EUROPEAN RESPIRATORY *journal*

FLAGSHIP SCIENTIFIC JOURNAL OF ERS

Early View

Original research article

Cluster Analysis to Identify Long COVID Phenotypes Using ^{129}Xe Magnetic Resonance Imaging: A Multi-centre Evaluation

Rachel L Eddy, David Mummy, Shuo Zhang, Haoran Dai, Aryil Bechtel, Alexandra Schmidt, Bradie Frizzell, Firoozeh V Gerayeli, Jonathon A Leipsic, Janice M Leung, Bastiaan Driehuys, Loretta G Que, Mario Castro, Don D Sin, Peter J Niedbalski

Please cite this article as: Eddy RL, Mummy D, Zhang S, *et al*. Cluster Analysis to Identify Long COVID Phenotypes Using ^{129}Xe Magnetic Resonance Imaging: A Multi-centre Evaluation. *Eur Respir J* 2024; in press (https://doi.org/10.1183/13993003.02301-2023).

Cluster Analysis to Identify Long COVID Phenotypes Using ^{129}Xe Magnetic Resonance Imaging: A Multi-centre Evaluation

Rachel L Eddy[1,2], David Mummy[4], Shuo Zhang[4], Haoran Dai[5], Aryil Bechtel[4], Alexandra Schmidt[1], Bradie Frizzell[8], Firoozeh V Gerayeli[1], Jonathon A Leipsic[1,3], Janice M Leung[1,2], Bastiaan Driehuys[4,5,6], Loretta G Que[7], Mario Castro[8], Don D Sin[1,2], Peter J Niedbalski[8]

[1]Centre for Heart Lung Innovation, St. Paul's Hospital; [2]Division of Respiratory Medicine, Department of Medicine; [3]Department of Radiology, University of British Columbia, Vancouver, Canada
[4]Department of Radiology, [5]Department of Medical Physics, [6]Department of Biomedical Engineering, [7]Division of Pulmonary, Department of Medicine, Duke University, Durham, NC, USA
[8]Division of Pulmonary and Critical Care Medicine, University of Kansas Medical Center, Kansas City, Kansas, USA

Correspondence to:
PJ Niedbalski PhD
University of Kansas Medical Center
3901 Rainbow Blvd, MS 3007
Kansas City, KS USA, 66160
Email: pniedbalski@kumc.edu

Original Research

Body Word Count: 3663 words
Reference Count: 40
Figures and Tables Count: 7

Take Home Message (256/256 characters)
Cluster analysis of ^{129}Xe MRI metrics identifies four phenotypes of long COVID with distinct functional MRI and clinical characteristics. MRI-based clusters can be used to dissect long COVID heterogeneity, enabling personalised clinical care and treatment.

ABSTRACT (250/250 words)

Background: Long COVID impacts ~10% of people diagnosed with COVID-19, yet the pathophysiology driving ongoing symptoms is poorly understood. We hypothesised that ^{129}Xe magnetic resonance imaging (MRI) could identify unique pulmonary phenotypic subgroups of long COVID, therefore we evaluated ventilation and gas exchange measurements with cluster analysis to generate imaging-based phenotypes.

Methods: COVID-negative controls and participants who previously tested positive for COVID-19 underwent ^{129}XeMRI ~14-months post-acute infection across three centres. Long COVID was defined as persistent dyspnea, chest tightness, cough, fatigue, nausea and/or loss of taste/smell at MRI; participants reporting no symptoms were considered fully-recovered. ^{129}XeMRI ventilation defect percent (VDP) and membrane (Mem)/Gas, red blood cell (RBC)/Mem and RBC/Gas ratios were used in k-means clustering for long COVID, and measurements were compared using ANOVA with post-hoc Bonferroni correction.

Results: We evaluated 135 participants across three centres: 28 COVID-negative (40±16yrs), 34 fully-recovered (42±14yrs) and 73 long COVID (49±13yrs). RBC/Mem ($p=0.03$) and FEV_1 ($p=0.04$) were different between long- and COVID-negative; FEV_1 and all other pulmonary function tests (PFTs) were within normal ranges. Four unique long COVID clusters were identified compared with recovered and COVID-negative. Cluster1 was the youngest with normal MRI and mild gas-trapping; Cluster2 was the oldest, characterised by reduced RBC/Mem but normal PFTs; Cluster3 had mildly increased Mem/Gas with normal PFTs; and Cluster4 had markedly increased Mem/Gas with concomitant reduction in RBC/Mem and restrictive PFT pattern.

Conclusion: We identified four ^{129}XeMRI long COVID phenotypes with distinct characteristics. ^{129}XeMRI can dissect pathophysiologic heterogeneity of long COVID to enable personalised patient care.

Key Words: Long COVID, ^{129}Xe MRI, cluster analysis, imaging phenotypes

INTRODUCTION

As the world transitions to a post-pandemic stage of coronavirus disease 2019 (COVID-19), clinical and research focus have shifted from the acute to chronic effects. Long COVID (also known as post-COVID syndrome or post-acute sequelae of COVID) is broadly accepted as signs, symptoms and/or conditions that continue or develop >4-12 weeks following acute infection [1-3]. Long COVID has been estimated to affect approximately 10% of the >770M individuals who experienced SARS-CoV-2 infection [4-7] and may persist two years or longer [8]. While it is now clear that long COVID is a multi-organ condition with a wide range of symptoms frequently involving the respiratory system [6,9], pulmonary function tests (PFTs) and conventional computed tomography (CT) are typically normal. Since the physiological drivers of persistent symptoms in these people are not fully understood, it remains difficult to manage long COVID patient's varied symptoms and to identify effective therapies.

One important challenge in identifying novel therapies in long COVID is its heterogeneity. Imaging has the potential to reveal unique phenotypes of long COVID. Functional evaluation using expiratory CT and hyperpolarised ^{129}Xe magnetic resonance imaging (MRI) have revealed air trapping [10] and ventilation abnormalities [11,12] in people reporting long COVID, suggesting the importance of small airway dysfunction in the pathophysiology. ^{129}Xe MRI further enables measurement of regional gas exchange as inhaled ^{129}Xe gas uptake in the alveolar membrane and transfer to capillary red blood cells (RBC), and in long COVID has characteristically shown reduced ratio of ^{129}Xe RBC-to-membrane [13-17]. These findings have been postulated to reflect microvascular abnormalities [13,16,18] and have been observed in people with dyspnea [13-15] or any other persistent symptoms [16,17], those who have been hospitalised [13-17] or not [16,17], for one-year or longer after acute infection [15,17], and importantly when conventional CT and PFTs have been mostly normal.

Given the broad range of long COVID presentation, there have been efforts to define phenotypes or clusters using symptoms or clinical factors [19-21]. Such phenotypic clustering has also been pursued in asthma and COPD using CT and ^{129}Xe MRI metrics to identify distinct lung structure-function and clinical characteristics that were not distinguishable using clinical evaluations alone [22-24]. We therefore hypothesised that ^{129}Xe MRI could also identify unique pulmonary phenotypic subgroups of long COVID. Thus, in this study we evaluated MRI gas exchange and ventilation measurements in participants with long COVID across three centres and employed a cluster analysis to generate imaging-based phenotypes of long COVID, and to determine the relation of these phenotypes to clinical endpoints such as symptoms. Preliminary results have been reported in an abstract form [25].

MATERIALS AND METHODS

Study Participants and Design

Participants aged ≥19-years who self-reported a previous positive test for COVID-19 by a polymerase chain reaction or lateral flow test provided written informed-consent to ethics board-approved protocols between July 2021 and February 2023 across three academic centres: University of British Columbia (Vancouver, Canada; Site1), University of Kansas Medical Center (Kansas City, USA; Site2), and Duke University (Durham, USA; Site3). At all centres participants were recruited through referral from healthcare providers or community advertising, and study assessments were completed at least three months after the COVID test date. Participants were not recruited by any specific symptoms, and participants reporting no symptoms (ie, no long COVID) were also recruited (Site1, Site2). Participants with pre-existing respiratory disease were not excluded. COVID-negative healthy controls were prospectively recruited (Site1) or retrospectively included from pre-existing datasets (Site2, Site3).

All participants underwent hyperpolarised ^{129}Xe gas exchange MRI, while PFT measurements and chest CT performed in the respective clinical laboratories/departments were included where available. Board-certified radiologist reports for inspiratory CT performed within six-months of MRI were reviewed for the presence of ground glass opacities, reticulation, honeycombing, consolidation and emphysema. The following self-reported (new or worsened) persistent symptoms for >3 months post-COVID (yes/no) were recorded across all sites at the time of MRI: dyspnea, chest tightness, cough, fatigue, nausea and loss of taste or smell. Long COVID was defined when at least one of these symptoms was present, new/worse post-COVID, and not explained by another cause as adjudicated by site investigators. Participants reporting no symptoms were considered fully recovered. The St. George's Respiratory Questionnaire (SGRQ) was also administered at Site1 to measure respiratory-related quality of life [26].

Pulmonary Function Tests

Spirometry was performed according to ATS/ERS guidelines [27] to determine the forced expiratory volume in one-second (FEV_1) and forced vital capacity (FVC), and plethysmography [28] was performed to measure residual volume (RV) and total lung capacity (TLC). The diffusion capacity for carbon monoxide (DL_{CO}) was measured using the single-breath method [29]. All values were reported using race-neutral Global Lung Function Initiative reference equations [30].

MRI Acquisition and Analysis

MRI was performed using 3T MAGNETOM systems (Site 1, Vida; Site 2, Skyra; Site 3; PrismaFIT; Siemens Healthineers, Erlangen, Germany) and a flexible ^{129}Xe vest coil (Clinical MR Solutions, Brookfield, USA). Up to 1.0L of isotopically enriched (≥85%) ^{129}Xe gas was polarised using 9820 hyperpolarisers (Polarean Imaging Inc., Durham, USA). Total inhaled doses were 1.0L (Site1) or 20% of the participants' FVC up to a maximum of 1.25L (Site2,

Site3), and participants were instructed to inhale the gas mixture from functional residual capacity for imaging under breath-hold conditions. Gas exchange imaging was performed using single-breath 1-point Dixon methods (Site1 [31]; Site2 [32]; Site3 [33]) to generate simultaneous ^{129}Xe gas and dissolved phase images (xenon dissolved in interstitial membrane [Mem] and red blood cells [RBC] within the pulmonary capillaries). Minor variations in flip angle, repetition time (TR) and sampling strategy were used in the 1-point Dixon methods across sites. The modified methods primarily reduced acquisition time [32,33] and have been validated against the gold-standard [31] to maintain quantitative gas exchange measurements. Importantly, the equivalent repetition time for a hypothetical 90° flip angle ($TR_{90,equiv}$) which determines gas exchange image contrast was 249ms across all sites, enabling meaningful aggregation of the multi-centre data. Detailed MRI protocols are described in the online supplement.

