, P=0·103	**0·171, P=0·018**
BMI	0·002, p=0·978	**-0·128, P=0·025**	**0·185, P=0·014**
Dyspnoea-12	-0·076, p=0·372	-0·050, P=0·455	-0·063, P=0·488

Table S 1: Table to show the relationship between the RBC:Gas, RBC:M and M:Gas with lung function measurements, BMI and the Dyspnoea-12 questionnaire. Correlations estimated from a regression adjusted for age and sex. FEV_1 - Forced Expired Volume in one second; FVC - Forced Vital Capacity; TLCO - Transfer factor for Lung carbon monoxide; KCO - carbon monoxide transfer coefficient; Va - Alveolar Volume, BMI - Body mass index.

RBC:Gas z-score	Controls (N=44)	Breathless Long COVID (N=109)	Non-breathless Long COVID(N=43)	p-value*
Lower z-score	-0·32 (0·90) [-1·88 to 1·98]	-0·48 (0·92) [-2·38 to 2·88]	-0·27 (1·07) [-1·89 to 2·93]	0·40
Mid z-score	-0·71 (0·81) [-2·30 to 0·83]	-0·79 (0·92) [-2·66 to 2·07]	-0·59 (1·10) [-2·43 to 2·62]	0·52
Upper z-score	-0·63 (1·05) [-2·92 to 2·04]	-0·72 (1·08) [-3·36 to 2·58]	-0·47 (1·33) [-2·81 to 3·05]	0·47
Anterior z-score	-0·54 (0·74) [-1·90 to 1·56]	-0·59 (0·76) [-2·14 to 2·10]	-0·48 (0·90) [-2·02 to 1·84]	0·72
Posterior z-score	-0·60 (0·97) [-2·45 to 1·30]	-0·74 (1·01) [-3·16 to 2·42]	-0·48 (1·28) [-2·58 to 2·87]	0·38
Missing, n(%)	1(2·3%)	0 (0·0%)	0 (0·0%)	
RBC:M z-score				
Lower z-score	0·34 (0·08) [0·22 to 0·53]	0·34 (0·08) [0·16 to 0·59]	0·35 (0·10) [0·16 to 0·67]	0·81
Mid z-score	0·35 (0·08) [0·22 to 0·50]	0·35 (0·09) [0·16 to 0·60]	0·35 (0·11) [0·18 to 0·72]	0·94
Upper z-score	0·37 (0·08) [0·22 to 0·58]	0·37 (0·09) [0·18 to 0·66]	0·38 (0·11) [0·20 to 0·78]	0·83
Anterior z-score	0·35 (0·08) [0·20 to 0·52]	0·35 (0·09) [0·17 to 0·60]	0·35 (0·10) [0·19 to 0·67]	0·99
Posterior z-score	0·36 (0·08) [0·25 to 0·54]	0·35 (0·09) [0·16 to 0·63]	0·36 (0·10) [0·18 to 0·75]	0·84
Missing, n(%)	1 (2·3%)	0 (0·0%)	0 (0·0%)	
M:Gas z-score				
Lower z-score	-0·21 (0·86) [-1·43 to 1·47]	-0·30 (1·02) [-2·34 to 3·68]	0·04 (1·17) [-1·90 to 3·32]	0·18
Mid z-score	-0·31 (1·09) [-1·83 to 2·91]	-0·42 (1·30) [-3·20 to 4·68]	0·08 (1·63) [-2·55 to 5·65]	0·12
Upper z-score	-0·27 (0·98)	-0·35 (1·05)	0·02 (1·35)	0·18

	[-1·63 to 3·56]	[-2·52 to 3·25]	[-2·12 to 4·90]	
Anterior z-score	-0·20 (0·94)	-0·31 (1·02)	0·06 (1·34)	0·16
	[-1·51 to 3·43]	[-2·48 to 4·30]	[-1·91 to 4·92]	
Posterior z-score	-0·28 (0·85)	-0·35 (1·03)	0·05 (1·28)	0·10
	[-1·49 to 1·87]	[-2·49 to 3·13]	[-2·04 to 4·03]	
Missing, n(%)	1 (2·3%)	0 (0·0%)	0 (0·0%)	

Table S 2: Regional RBC:Gas, RBC:M and M:Gas z-scores in controls, breathless Long COVID and non-breathless Long COVID. RBC:Gas, RBC:M and M:Gas z-scores examined separately in the lower, mid and upper lung zones and anterior and posterior portions of lung. Summaries presented as mean, standard deviation and minimum and maximum. *P-values derived from ANOVA.