Quantitative MRI analysis was performed centrally at Site3 as previously described [34]. Ventilation defect percent (VDP) was quantified as the percentage of lung with absent ^{129}Xe gas signal, reflecting airway obstruction and flow limitation. Gas exchange was quantified as whole-lung ratios of Mem/Gas, RBC/Mem and RBC/Gas. Mem/Gas, or membrane tissue uptake, measures ^{129}Xe dissolved in the interstitial membrane normalised to the gas signal and reflects parenchymal integrity; high membrane uptake is typically due to fibrosis or inflammation whereas low membrane uptake is typically due to tissue destruction or emphysema. RBC/Gas, or RBC transfer, measures ^{129}Xe dissolved in pulmonary capillary RBCs normalised to the gas signal and reflects microvascular integrity; low RBC/Gas is typically due to diffusion/perfusion limitation or vascular destruction. RBC/Mem measures the ratio of ^{129}Xe dissolved in RBCs to that in interstitial tissue and represents the efficiency of transfer of gas from the interstitium to capillary RBCs as a measure of overall gas exchange; reduced RBC/Mem could be due to reduced RBC transfer and/or increased membrane uptake.

RBC/Mem has been most widely applied to date to evaluate long COVID [13-17]. MRI measurements were not adjusted for hemoglobin, consistent with prior work in this area. VDP, Mem/Gas, RBC/Mem and RBC/Gas were included for clustering to represent the range of pulmonary pathophysiology including airway, parenchymal and microvascular measurements from a single ^{129}Xe scan.

Statistical Analysis

Clustering was performed for long COVID participants using MATLAB R2021a (Mathworks, Natick, USA). MRI variables were z-transformed to standardise scales and k-means clustering was performed for 2-5 clusters. Internal cluster validation to determine the optimal number of clusters was evaluated using the following parameters: Davies-Bouldin index, measuring the average similarity between clusters (minimised); silhouette width, measuring the average distance between clusters (maximised); and Dunn's index, measuring the ratio of smallest distance between observations in different clusters to the largest cluster diameter (maximised). All statistical analysis was performed using SPSS Statistics 25.0 (IBM Corporation, Armonk, USA). Data were tested for normality using Shapiro-Wilk tests and when not normally distributed, non-parametric tests were performed. Numerical variables were compared using ANOVA or Kruskal-Wallis tests with a post-hoc Bonferroni correction for multiple comparisons unless otherwise stated. Categorical variables were compared using a Fisher's Exact test. Relationships were evaluated using Spearman correlation coefficients. Results were considered statistically significant when $p<0.05$.

RESULTS

Figure 1 shows a CONSORT flow diagram of participant enrollment from each site; in total, we evaluated 135 participants across three centers, including 28 COVID-negative participants (40±16yrs, 16F), 34 fully recovered from COVID (42±14yrs, 19F) and 73 with long COVID (49±13yrs, 43F). Table 1 shows participant demographics, lung function and imaging

measurements by group. All groups were predominantly white Caucasians (>75%); 12% of the fully recovered and 25% of long COVID participants were hospitalised during the acute infection. Days between COVID diagnosis and MRI were not significantly different between the fully recovered and long COVID participants (mean±SD, 424±243 vs. 414±226; median [IQR], 399 [223,610] vs. 405 [217,565]; both p=0.8). Long COVID participants were significantly older than COVID-negative (p=0.01) with significantly higher SGRQ scores (p<0.001 vs. COVID-negative and recovered) and significantly lower ^{129}Xe RBC/Mem (p=0.03). FEV_1 was also significantly lower in long COVID (p=0.04) but was within normal limits; all other PFT measurements were also within normal limits and not significantly different. ^{129}Xe VDP, Mem/Gas, RBC/Mem and RBC/Gas plotted by group are shown in supplemental Figure E1. Tables E1-E3 in the online supplement show study measurements by site for each group.

Figure 2A shows that the optimal number of clusters was four, all of which contained participants from each site (supplemental Table E4). Figure 2B shows a radar plot of the primary MRI signatures associated with each of the four clusters. Qualitatively, VDP did not largely contribute to any group but was greatest in Cluster2. Mem/Gas was greatest in Cluster3 and Cluster4, and RBC/Mem and RBC/Gas were lowest in Cluster2 and Cluster4.

Table 2 provides a full listing of measurements, self-reported pre-existing comorbidities and persistent symptoms, and CT patterns by cluster. The days between COVID diagnosis and MRI were not significantly different across the clusters, nor were the proportion of participants by sex or SGRQ total, symptom and impact scores (p>0.1). SGRQ activity score was significantly different (p=0.02), with the smallest score in Cluster1 and greatest in Cluster4 (post-hoc p=0.02). The proportion of participants hospitalised during the acute infection phase was significantly different across clusters (p=0.03), with the greatest proportion (63%) and the longest hospital stays (mean 28 days) in Cluster4. Oxygen use (p=0.005), intensive care unit

admission (p=0.001), and use of ventilation (p<0.001) during hospitalisation were significantly different across the clusters, again with the greatest proportions in Cluster4 (80-100%). Dexamethasone and other steroids (60-100%) and remdesivir (antiviral, 14-67%) were the most common in-patient medical therapies across all clusters. Only the use of steroids was significantly different but borderline (p=0.045), with 100% of Cluster1 and Cluster3 patients prescribed steroids followed by 71% of Cluster2 and 60% of Cluster4. Full details for acute hospitalisation interventions are provided in supplemental Table E5. All ^{129}Xe MRI and most PFT measurements were significantly different across the clusters (p<0.05).

Up to 27% of all long COVID participants had at least one of the comorbidities: hypertension, diabetes, gastroesophageal reflux disease, obstructive sleep apnea, asthma or COPD. There were no significant differences in the prevalence of specific comorbidities across clusters, although the median number of comorbidities per participant was significantly different (p=0.04, supplemental Table E6), with Cluster3 and Cluster4 trending to the most comorbidities (p=0.055 and p=0.054, respectively vs Cluster1). For those participants with asthma and COPD, prescribed respiratory medications are listed in supplemental Table E7; most participants with asthma and COPD had mild disease based on international GINA and GOLD guidelines, respectively, except two participants with asthma who had been prescribed dupilumab (Cluster3) or prednisone (Cluster4). All had well-controlled disease prior to the initial SARS-CoV-2 infection. There were no differences in respiratory medications across the clusters.

Dyspnea (85%) and fatigue (77%) were the most commonly reported symptoms among all long COVID participants and were common in all clusters (63-100%). The occurrence of ongoing symptoms was not significantly different across clusters.

A subset of 56 long COVID participants underwent chest CT for post-COVID assessment (range 180-days before, 99-days after MRI). Presence of ground glass opacities, reticulation

and consolidation was significantly different across the clusters ($p<0.05$), with the most CT abnormalities being present in Cluster4. Of note, all eight participants in Cluster4 had evidence of ground glass opacities. Representative ^{129}Xe MR and CT images from each cluster as well as COVID-negative and recovered participants are shown in supplemental Figure E2. CT was acquired with varied protocols across sites thus further quantitative analysis could not be performed.

Figure 3 and Figure 4 show cluster-wise comparisons with COVID-negative and recovered groups for MRI and clinical measurements respectively, and full cluster-wise comparisons are provided in supplemental Tables E8-19. Of note, all clusters had a unique MRI feature compared with COVID-negative and fully recovered participants. SGRQ was different in all long COVID clusters, and trending greatest in Cluster4. FEV_1, FVC, DL_{CO}, RV, and TLC for most participants were above the lower limit of normal and were significantly different in Cluster4 only. As a proof-of-concept to explain Cluster1 pathophysiology, we tested RV/TLC for Cluster1 only, which was significantly greater than COVID-negative participants ($p=0.04$, independent samples t-test).

Altogether, the MRI and clinical features of each cluster are summarised in Figure 5 with representative ^{129}Xe images. Cluster1 exhibited 'normal' ^{129}Xe MRI metrics but mildly increased RV/TLC relative to COVID-negative participants, and was the youngest group with no inspiratory CT parenchymal abnormalities and normal spirometry. Cluster2 was characterised by low RBC/Mem only, and was the oldest group with mildly reduced DL_{CO}, no inspiratory CT parenchymal abnormalities and normal spirometry and lung volumes. Cluster3 was characterised by mildly elevated membrane uptake and preserved RBC/Mem, with minimal CT abnormalities and normal PFTs. Finally, Cluster4 was characterised by markedly elevated membrane uptake and corresponding low RBC/Mem, with high BMIs, CT abnormalities, and reduced flow rates on spirometry, DL_{CO} and lung volumes.

DISCUSSION

In this multi-centre study, we identified four long COVID clusters with unique ^{129}Xe MRI signatures and corresponding distinct demographic and clinical features. The key strength of this work is the large multi-centre population with harmonised ^{129}Xe imaging acquisition and analysis methods, thereby enabling cluster analysis to discover data-driven phenotypic subtypes of long COVID. Previously published ^{129}Xe MRI studies have revealed low RBC/Mem in people with post-acute or long COVID [13-17,35], consistent with Cluster2 phenotype in our work [15,17]. We further identified three other functional pulmonary long COVID groups. Cluster1 exhibited normal ^{129}Xe MRI with mildly increased RV/TLC and elevated SGRQ. Cluster3 and Cluster4 were uniquely driven by elevated ^{129}Xe membrane uptake that has not been previously noted in long COVID participants [13-17,35]. Prior reports have also shown abnormal VDP that was related to worse quality of life and exercise limitation in long COVID [11,12]; while VDP was greatest in Cluster2, the burden of ventilation abnormalities was low in all clusters. Thus, gas exchange abnormalities, and not airway or ventilation, were the greatest pathophysiologic drivers of long COVID in these participants. We additionally showed that the characteristic ^{129}Xe gas exchange metrics in Clusters2-4 were different from both COVID-negative controls and fully recovered participants.