Characteristics of breathless patients with TLCO z-score >-1.64 and <-1.64

Outcome	BLC with TLCO >-1·64 (N=94)	BLC with TLCO <-1·64 (N=14)	p-value
FVC z-score	0·1 (0·9) [-2·1 to 2·6]	-0·3 (1·0) [-1·9 to 1·9]	0·20
FEV_1 z-score	-0·0 (0·9) [-2·7 to 1·9]	-0·3 (1·0) [-1·9 to 1·8]	0·33
KCO z-score	0·2 (0·8) [-1·5 to 2·3]	-1·1 (1·1) [-3·7 to 1·4]	**<0·001**
Va z-score	-0·8 (1·0) [-2·9 to 1·7]	-1·4 (1·0) [-3·5 to 0·6]	**0·02**
BMI (kg/m^2)	28·3 (6·0) [17·2 to 49·5]	30·1 (7·9) [16·4 to 40·9]	0·31
Smoking history, n (%)	23 (24·5%)	6 (42·9%)	p(exact) = 0·20
Dyspnoea-12	14 (8) [1 to 36]	22 (10) [2 to 36]	**0·001**
Missing, n (%)	17 (18·1%)	3 (21·4%)	
FACIT F	22 (10) [2 to 52]	18 (18) [7 to 47]	0·34
Missing, n (%)	17 (18·1%)	3 (21·4%)	
GAD-7	6 (6) [0 to 2]	10 (6) [0 to 19]	**0·04**
Missing, n (%)	8 (8·5%)	2 (14·3%)	
Baseline mBORG	1·2 (1·2) [0·0 to 6·0]	1·8 (1·5) [0·0 to 4·0]	0·11
Missing, n (%)	1 (1·1%)	0 (0·0%)	
mBORG post STS test	4·4 (1·9) [1·0 to 10·0]	5·5 (2·8) [0·5 to 10·0]	0·08
Missing, n (%)	2 (2·1%)	0 (0·0%)	

Number STS movements patient completed	27·6 (10·3) [9·0 to 71·0]	22·1 (7·0) [11·0 to 40·0]	0·06
Missing, n (%)	2 (2·1%)	0 (0·0%)	
RBC:Gas z-score	-0·7 (1·0) [-3·3 to 2·4]	-0·8 (1·2) [-2·9 to 2·0]	**0·01**
RBC:M z-score	-0·9 (1·2) [-3·9 to 2·5]	-1·0 (1·2) [-3·9 to 2·5]	**0·049**
M:Gas z-score	-0·3 (1·0) [-2·6 to 2·2]	-0·7 (1·5) [-2·2 to 3·7]	0·16

Table S 3: A comparison of the characteristics of BLC with a TLCO z-score below-1·64 and above -1·64 including lung function values, questionnaires, exercise data, and ^{129}Xe-MRI findings. NBLC and controls not analysed due to low numbers (n=3 and 2, respectively). Summaries presented as mean, standard deviation and minimum and maximum. P-values derived from ANOVA for between group comparisons. TLCO - Transfer factor for Lung carbon monoxide; FVC - Forced Vital Capacity; FEV_1 - Forced Expired Volume in one second; KCO - carbon monoxide transfer coefficient; Va – Alveolar Volume; BMI – Body mass index; FACIT-F – Functional Assessment of Chronic Illness Therapy – Fatigue; GAD-7 – Generalised Anxiety Disorder-7; STS – 1-minute Sit-to-stand test.

Outcome	*Control subjects (N=44)*	*Breathless Long Covid (N=109)*	*Non-breathless Long Covid (N=43)*	*P-value†*
Number STS repetitions	39·2 (13·5) [20 to 68]	26·8 (10·1) [9 to 71]	31·3 (12·2) [8 to 59]	**<0·001**
STS Percentile (n^{th})	25 (2·5 to 56·3) [2·5 to 97·5]	2·5 (2·5 to 2·5) [2·5 to 97·5]	2·5 (2·5 to 25·0) [2·5 to 75·0]	**<0·001**
mBORG pre-STS test	0 (0 to 0) [0 to 2]	1 (0·2 to 2) [0 to 6]	0 (0 to 0·5) [0 to 2]	**<0·001**
mBORG post-STS test‡	3 (2 to 3) [0 to 9]	4 (3 to 5·5) [0·5 to 10]	3 (2 to 4) [0·5 to 7]	**<0·001**
Oxygen saturation pre-STS test (%)	98 (1·4) [94 to 100]	97·8 (1·7) [92 to 100]	97·5 (1·7) [94 to 100]	0·420
Oxygen saturation post-STS test (%)	97·4 (1·6) [93 to 100]	97·3 (2·1) [85 to 100]	97·8 (1·3) [95 to 100]	0·247
Heart rate pre-STS test (bpm)	75·4 (13·5) [52 to 112]	78·4 (13·4) [50 to 120]	74·7 (12·2) [54 to 104]	0·207
Heart rate post-STS test (bpm)	100·5 (85·5 to 116) [64 to 152]	101 (91 to 113) [55 to 160]	90 (81 to 115) [56 to 171]	0·368
Missing, n(%)	0 (0%)	2 (1.8%)	0 (0%)	

‡Median and interquartile range presented for non-normally distributed continuous variables
†ANOVA used for normally distributed variables and Kruskal-Wallis for non-normally distributed variables.

Table S 4: Summary of number of repetitions completed during the 1-minute sit-to-stand test (STS) across groups, plus measures of breathlessness (mBORG), oxygen saturations, and heart rate pre-and post the STS test. Summaries presented

as mean, standard deviation and minimum and maximum except for non-normally distributed continuous variables‡. mBORG – modified BORG dyspnoea score; bpm – beats per minute.

	Control subjects (N=27)	*Breathless Long COVID (N=66)*	*Non-breathless Long COVID (N=28)*	*P-value**
Age (years)	48·2 (14·8) [27·4 to 72·6]	50·0 (12·6) [23·3 to 81·4]	46·8 (10·0) [28·5 to 63·7]	0·510
BMI (kg/m²)	29·9 (6·3) [19·4 to 45·5]	29·9 (6·6) [16·4 to 49·5]	27·5 (4·6) [20·8 to 37·1]	0·209
Sex, n (%)				p = 0·696
Female	20 (74·1%)	43 (65·2%)	19 (67·9%)	
Male	7 (25·9%)	23 (34·8%)	9 (32·1%)	
Time from infection (days)‡	418 (206 to 628) [110 to 1319]	574 (347 to 900) [125 to 1489]	298 (195 to 518) [141 to 1064]	**0·004**
Smoking, n (%)				p(exact) = 0·265
Former	6 (22·2%)	20 (30·3%)	4 (14·3%)	
FEV_1 z-score	0·27 (0·79) [-1·21 to 1·67]	-0·04 (0·94) [-2·66 to 1·83]	0·43 (1·04) [-1·28 to 4·02]	0·063
FVC z-score	0·28 (0·95) [-1·48 to 2·86]	-0·02 (0·96) [-2·10 to 2·36]	0·36 (0·94) [-1·20 to 3·26]	0·146
TLCO z-score	-0·31 (0·90) [-2·93 to 1·39]	-0·99 (0·83) [-3·33 to 0·41]	-0·47 (0·88) [-2·09 to 1·23]	**<0·001**
KCO z-score	0.51 (1·09) [-1·16 to 3·10]	-0·17 (1·05) [-3·68 to 1·51]	0·07 (0·98) [-1·51 to 2·46]	**0·020**
Va z-score	-0·96 (0·93) [-2·78 to 1·26]	-1·03 (0·94) [-3·46 to 1·36]	-0·69 (0·94) [-3·25 to 1·51]	0·270
CT score	0·0 (0·0 to 0·0) [0·0 to 3·0]	0·0 (0·0 to 1·0) [0·0 to 4·0]	0·0 (0·0 to 0·0) [0·0 to 2·0]	0·134