The heterogeneity of long COVID has made it challenging to advance clinical understanding and develop novel therapies. Here, we used an emerging technology, ^{129}Xe MRI, to dissect the heterogeneity of long COVID and found that there are several clusters of physiologic processes that underpin the condition. There is a cluster of patients (Cluster1, ~30% of participants) who have mildly elevated RV/TLC ratio indicating air trapping despite 'normal' MRI and inspiratory CT. This group of long COVID patients may be similar to those from prior work that showed air trapping using expiratory CT [10]. For Cluster1, we posit that bronchodilators

could be effective in relieving gas trapping [12], though future studies are needed to validate this notion.

We were also able to separate participants with low RBC/Mem alone (Cluster2, ~30% of participants) from those with low RBC/Mem concomitant with elevated membrane uptake (Cluster4, ~10% of participants), and those with elevated membrane uptake but preserved RBC/Mem (Cluster3, ~30% of participants). Low RBC/Mem in Cluster2 in the absence of airway or parenchymal abnormalities adds to the growing evidence [13-17,35] of post-COVID microvascular abnormalities such as pulmonary capillary inflammation, vasoconstrictive remodeling or thrombosis that inhibit transfer of gas [18,37]. Cluster3 was only differentiated by mildly elevated Mem/Gas (with preserved RBC/Mem), which may represent (residual) interstitial inflammation or edema. Cluster2 and Cluster3 signatures demonstrate the high sensitivity of ^{129}Xe MRI in revealing abnormalities in specific components of pulmonary gas exchange pathophysiology. In contrast, DL_{CO} provides a whole-lung average of the alveolar epithelial-capillary structure, which may explain why participants in Cluster2 and Cluster3 had DL_{CO} values all within normal limits. These data also raise the possibility of evaluating specific anti-inflammatory therapies targeted at the alveolar-capillary units as well as non-pharmacologic interventions such as pulmonary rehabilitation and exercise for patients in these clusters to enhance gas exchange and reduce symptoms.

Cluster4 had considerably elevated ^{129}Xe membrane uptake with concomitant low RBC/Mem and gas transfer, restrictive pattern on PFT, ground glass opacities on CT, and the highest SGRQ, indicating worse respiratory health status. Although small, Cluster4 was largely comprised of participants with the greatest severity of acute COVID-19 (based on hospitalisation and in-patient interventions), and were left with persistent post-COVID interstitial lung abnormalities [38]. It should be noted however that ~1/3 of all participants in Cluster4 did not require hospitalisation for acute COVID-19; thus, while severe COVID-19

pneumonia is a risk factor for Cluster4, patients who had milder COVID-19 infection can remain highly symptomatic 1-2 years following their initial infection. ^{129}Xe gas exchange MRI is known to be sensitive in detecting abnormalities in advanced interstitial lung disease [36,39,40]; however, the observed increase in ^{129}Xe membrane uptake and CT patterns in Cluster4 were not as extensive, suggesting early or mild stages of any post-COVID interstitial abnormalities. Cluster3 and Cluster4 long COVID participants could be targeted for monitoring to determine whether ^{129}Xe membrane uptake in this context is a treatable trait and whether these changes are permanent or transient. Long-term longitudinal studies will also be of value in determining whether there is a natural progression of clusters from 1 to 4 (or from 4 to 1) over time or whether each of these clusters are distinct phenotypes with different molecular drivers.

The impact of persistent symptoms was corroborated with elevated SGRQ across all clusters compared with COVID-negative and fully recovered post-COVID participants, however there were no significant differences in self-reported symptoms across the clusters. Our study was likely underpowered to draw such conclusions because of the symptoms which were recorded and could be pooled from all sites. For example, 'brain fog' has more recently been recognised as a common long COVID symptom [9], but neurocognitive symptoms were not recorded. However, respiratory symptoms relevant to ^{129}Xe MRI including dyspnea, chest tightness and cough were highly prevalent in these participants. We note that although asthma was the most commonly reported respiratory comorbidity, the overall burden of pre-existing respiratory disease was low and non-severe. Ongoing respiratory symptoms in all patients were adjudicated as new or worse post-COVID and thus mostly likely to be COVID-related. Furthermore, Cluster4 had the greatest SGRQ total score and significantly greater SGRQ activity score (vs Cluster1) indicating worse respiratory-related quality-of-life and activity

limitation related to dyspnea, and consistent with the most severe ^{129}Xe MRI changes in this group.

Limitations

Limitations of our study include the site-specific study designs and retrospective aggregation of multi-centre data, which limited comparisons of some clinical information. Importantly, MRI data quality and measurements were validated across sites to ensure data compatibility for clustering. This study was also cross-sectional where MRI was not performed at a pre-specified interval following SARS-CoV-2 infection at any site, and PFT and CT were not always performed contemporaneously with MRI. As a result, we cannot be certain on how measurements in each of the clusters or recovered group evolved over time, nor whether MRI differences might be related to different pulmonary effects of SARS-CoV-2 between participants. We note however that prior longitudinal studies showed negligible change in RBC/Mem between ~3-15-months post-infection [15,17]. CT was not performed with a consistent protocol at all sites, so quantification nor regional comparisons of CT patterns could not be performed. It is unknown how post-COVID MRI findings compare with acute lung injury caused by other infectious agents such as influenza. Peripheral blood was also not collected to help further characterise MRI measurements and phenotypes. However, a subset of participants from Site1 underwent bronchoscopy and sample of small airways; these results will be reported separately. Furthermore, like most long COVID research to date, pre-COVID measurements were not available for comparison in these participants.

We were also unable to confirm COVID variants to compare whether different variants contribute to differences across clusters. We posit based on the self-reported COVID testing dates that most infections were pre-omicron. Acute SARS-CoV-2 infection dates spanned timelines before vaccinations were available in the local regions and after up to two vaccine doses, therefore we did not evaluate the impact of vaccination on MRI abnormalities or

differences across the clusters. Finally, COVID antibody testing was not performed to verify self-reported SARS-CoV-2 infection status.

Conclusion

In conclusion, we demonstrate four unique long COVID phenotypes defined by ^{129}Xe MRI metrics, with distinct functional MRI and corresponding clinical characteristics. Such differences were otherwise not evident in the total long COVID group compared with COVID-negative and fully recovered post-COVID participants, especially using PFTs and CT. These four clusters reveal distinct (patho-) physiological changes that can be potential targets for therapy. Although a relatively new technology, ^{129}Xe MRI is now approved for clinical use in the United Kingdom and United States [31], providing new clinical opportunities for improved characterisation of respiratory diseases [36] and targeting of therapeutics for patients with long COVID. Future work can build from this foundation towards personalised medicine to evaluate the biological underpinnings of ^{129}Xe MRI signatures of long COVID clusters, and to determine their disease trajectories and responses to therapy.

FUNDING

This work was supported by: Canadian Institutes of Health Research (CIHR) and St. Paul's Foundation (Site1, University of British Columbia); Translating Duke Health – Keeping the Heart Young Pilot Award and NHLBI R01HL126771 (Site3, Duke University).

Table 1. Participant Demographic, Pulmonary Function and MRI Measurements.

Parameter ± SD**	COVID-negative (n=28)	Recovered (n=34)	Long COVID (n=73)	p-value*
Age years	40 ± 16	42 ± 14	49 ± 12†	**0.004**
Female n (%)	16 (57)	20 (59)	44 (60)	1.0
Caucasian n (%)	22 (79)	28 (82)	62 (85)	0.3
BMI kg/m^2	27 ± 4	27 ± 5	29 ± 6	0.2
COVID Hosp n (%)	-	4 (12)	18 (25)	0.2
Dx days n	-	424 ± 243	414 ± 226	0.8
SGRQ Total***	5 ± 5	7 ± 7	36 ± 20†‡	**<0.001**
SGRQ Symptom***	5 ± 7	11 ± 14	42 ± 22†‡	**<0.001**
SGRQ Activity***	10 ± 12	12 ± 11	51 ± 26†‡	**<0.001**
SGRQ Impact***	1 ± 3	3 ± 6	26 ± 21†‡	**<0.001**
FEV_1 $\%_{pred}$	105 ± 15	109 ± 16	99 ± 18‡	**0.04**
FVC $\%_{pred}$	107 ± 13	112 ± 15	102 ± 20	0.06
FEV_1/FVC %	81 ± 8	80 ± 7	79 ± 6	0.5
RV $\%_{pred}$	108 ± 25	110 ± 21	106 ± 30	0.8
TLC $\%_{pred}$	99 ± 13	102 ± 11	94 ± 18	0.09
RV/TLC $\%_{pred}$	107 ± 19	109 ± 19	112 ± 22	0.6
DL_{CO} $\%_{pred}$	103 ± 14	108 ± 10	99 ± 24	0.1
VDP %	0 ± 1	0 ± 1	1 ± 2	0.2
Mem/Gas x10^{-2}	0.71 ± 0.15	0.73 ± 0.18	0.77 ± 0.24	0.4
RBC/Mem	0.47 ± 0.13	0.43 ± 0.11	0.40 ± 0.12†	**0.04**
RBC/Gas x10^{-2}	0.35 ± 0.12	0.32 ± 0.09	0.31 ± 0.09	0.2

BMI=body mass index; COVID Hosp=hospitalised for COVID during acute infection; Dx days=days between COVID diagnosis and MRI; DL_{CO}=diffusion capacity of the lungs for carbon monoxide; FEV_1=forced expiratory volume in one second; FVC=forced vital capacity; Mem=interstitial membrane; RBC=red blood cells; RV=residual volume; SGRQ=St. George's Respiratory Questionnaire; TLC=total lung capacity; VDP=ventilation defect percent; $\%_{pred}$=percent predicted.

*Using one-way ANOVA for parametric variables, Kruskal-Wallis tests for non-parametric variables, or Fisher's exact test for categorial variables.

**Pulmonary function test measurements included where available.

COVID-negative: n=27 for FEV_1, FVC; n=15 for RV; n=20 for TLC; n=21 for DL_{CO}

Recovered: n=27 for all pulmonary function tests.

Long COVID: n=66 for FEV_1, FVC; n=57 for RV; n=61 for TLC; n=62 for DL_{CO}.

***SGRQ administered at Site1 only; n=11 COVID-negative, n=26 Recovered, n=36 Long COVID.

†Significantly different from COVID-negative ($p<0.05$)

‡Significantly different from Recovered ($p<0.05$)

Table 2. Long COVID MRI Cluster Summary.