RBC:M z-score	-0·67 (0·90) [-2·01 to 1·61]	-0·75 (1·18) [-3·58 to 2·63]	-0·62 (1·24) [-3·21 to 2·68]	0·866
RBC:Gas z-score	-0·84 (1·14) [-2·41 to 1·52]	-1·08 (1·14) [-3·28 to 2·37]	-0·86 (1·27) [-2·53 to 2·68]	0·558
M:Gas z-score	-0·28 (1·06) [-1·52 to 2·38]	-0·49 (1·15) [-2·61 to 3·65]	-0·24 (1·32) [-2·11 to 4·48]	0·552
Missing, n (%) for ^{129}Xe-MRI	1 (3·7%)	0 (0·0%)	0 (0%)	
Dyspnoea-12‡	NA	14 (11 to 20) [1 to 36]	3 (0 to 6) [0 to 14]	**<0·001**
Missing, n (%)	NA	8 (12·1%)	3 (10·7%)	

‡ Median, interquartile range and minimum and maximum presented for non-normally distributed continuous variables
†ANOVA used for normally distributed variables and Kruskal-Wallis for non-normally distributed variables.

Table S 5: Demographics, pulmonary function tests, Dyspnoea-12, CT, and global ^{129}Xe-MRI (using 3T scanner) measurements across groups in the Oxford and Cardiff cohort (all study procedures undertaken in Oxford). Summaries presented as mean, standard deviation and minimum and maximum except for non-normally distributed continuous variables‡. BMI – Body mass index; FEV_1 - Forced Expired Volume in one second; FVC - Forced Vital Capacity; TLCO - Transfer factor for Lung carbon monoxide; KCO - carbon monoxide transfer coefficient; Va – Alveolar Volume; CT – Computed Tomography.

	Control subjects (N=17)	*Breathless Long COVID (N=43)*	*Non-breathless Long COVID (N=15)*	*P-value**
Age (years)	41·8 (10·2) [23·3 to 60·5]	42·2 (11·7) [18·8 to 69·1]	44·3 (12·4) [22·1 to 64·1]	0·802
BMI (kg/m²)	26·2 (4·2) [19·1 to 34·4]	26·3 (5·3) [16·6 to 41·5]	28·2 (6·1) [21·1 to 40·0]	0·433
Sex, n (%)				**p(exact) = 0·042**
Female	14 (82·4%)	28 (65·1%)	6 (40·0%)	
Male	3 (17·6%)	15 (34·9%)	9 (60·0%)	
Time from infection (days)‡	183 (172 to 237) [130 to 472]	723 (442 to 910) [189 to 1144]	645 (400 to 892) [235 to 1336]	**<0·001**
Missing, n (%)	0 (0·0%)	2 (4·2%)	0 (0·0%)	
Smoking, n (%)				p(exact) = 0·451
Former	4 (23·5%)	8 (18·6%)	5 (33·3%)	
FEV_1 z-score	0·34 (0·85) [-0·85 to 2·11]	-0·04 (0·86) [-2·39 to 1·89]	0·20 (0·91) [-1·04 to 2·60]	0·297
FVC z-score	0·63 (0·94) [-1·04 to 2·36]	0·14 (0·97) [-1·82 to 2·58]	0·35 (0·92) [-0·67 to 0·80]	0·203
TLCO z-score	0·49 (0·86) [-1·75 to 1·72]	-0·12 (0·81) [-1·50 to 1·53]	0·18 (0·80) [-0·67 to 1·84]	**0·038**
KCO z-score	0·33 (1·00) [-0·90 to 3·29]	0·32 (0·76) [-1·04 to 2·32]	0·42 (0·75) [-0·83 to 1·45]	0·925
Va z-score	0·18 (0·78) [-1·58 to 1·30]	-0·53 (1·02) [-2·80 to 1·70]	-0·27 (0·75) [-1·27 to 1·27]	**0·031**
Missing, n (%) for all PFTs	0 (0·0%)	1 (2·3%)	1 (6·7%)	
CT score	0·0 (0·0 to 0·0) [0·0 to 0·0]	0·0 (0·0 to 0·0) [0·0 to 1·0]	0·0 (0·0 to 0·0) [0·0 to 0·0]	0·683
Missing, n (%)	0 (0·0%)	1 (2·1%)	0 (0%)	
RBC:M z-score	-0·27 (1·12) [-1·95 to 2·15]	-0·41 (1·14) [-2·34 to 2·52]	-0·75 (1·43) [-3·47 to 2·88]	0·511
RBC:Gas z-score	-0·30 (0·64) [-1·16 to 1·17]	-0·36 (0·72) [-2·18 to 1·31]	0·03 (1·20) [-1·85 to 1·73]	0·291
M:Gas z-score	-0·23 (0·63) [-1·46 to 0·67]	-0·10 (0·92) [-1·79 to 1·90]	0·59 (1·25) [-1·73 to 2·79]	0·030
Dyspnoea-12‡	NA	14 (8 to 19) [1 to 28]	4 (0 to 5) [0 to 18]	**0·002**
Missing, n (%)	NA	19 (44·2%)	2 (13·3%)	

‡ Median, interquartile range and minimum and maximum presented for non-normally distributed continuous variables
†ANOVA used for normally distributed variables and Kruskal-Wallis for non-normally distributed variables.

Table S 6: Demographics, pulmonary function tests, Dyspnoea-12, CT, and global ^{129}Xe-MRI (using 1.5T scanner) measurements across groups in the Sheffield and Manchester cohort (all study procedures undertaken in Sheffield). Summaries presented as mean, standard deviation and minimum and maximum except for non-normally distributed continuous variables‡. BMI – Body mass index; FEV_1 - Forced Expired Volume in one second; FVC - Forced Vital Capacity; TLCO - Transfer factor for Lung carbon monoxide; KCO - carbon monoxide transfer coefficient; Va – Alveolar Volume; PFTs – Pulmonary function tests; CT – Computed Tomography.

References

1 Francone M, Iafrate F, Masci GM, et al. Chest CT score in COVID-19 patients: correlation with disease severity and short-term prognosis. *Eur Radiol* 2020; 30(12): 6808–6817.

2 Strassmann A, Steurer-Stey C, Lana KD, et al. Population-based reference values for the 1-min sit-to-stand test. *Int J Public Health* 2013; 58(6): 949–953.

3 Chen M, Doganay O, Matin T, et al. Delayed ventilation assessment using fast dynamic hyperpolarised Xenon-129 magnetic resonance imaging. *Eur Radiol* 2020; 30(2): 1145–1155.

4 Norquay G, Collier GJ, Rodgers OI, et al. Standalone portable xenon-129 hyperpolariser for multicentre clinical magnetic resonance imaging of the lungs. *Br J Radiol* 2022; 95(1132): 20210872.

5 Saunders LC, Collier GJ, Chan HF, et al. Longitudinal Lung Function Assessment of Patients Hospitalized With COVID-19 Using (1)H and (129)Xe Lung MRI. *Chest* 2023; 164(3): 700–716.

6 Smith LJ, Horsley A, Bray J, et al. The assessment of short- and long-term changes in lung function in cystic fibrosis using ^{129}Xe MRI. *Eur Respir J* 2020; 56(6): 2000441.

7 Collier GJ, Eaden JA, Hughes PJ, et al. Dissolved 129Xe lung MRI with four-echo 3D radial spectroscopic imaging: quantification of regional gas transfer in idiopathic pulmonary fibrosis. *Magn Reson Med* 2021; 85(5): 2622–2633.

8 Collier GJ, Smith LJ, Saunders LC, et al. Age, sex, and lung volume dependence of dissolved xenon-129 MRI gas exchange metrics. *Magn Reson Med* 2024; 92(4): 1471–1483.