Parameter ± SD**	**Cluster1 (n=24)**	**Cluster2 (n=22)**	**Cluster3 (n=19)**	**Cluster4 (n=8)**	**p-value***
Age years	43 ± 13	55 ± 10	49 ± 11	52 ± 13	**0.004**
Female n (%)	12 (50)	15 (68)	12 (63)	5 (63)	0.6
Caucasian n (%)	20 (83)	19 (86)	17 (89)	6 (75)	0.2
BMI kg/m^2	26 ± 4	28 ± 5	31 ± 8	33 ± 7	**0.02**
COVID Hosp n (%)	3 (13)	7 (32)	3 (16)	5 (63)	**0.03**
Dx days n	429 ± 222	446 ± 246	381 ± 191	358 ± 276	0.7
SGRQ Total***	31 ± 23	36 ± 14	37 ± 16	60 ± 13	0.2
SGRQ Symptom***	42 ± 24	39 ± 17	39 ± 26	59 ± 14	0.6
SGRQ Activity***	37 ± 29	57 ± 16	57 ± 17	84 ± 4	**0.02**
SGRQ Impact***	24 ± 24	22 ± 17	23 ± 16	46 ± 21	0.4
FEV_1 $\%_{pred}$	100 ± 15	101 ± 19	104 ± 18	82 ± 22	0.057
FVC $\%_{pred}$	107 ± 15	103 ± 20	105 ± 18	79 ± 23	**0.005**
FEV_1/FVC %	76 ± 6	79 ± 6	81 ± 6	85 ± 4	**0.005**
RV $\%_{pred}$	118 ± 24	104 ± 34	103 ± 24	71 ± 29	**0.006**
TLC $\%_{pred}$	100 ± 14	95 ± 19	95 ± 15	70 ± 20	**0.002**
RV/TLC $\%_{pred}$	118 ± 17	111 ± 30	110 ± 21	103 ± 24	0.4
DL_{CO} $\%_{pred}$	109 ± 18	93 ± 20	108 ± 18	61 ± 29	**<0.001**
VDP %	0 ± 0	2 ± 3	1 ± 1	1 ± 1	**0.004**
Mem/Gas ($x10^{-2}$)	0.62 ± 0.09	0.64 ± 0.14	0.92 ± 0.14	1.23 ± 0.19	**<0.001**
RBC/Mem	0.51 ± 0.07	0.33 ± 0.07	0.42 ± 0.08	0.22 ± 0.06	**<0.001**
RBC/Gas ($x10^{-2}$)	0.33 ± 0.06	0.22 ± 0.05	0.40 ± 0.07	0.29 ± 0.09	**<0.001**
Comorbidities n (%)					
Hypertension	2 (8)	7 (32)	6 (32)	4 (50)	0.052
Diabetes	0 (0)	1 (5)	4 (21)	2 (25)	0.06
GERD	3 (13)	9 (41)	4 (21)	1 (13)	0.1
OSA	4 (17)	4 (18)	6 (32)	2 (25)	0.7
Asthma	5 (21)	8 (36)	6 (32)	1 (13)	0.5
COPD	0 (0)	1 (5)	1 (5)	0 (0)	0.6
Persistent Symptoms n (%)					
Dyspnea	19 (79)	21 (95)	14 (74)	8 (100)	0.5
Chest Tightness	6 (25)	9 (41)	3 (16)	1 (13)	0.2
Cough	6 (25)	6 (27)	5 (26)	3 (38)	0.6
Fatigue	15 (63)	19 (86)	17 (89)	5 (63)	0.2
Nausea	2 (8)	3 (14)	3 (16)	0 (0)	0.8
Loss of Taste/Smell	9 (38)	5 (23)	2 (11)	2 (25)	0.4
CT Patterns n (%)[§]	(n=19)	(n=14)	(n=15)	(n=8)	
Ground Glass	0 (0)	1 (7)	1 (7)	8 (100)	**<0.001**
Reticulation	0 (0)	2 (14)	2 (13)	3 (38)	**0.046**
Honeycombing	0 (0)	0 (0)	0 (0)	1 (13)	0.1
Consolidation	0 (0)	0 (0)	0 (0)	2 (25)	**0.02**
Emphysema	1 (5)	1 (7)	0 (0)	0 (0)	0.8

BMI=body mass index; COVID Hosp=hospitalised for COVID during acute infection; CT=computed tomography; Dx days=days between COVID diagnosis and MRI; DL_{CO}=diffusion capacity of the lungs for carbon monoxide; FEV_1=forced expiratory volume in one second; FVC=forced vital capacity; Mem=interstitial membrane;

GERD=gastroesophageal reflux disease; OSA=obstructive sleep apnea; RBC=red blood cells; RV=residual volume SGRQ=St. George's Respiratory Questionnaire; TLC=total lung capacity; VDP=ventilation defect percent; $\%_{pred}$=percent predicted.
*P-values shown for long COVID cluster comparisons only. Using one-way ANOVA for parametric variables, Kruskal-Wallis tests for non-parametric variables, or Fisher's exact test for categorial variables.
**Pulmonary function test measurements included where available.
All long COVID: n=66 for FEV_1, FVC; n=57 for RV; n=61 for TLC; n=62 for DL_{CO}.
Cluster1: n=23 for FEV_1, FVC; n=21 for RV; n=22 for TLC; n=22 for DL_{CO}.
Cluster2: n=18 for FEV_1, FVC; n=15 for RV; n=17 for TLC; n=17 for DL_{CO}.
Cluster3: n=18 for FEV_1, FVC; n=15 for RV; n=16 for TLC; n=16 for DL_{CO}.
Cluster4: n=7 for FEV_1, FVC; n=6 for RV, TLC; n=7 for DL_{CO}.
***SGRQ administered at Site1 only; n=15 Cluster1, n=6 Cluster2, n=8 Cluster3, n=3 Cluster4.
§CT patterns identified as present or absent by qualitative radiologist review. Percentages shown as fraction of subset of patients with CT in each cluster.

REFERENCES

1. National Institute for Health and Care Excellence (NICE). COVID-19 rapid guideline: Managing the long-term effects of COVID-19. London; 2021.

2. US Centers for Disease Control and Prevention (CDC). Long COVID or post-COVID conditions. 2023 [accessed 2023 September 4]; Available from: https://www.cdc.gov/coronavirus/2019-ncov/long-term-effects/index.html

3. World Health Organisation (WHO). A clinical case definition of post COVID-19 condition by a Delphi consensus; 2021.

4. Wulf Hanson S, Abbafati C, Aerts JG, et al. Estimated global proportions of individuals with persistent fatigue, cognitive, and respiratory symptom clusters following symptomatic COVID-19 in 2020 and 2021. *JAMA* 2022: 328(16): 1604-1615.

5. Chen C, Haupert SR, Zimmermann L, et al. Global prevalence of post-coronavirus disease 2019 (COVID-19) condition or long COVID: A meta-analysis and systematic review. *J Infect Dis* 2022: 226(9): 1593-1607.

6. Davis HE, McCorkell L, Vogel JM, et al. Long COVID: Major findings, mechanisms and recommendations. *Nat Rev Microbiol* 2023: 21(3): 133-146.

7. Ford ND, Slaughter D, Edwards D, et al. Long COVID and significant activity limitation among adults, by age - united states, June 1-13, 2022, to June 7-19, 2023. *MMWR Morb Mortal Wkly Rep* 2023: 72(32): 866-870.

8. Bowe B, Xie Y, Al-Aly Z. Postacute sequelae of COVID-19 at 2 years. *Nat Med* 2023.

9. Perlis RH, Santillana M, Ognyanova K, et al. Prevalence and correlates of long COVID symptoms among US adults. *JAMA Netw Open* 2022: 5(10): e2238804.

10. Cho JL, Villacreses R, Nagpal P, et al. Quantitative chest CT assessment of small airways disease in post-acute sars-cov-2 infection. *Radiology* 2022: 304(1): 185-192.

11. Kooner HK, McIntosh MJ, Matheson AM, et al. (129)Xe MRI ventilation defects in ever-hospitalised and never-hospitalised people with post-acute COVID-19 syndrome. *BMJ Open Respir Res* 2022: 9(1): e001235.

12. Kooner HK, McIntosh MJ, Matheson AM, et al. Post-acute COVID-19 syndrome: (129)Xe MRI ventilation defects and respiratory outcomes one year later. *Radiology* 2023: 307(2): e222557.

13. 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.

14. Grist JT, Collier GJ, Walters H, et al. Lung abnormalities depicted with hyperpolarized xenon MRI in patients with long COVID. *Radiology* 2022: 305(5): 709-717.

15. Saunders LC, Collier GJ, Chan HF, et al. Longitudinal lung function assessment of patients hospitalised with COVID-19 using (1)H and (129)Xe lung MRI. *Chest* 2023.

16. Matheson AM, McIntosh MJ, Kooner HK, et al. Persistent (129)Xe MRI pulmonary and CT vascular abnormalities in symptomatic individuals with post-acute COVID-19 syndrome. *Radiology* 2022: 305(2): 466-476.

17. 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(4): 418-421.

18. Ackermann M, Verleden SE, Kuehnel M, et al. Pulmonary vascular endothelialitis, thrombosis, and angiogenesis in covid-19. *N Engl J Med* 2020: 383(2): 120-128.

19. Subramanian A, Nirantharakumar K, Hughes S, et al. Symptoms and risk factors for long COVID in non-hospitalized adults. *Nat Med* 2022: 28(8): 1706-1714.

20. Zhang H, Zang C, Xu Z, et al. Data-driven identification of post-acute sars-cov-2 infection subphenotypes. *Nat Med* 2023: 29(1): 226-235.

21. Danesh V, Arroliga AC, Bourgeois JA, et al. Symptom clusters seen in adult COVID-19 recovery clinic care seekers. *J Gen Intern Med* 2023: 38(2): 442-449.

22. Choi S, Hoffman EA, Wenzel SE, et al. Quantitative computed tomographic imaging-based clustering differentiates asthmatic subgroups with distinctive clinical phenotypes. *J Allergy Clin Immunol* 2017: 140(3): 690-700.e698.

23. Eddy RL, McIntosh MJ, Matheson AM, et al. Pulmonary MRI and cluster analysis help identify novel asthma phenotypes. *J Magn Reson Imaging* 2022: 56(5): 1475-1486.

24. Billatos E, Ash SY, Duan F, et al. Distinguishing smoking-related lung disease phenotypes via imaging and molecular features. *Chest* 2021: 159(2): 549-563.

25. Eddy RL, Mummy D, Dai H, et al. 129Xe magnetic resonance imaging-based phenotypes of long COVID: A multi-center evaluation. *Am J Respir Crit Care Med* 2023: 207: A6786.

26. Jones PW, Quirk FH, Baveystock CM. The St George's Respiratory Questionnaire. *Respir Med* 1991: 85 Suppl B: 25-31; discussion 33-27.

27. Miller MR, Hankinson J, Brusasco V, et al. Standardisation of spirometry. *Eur Respir J* 2005: 26(2): 319-338.

28. Wanger J, Clausen JL, Coates A, et al. Standardisation of the measurement of lung volumes. *Eur Respir J* 2005: 26(3): 511-522.

29. Graham BL, Brusasco V, Burgos F, et al. 2017 ERS/ATS standards for single-breath carbon monoxide uptake in the lung. *Eur Respir J* 2017: 49(1): 1600016.

30. Bowerman C, Bhakta NR, Brazzale D, et al. A race-neutral approach to the interpretation of lung function measurements. *Am J Respir Crit Care Med* 2023: 207(6): 768-774.

31. Niedbalski PJ, Hall CS, Castro M, et al. Protocols for multi-site trials using hyperpolarized (129) Xe MRI for imaging of ventilation, alveolar-airspace size, and gas

exchange: A position paper from the (129) Xe MRI clinical trials consortium. *Magn Reson Med* 2021: 86(6): 2966-2986.

32. Niedbalski PJ, Willmering MM, Thomen RP, et al. A single-breath-hold protocol for hyperpolarized (129) Xe ventilation and gas exchange imaging. *NMR Biomed* 2023: e4923.

33. Niedbalski PJ, Lu J, Hall CS, et al. Utilizing flip angle/TR equivalence to reduce breath hold duration in hyperpolarized (129) Xe 1-point Dixon gas exchange imaging. *Magn Reson Med* 2022: 87(3): 1490-1499.

34. Wang Z, Robertson SH, Wang J, et al. Quantitative analysis of hyperpolarized (129) Xe gas transfer MRI. *Med Phys* 2017: 44(6): 2415-2428.

35. 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(1): eabc8180.

36. Wang Z, Bier EA, Swaminathan A, et al. Diverse cardiopulmonary diseases are associated with distinct xenon magnetic resonance imaging signatures. *Eur Respir J* 2019: 54(6).

37. Varga Z, Flammer AJ, Steiger P, et al. Endothelial cell infection and endotheliitis in COVID-19. *Lancet* 2020: 395(10234): 1417-1418.

38. 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.

39. Collier GJ, Eaden JA, Hughes PJC, et al. Dissolved (129) Xe 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.

40. Mummy DG, Bier EA, Wang Z, et al. Hyperpolarized (129)Xe MRI and spectroscopy of gas-exchange abnormalities in nonspecific interstitial pneumonia. *Radiology* 2021: 301(1): 211-220.

FIGURE CAPTIONS

Figure 1. CONSORT Diagram.
Of 80 participants enrolled at Site1, nine were excluded; 4 did not complete MRI, 5 were COVID-negative or fully recovered with comorbidities or exposures that could influence ^{129}Xe MRI (severe COPD [n=2], rheumatoid arthritis [n=1], HIV [n=1], heavy current daily cannabis and tobacco smoking [n=1]), and 2 were deemed misclassified (no positive COVID test and negative antibody test [n=1], positive COVID test between pulmonary function tests and MRI [n=1]). Site2 and Site3 enrolled 40 and 18 participants respectively, of which 2 at Site2 and 3 at Site3 did not complete MRI and were excluded. Site2 (n=7) and Site3 (n=10) retrospectively included COVID-negative controls, whereas Site1 prospectively enrolled. From 139 participants pooled for centralised analysis, four were further excluded (low ^{129}Xe gas SNR [n=3], missing COVID symptoms for classification as recovered or long COVID [n=1]), and the final study group consisted of 135 participants in total (n=28 COVID-negative, n=34 fully recovered, n=73 long COVID).

Figure 2. Internal Cluster Validation and Primary ^{129}Xe MRI Signatures for Final Clusters.
(A) Internal cluster validation using Davies-Bouldin index (minimised; 1.1644, 0.9236, **0.9132**, 0.9167), Dunn's index (maximised; 0.0356, 0.0535, **0.0832**, 0.0786) and silhouette width (maximised; 0.4521, 0.5122, **0.5223**, 0.4984) for 2-5 clusters. Four was the optimal number of clusters across all three indices (dotted box). (B) Radar plot showing the primary ^{129}Xe MRI signatures associated with each cluster. Scales show minimum and maximum for each MRI metric on one of the four radial axes, and points represent mean cluster values.
Mem=interstitial membrane; RBC=red blood cells; VDP=ventilation defect percent.

Figure 3. MRI Measurement Comparison by Cluster.
Comparison of four MRI measurements used to generate long COVID clusters against COVID-negative and fully recovered participants for (A) ventilation defect percent (VDP), (B) interstitial membrane (Mem)/Gas ratio or membrane uptake, (C) red blood cell (RBC)/Mem ratio, and (D) RBC/Gas ratio or RBC transfer. Symbols signify significant differences ($p<0.05$): † vs. COVID-negative, ‡ vs. recovered, ∓ vs, Cluster1, ¥ vs. Cluster2, § vs. Cluster3

Figure 4. Demographic and Pulmonary Function Test Measurement Comparison by Cluster.
Comparison of clinical characteristics across COVID-negative, recovered and long COVID clusters for (A) participant age, (B) body mass index (BMI), (C) St. George's Respiratory Questionnaire (SGRQ) Total Score, (D) forced expiratory volume in one second (FEV_1), (E) forced vital capacity (FVC), (F) diffusion capacity of the lung for carbon monoxide (DL_{CO}), (G) residual volume (RV), (H) total lung capacity (TLC), and (I) RV/TLC. Dotted grey lines show lower and/or upper limits of normal for pulmonary function tests. P-values in graphs are from ANOVA comparing all six groups. Symbols signify significant group-wise differences after Bonferroni correction ($p<0.05$): † vs. COVID-negative, ‡ vs. recovered, ∓ vs. Cluster1, ¥ vs. Cluster2, § vs. Cluster3.
†*As a proof-of-concept using an independent samples t-test, Cluster1 RV/TLC was significantly greater than COVID-negative (p=0.04).

Figure 5. Long COVID Cluster Summary.
(A) ^{129}Xe MR gas, membrane and red blood cell (RBC) images for representative participants in COVID-negative participants, recovered participants, and resulting long COVID clusters. Site location for each participant embedded in images. (B) Qualitative cluster comparison for imaging measurements used to generate clusters and clinical characteristics.
D=defect (no signal); L=low-intensity signal; H=high-intensity signal; RBC=red blood cells; BMI=body mass index; CT=computed tomography; GGO=ground glass opacities; Ret=reticulation.

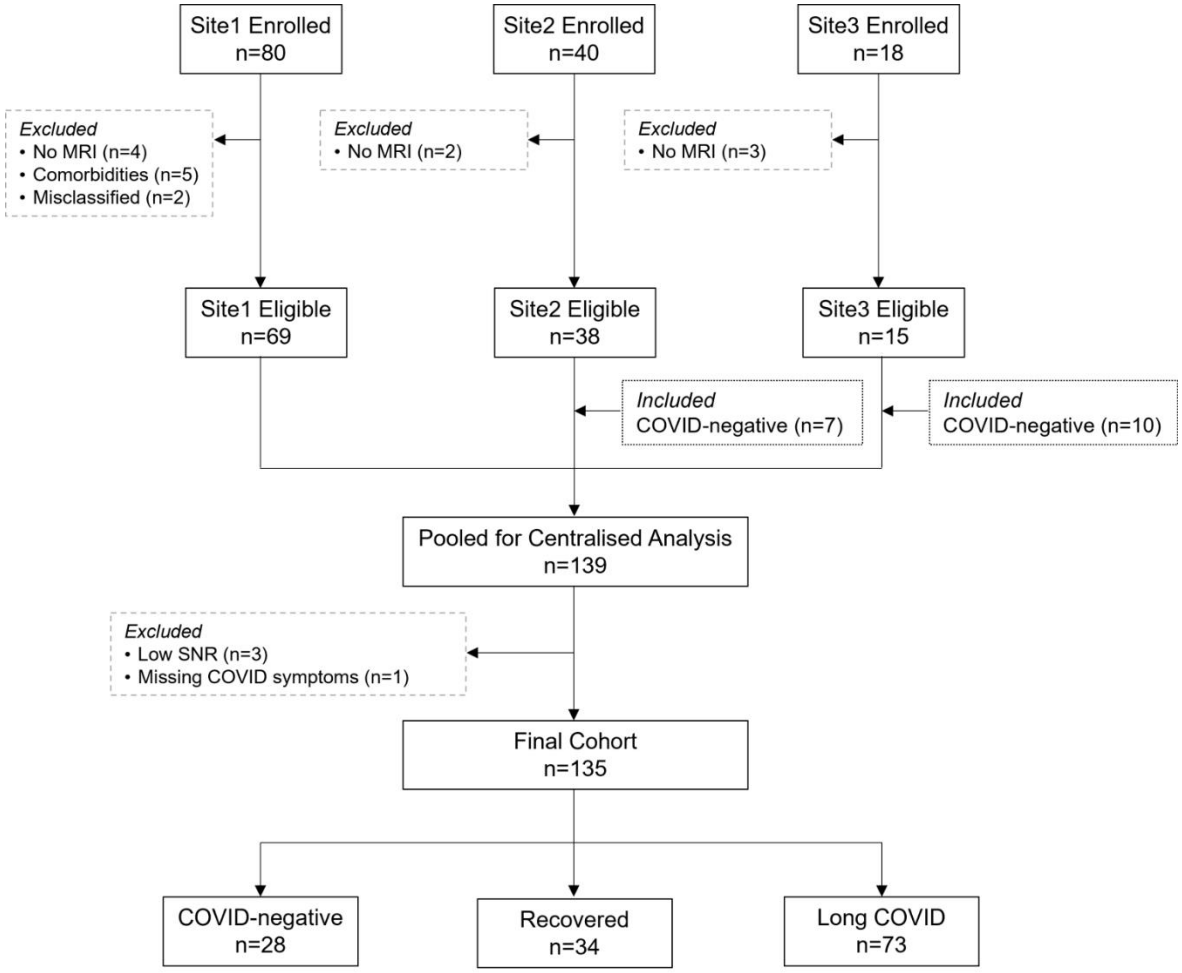

Site1 Enrolled
n=80
Site2 Enrolled
n=40
Site3 Enrolled
n=18
Excluded
• No MRI (n=4)
• Comorbidities (n=5)
• Misclassified (n=2)
Excluded
• No MRI (n=2)
Excluded
• No MRI (n=3)
Site1 Eligible
n=69
Site2 Eligible
n=38
Site3 Eligible
n=15
Included
COVID-negative (n=7)
Included
COVID-negative (n=10)
Pooled for Centralised Analysis
n=139
Excluded
• Low SNR (n=3)
• Missing COVID symptoms (n=1)
Final Cohort
n=135
COVID-negative
n=28
Recovered
n=34
Long COVID
n=73

Figure 1

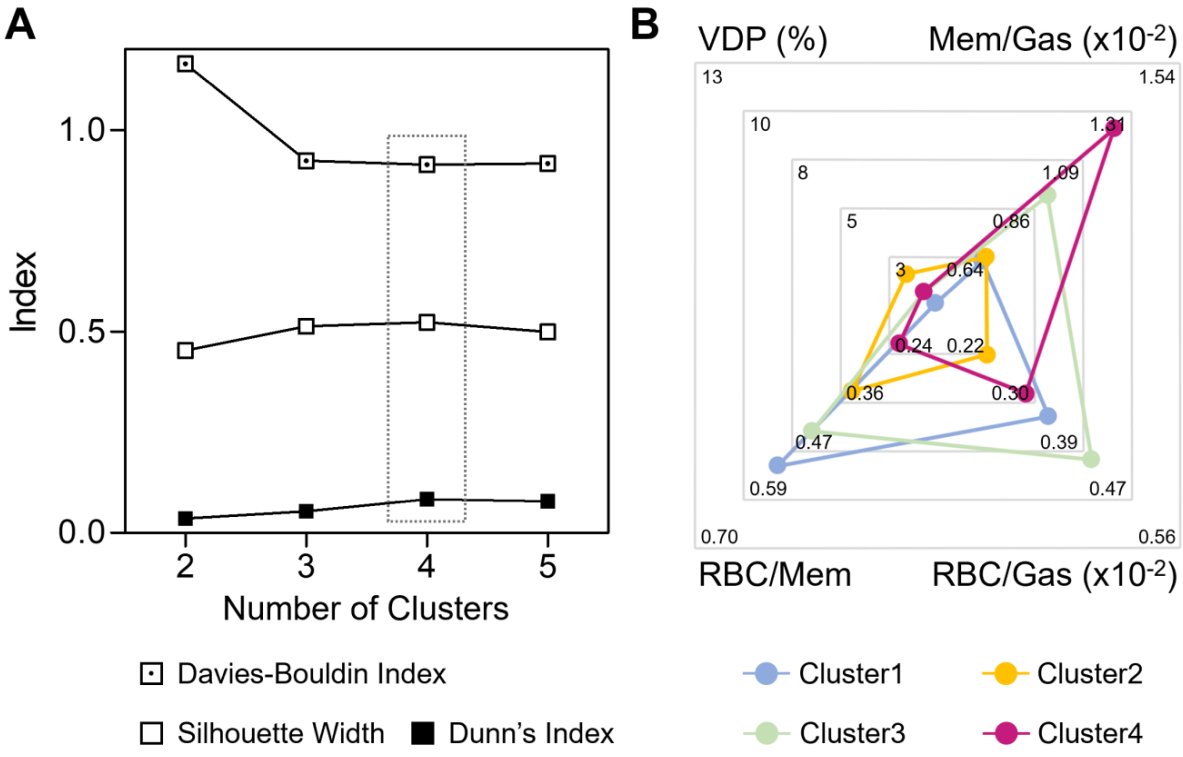

A
1.0
0.5
0.0
Index
2
3
4
5
Number of Clusters
Davies-Bouldin Index
Silhouette Width
Dunn's Index
B
VDP (%)
Mem/Gas (x10-2)
13
1.54
10
1.31
8
1.09
5
0.86
3
0.64
0.24
0.22
0.36
0.30
0.47
0.39
0.59
0.47
0.70
0.56
RBC/Mem
RBC/Gas (x10-2)
Cluster1
Cluster2
Cluster3
Cluster4

Figure 2

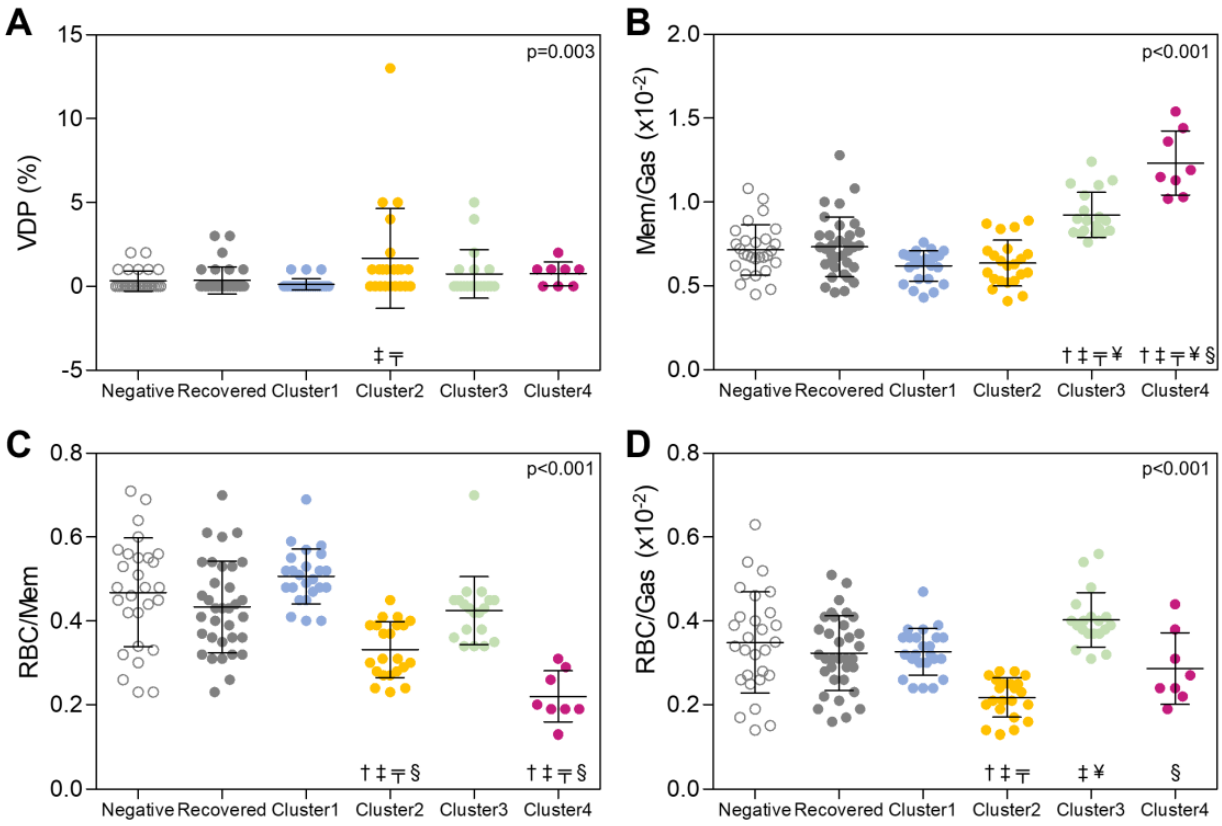

A
VDP (%)
p=0.003
‡ ∓
B
Mem/Gas (x10-2)
p<0.001
† ‡ ∓ ¥
† ‡ ∓ ¥ §
C
RBC/Mem
p<0.001
† ‡ ∓ §
† ‡ ∓ §
D
RBC/Gas (x10-2)
p<0.001
† ‡ ∓
‡ ¥
§
Negative
Recovered
Cluster1
Cluster2
Cluster3
Cluster4

Figure 3

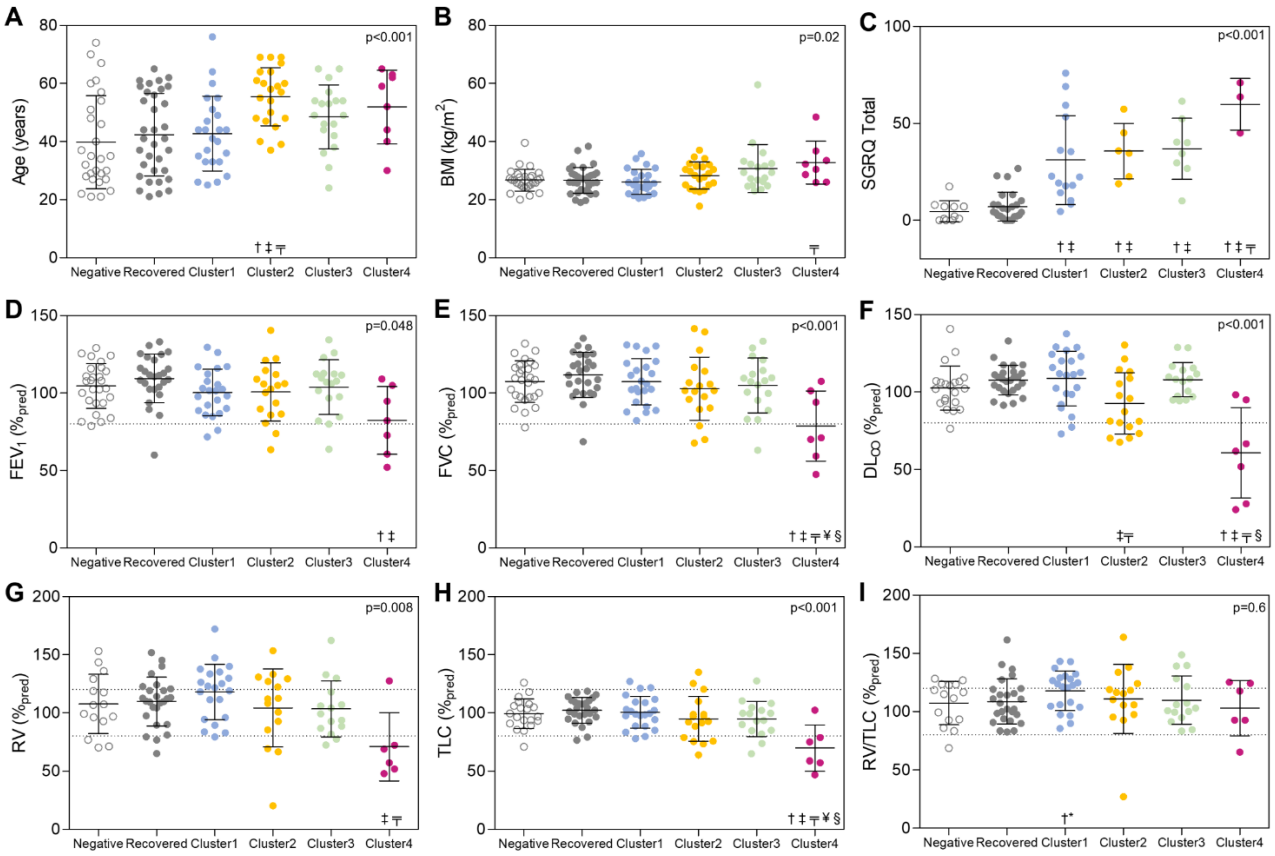

A
Age (years)
p<0.001
B
BMI (kg/m2)
p=0.02
C
SGRQ Total
p<0.001
D
FEV1 (%pred)
p=0.048
E
FVC (%pred)
p<0.001
F
DLCO (%pred)
p<0.001
G
RV (%pred)
p=0.008
H
TLC (%pred)
p<0.001
I
RV/TLC (%pred)
p=0.6
Negative
Recovered
Cluster1
Cluster2
Cluster3
Cluster4

Figure 4

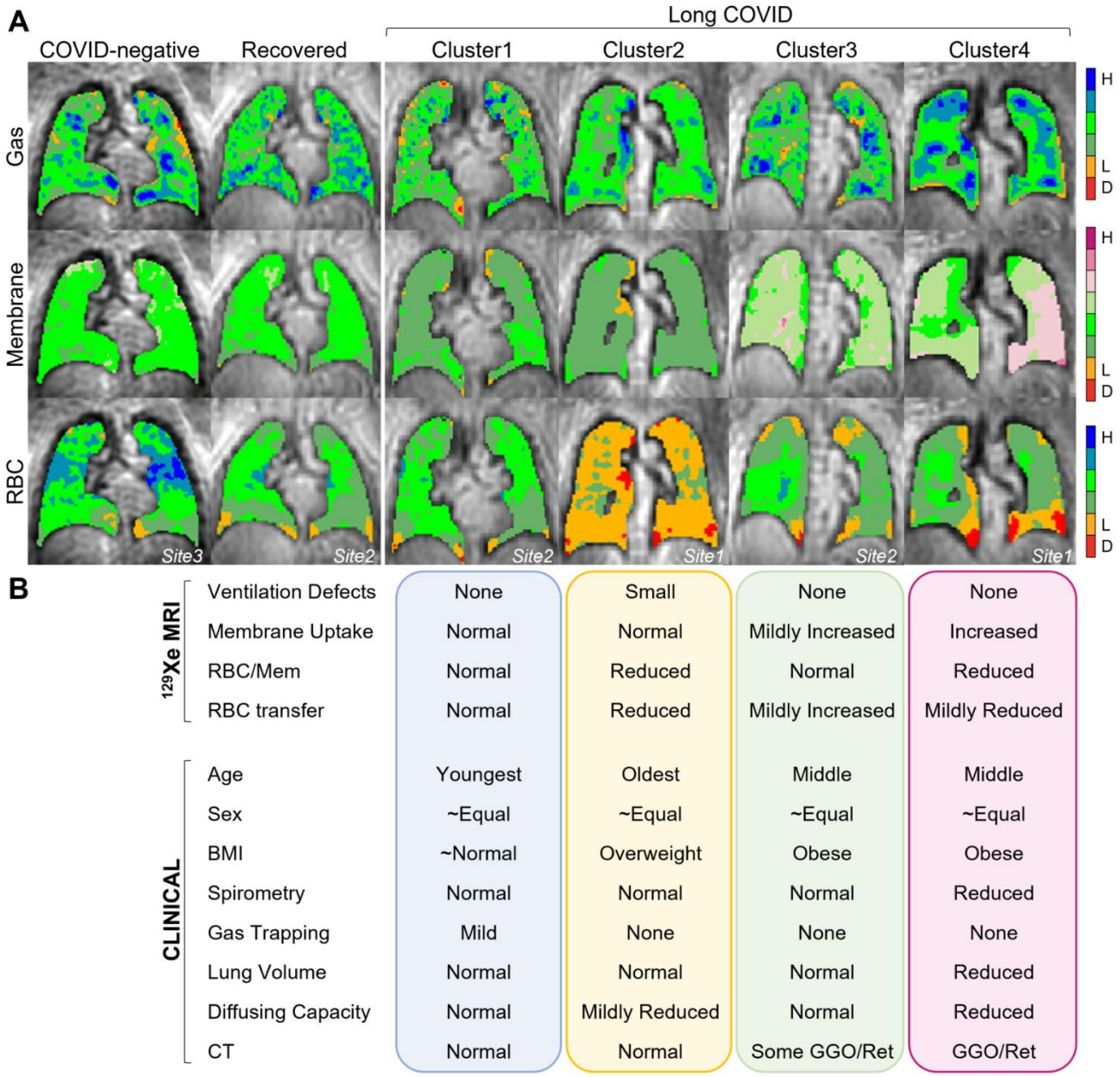

A
Long COVID
COVID-negative
Recovered
Cluster1
Cluster2
Cluster3
Cluster4
Gas
Membrane
RBC
H
L
D
Site3
Site2
Site2
Site1
Site2
Site1
B
^{129}Xe MRI
Ventilation Defects
None
Small
None
None
Membrane Uptake
Normal
Normal
Mildly Increased
Increased
RBC/Mem
Normal
Reduced
Normal
Reduced
RBC transfer
Normal
Reduced
Mildly Increased
Mildly Reduced
CLINICAL
Age
Youngest
Oldest
Middle
Middle
Sex
~Equal
~Equal
~Equal
~Equal
BMI
~Normal
Overweight
Obese
Obese
Spirometry
Normal
Normal
Normal
Reduced
Gas Trapping
Mild
None
None
None
Lung Volume
Normal
Normal
Normal
Reduced
Diffusing Capacity
Normal
Mildly Reduced
Normal
Reduced
CT
Normal
Normal
Some GGO/Ret
GGO/Ret

Figure 5

ONLINE SUPPLEMENT

Cluster Analysis to Identify Long COVID Phenotypes Using ^{129}Xe Magnetic Resonance Imaging: A Multi-centre Evaluation

SUPPLEMENTARY METHODS

Study Participants and Design

Participants aged ≥19-years provided written informed-consent to ethics board-approved protocols between July 2021 and February 2023 across the three centres. Tests were performed by real-time reverse transcription polymerase chain reaction or lateral flow test depending on local guidelines at the time of testing. Self-reported COVID positive test dates and subsequently MRI study enrollment dates at each were:

- Site1 University of British Columbia: infections August 2020 – September 2021, MRI enrollment January 2022 – February 2023
- Site2 University of Kansas Medical Center: infections September 2020 – November 2021, MRI enrollment September 2021 – November 2022
- Site3 Duke University: infections September 2020 – August 2021, MRI enrollment July 2021 – June 2022

MRI Acquisition and Analysis

Total inhaled doses were 1.0L (Site1) or 20% of the participants FVC rounded to the nearest 200mL up to a maximum of 1.25L (Site2, Site3), and participants were instructed to inhale the gas mixture from a Tedlar® bag (Jensen Inert Products, Coral Springs, USA) from functional residual capacity for imaging under breath-hold conditions. Up to 1.0L ^{129}Xe gas was polarised using 9820 hyperpolarisers (Polarean Imaging Inc., Durham, USA) and administered ^{129}Xe volumes were expressed as the dose equivalent volume (DEV), which represents an equivalent ^{129}Xe volume that is 100% isotopically enriched and 100% hyperpolarised [1] to enable comparisons between sites.

Participants first performed an in vivo calibration scan [2] using 100-200mL ^{129}Xe diluted to the total dose with N_2 to determine participant-specific parameters including xenon resonance frequency, transmitter power, TE90 (echo time at which the dissolved ^{129}Xe red blood cell [RBC] and membrane [Mem] signals are 90° out of phase) and RBC/Mem ratio. Gas exchange imaging was then performed using single-breath 1-point Dixon methods. Site1 employed a standard 1-point Dixon method with identical resolution for gas and dissolved phase according to current recommended guidelines [2,3]. Site2 employed an interleaved spiral/radial approach that encodes a high-resolution ventilation image alongside standard resolution dissolved images in a single breath-hold [4], and Site3 employed a flip angle/repetition time equivalence accelerated 1-point Dixon method [5]. While non-harmonised protocols were used because of the post-hoc aggregation of multi-centre data, in all cases the $TR_{90,equiv}$ was 249 ms, enabling comparison of gas exchanges values [5,6]. Anatomical ^{1}H imaging was performed directly following ^{129}Xe gas exchange; Site1 and Site3 acquired ^{1}H images in a separate breath-hold, and Site2 performed ^{1}H imaging during the same breath-hold as ^{129}Xe gas exchange to facilitate image quantification.

Image segmentation was first performed with an automated machine learning algorithm [7], and resulting thoracic cavity masks were manually edited by trained observers as needed. Membrane and RBC dissolved images were divided on a voxel-wise basis by the gas-phase image intensities to create normalised images of membrane uptake and RBC transfer, respectively, and normalised voxel intensities were averaged to get the mean signal value [8].

Long COVID Cluster Internal Validation

Internal cluster validation was evaluated using the Davies-Bouldin index [9], silhouette width [10] and Dunn's index [11], which are measures of the average similarity between clusters, average distance between clusters, and the ratio of the smallest distance between observations in different

clusters to largest cluster diameter, respectively. A minimal Davies-Bouldin index and maximal silhouette width and Dunn's index indicate better quality clustering.

SUPPLEMENTARY RESULTS

The number of participants in each study group significantly differed by study site ($p<0.001$), driven by the lack of recovered participants that were not recruited by design at Site3. Tables E1-E3 show study measurements by site for never-COVID, recovered and long COVID groups, respectively. Most noteworthy, MRI at Site1 was performed significantly later after COVID diagnosis than Site2 for fully recovered (479±231vs. 237±185 days, $p=0.008$) and long COVID participants (534±213 vs. 356±182 days, $p=0.002$), and significantly later than Site3 for long COVID participants (vs. 242±187 days, $p<0.001$). Long COVID participants at Site3 had generally reduced pulmonary function test measurements compared with the other two sites ($p<0.05$ vs. at least Site1 or Site2). Although significantly different across sites for some groups, SNR was adequate for analysis in all cases and was not significantly different between combined study groups, supporting aggregation of the multi-center data for clustering. We note that membrane and RBC SNR are dependent on the ^{129}Xe transfer efficiency (Mem/Gas, RBC/Gas, RBC/Mem) which are increased or reduced in some patients, thus membrane and RBC SNR are not directly interpretable.

Participants from each site were distributed throughout all clusters (Table E4). There were no differences in prevalence of specific comorbidities across the clusters, nor the total number of comorbidities by participant ($p=0.2$, Table E5). The median number of comorbidities was significantly different across the clusters ($p=0.04$, Table E5), with Cluster3 and Cluster4 trending the greatest number of comorbidities ($p=0.055$ and $p=0.054$, respectively vs Cluster1). Tables E6-

E17 show full cluster-wise comparisons with post-hoc Bonferroni corrected p-values. Because age has been shown to be significantly related to ^{129}Xe MRI RBC/Mem [12,13], we evaluated age relationships to separate potential contributions of age versus long COVID to differences in RBC/Mem across groups and clusters. As shown in Figure E3, RBC/Mem was significantly negatively correlated with age in never-COVID (ρ=-0.68, $p<0.001$), recovered (ρ=-0.35, p=0.04) and long COVID (ρ=-0.42, $p<0.001$). The age ranges were overlapping across the groups though, suggesting that age-RBC/Mem relationships did not largely contribute to cluster separation. BMI was not significantly related to RBC/Mem in the COVID-negative nor recovered groups (both p=0.7). For all long COVID participants, BMI was significantly but weakly correlated with RBC/Mem (ρ=-0.31, p=0.007), although likely driven by two participants with very obese BMIs (49, 60 kg/m2). The BMI ranges overlapped across all groups, suggesting that BMI differences across clusters are true physiological characteristics. BMI was highest in Cluster4, and could have could have been a risk factor for these patients for more severe SARS-CoV-2 infections (75% were hospitalised during their acute infection) as well as development of long COVID.

Table E1. COVID-negative Participant Demographics and Study Measurements by Site.

Parameter ± SD	ALL (n=28)	Site1 (n=11)[1]	Site2 (n=7)[2]	Site3 (n=10)[3]	p-value*
Age years	40 ± 16	44 ± 18	33 ± 13	41 ± 16	0.3
Female n (%)	16 (57)	8 (73)	3 (43)	5 (50)	0.4
Caucasian n (%)	22 (79)	7 (64)	7 (100)	8 (80)	0.5
BMI kg/m^2	27 ±4	25 ± 3	27 ± 6	28 ± 3	0.2
Dx days n	-	-	-	-	-
FEV_1 $\%_{pred}$	105 ± 15	106 ± 14	104 ± 15	103 ± 17	0.9
FVC $\%_{pred}$	107 ± 13	109 ± 13	105 ± 9	107 ± 16	0.9
FEV_1/FVC %	81 ± 8	81 ± 9	82 ± 6	80 ± 7	0.8
RV $\%_{pred}$	108 ± 25	113 ± 17	-	93 ± 40	0.2
TLC $\%_{pred}$	99 ± 13	102 ± 11	-	96 ± 14	0.4
RV/TLC $\%_{pred}$	107 ± 19	112 ± 15	-	94 ± 23	0.09
DL_{CO} $\%_{pred}$	103 ± 14	105 ± 16	-	99 ± 12	0.3
VDP %	0 ± 1	0 ± 0	0 ± 1	1 ± 1†	**0.04**
Mem/Gas	0.71 ± 0.15	0.69 ± 0.16	0.76 ± 0.18	0.70 ± 0.11	0.6
RBC/Mem	0.47 ± 0.13	0.39 ± 0.11	0.53 ± 0.12	0.51 ±0.13	**0.046**
RBC/Gas	0.35 ± 0.12	0.29 ± 0.11	0.40 ± 0.11	0.38 ± 0.13	0.1
DEV mL	190 ± 36	191 ± 21	204 ± 53	178 ± 34	0.4
Gas SNR	21 ± 6	22 ± 3	25 ± 7	16 ± 4†‡	**0.003**
Mem SNR	19 ± 8	19 ± 5	28 ± 7†	14 ± 6‡	**<0.001**
RBC SNR	9 ± 5	8 ± 3	15 ± 4†	7 ± 4‡	**<0.001**

BMI=body mass index; DEV=dose equivalent volume; Dx days=days between COVID diagnosis and MRI; DL_{CO}=diffusion capacity of the lungs for carbon monoxide; FEV_1=forced expiratory volume in one second; FVC=forced vital capacity; Mem=interstitial membrane; RBC=red blood cells; RV=residual volume; SNR=signal-to-noise ratio; TLC=total lung capacity; VDP=ventilation defect percent; $\%_{pred}$=percent predicted.

*Using one-way ANOVA for parametric variables or Kruskal-Wallis tests for non-parametric variables.

[1]All participants at Site1 have all measurements; [2]Site2 n=6 for FEV_1, FVC, FEV_1/FVC, n=0 for TLC and DL_{CO}; [3]Site3 n=7 for TLC.

†Significantly different from Site1 ($p<0.05$); ‡Significantly different from Site2 ($p<0.05$)

Table E2. Recovered Participant Demographics and Study Measurements by Site.

Parameter ± SD	ALL (n=35)	Site1 (n=25)[1]	Site2 (n=9)[2]	Site3 (n=0)[3]	p-value*
Age years	43 ± 14	45 ± 14	37 ± 14	-	0.1
Female n (%)	20 (57)	14 (56)	6 (67)	-	0.7
Caucasian n (%)	28 (80)	19 (76)	9 (100)		0.8
BMI kg/m^2	27 ± 5	27 ± 5	27 ± 4	-	0.8
Dx days n	417 ± 243	479 ± 231	237 ± 185	-	**0.008**
FEV_1 $\%_{pred}$	107 ± 20	107 ± 20	99	-	-
FVC $\%_{pred}$	110 ± 16	111 ± 16	93	-	-
FEV_1/FVC %	79 ± 9	79 ± 9	87	-	-
RV $\%_{pred}$	112 ± 23	113 ± 21	65	-	-
TLC $\%_{pred}$	102 ± 11	102 ± 10	79	-	-
RV/TLC $\%_{pred}$	112 ± 24	113 ± 24	83	-	-
DL_{CO} $\%_{pred}$	108 ± 10	108 ± 10	104	-	-
VDP %	0 ± 1	0 ± 1	1 ± 1	-	0.5
Mem/Gas	0.73 ± 0.18	0.74 ± 0.15	0.72 ± 0.25	-	0.7
RBC/Mem	0.44 ± 0.11	0.44 ± 0.10	0.42 ± 0.14	-	0.6
RBC/Gas	0.33 ± 0.09	0.34 ± 0.09	0.30 ± 0.09	-	0.3
DEV mL	197 ± 47	211 ± 32	155 ± 60	-	**0.001**
Gas SNR	22 ± 6	23 ± 5	20 ± 8	-	0.2
Mem SNR	22 ± 8	20 ± 5	25 ± 14	-	0.2
RBC SNR	9 ± 3	9 ± 3	10 ± 4	-	0.9

BMI=body mass index; DEV=dose equivalent volume; Dx days=days between COVID diagnosis and MRI; DL_{CO}=diffusion capacity of the lungs for carbon monoxide; FEV_1=forced expiratory volume in one second; FVC=forced vital capacity; Mem=interstitial membrane; RBC=red blood cells; RV=residual volume; SNR=signal-to-noise ratio; TLC=total lung capacity; VDP=ventilation defect percent; $\%_{pred}$=percent predicted.

*Using independent samples t-tests for parametric variables or Mann-Whitney U-tests for non-parametric variables.

[1]All participants at Site1 have all measurements; [2]Site2 n=1 for all pulmonary function tests; [3]Site3 did not recruit fully recovered COVID participants by design.

Table E3. Long COVID Participant Demographics and Study Measurements by Site.

Parameter ± SD	ALL (n=73)	Site1 (n=32)[1]	Site2 (n=28)[2]	Site3 (n=13)[3]	p-value*
Age years	49 ± 13	48 ± 11	50 ± 13	50 ± 12	0.8
Female n (%)	44 (60)	18 (56)	20 (71)	6 (46)	0.3
Caucasian n (%)	62 (85)	22 (69)	27 (96)	13 (100)	0.4
BMI kg/m^2	29 ± 6	28 ± 6	29 ± 7	29 ± 4	0.6
Dx days n	414 ± 226	534 ± 213	356 ± 182†	242 ± 187†‡	**<0.001**
FEV_1 $\%_{pred}$	99 ± 18	104 ± 18	98 ± 17	91 ± 19†	0.08
FVC $\%_{pred}$	102 ± 20	109 ± 19	99 ± 16	93 ± 21†	**0.02**
FEV_1/FVC %	79 ± 6	78 ± 6	81 ± 6†	80 ± 6‡	0.2
RV $\%_{pred}$	106 ± 30	117 ± 26	94 ± 18	85 ± 36	**0.001**
TLC $\%_{pred}$	94 ± 18	102 ± 17	87 ± 13	84 ± 19†	**0.001**
RV/TLC $\%_{pred}$	112 ± 22	116 ± 19	110 ± 16	103 ± 35	0.4
DL_{CO} $\%_{pred}$	99 ± 24	104 ± 20	102 ± 19	80 ± 32‡	0.06
VDP %	1 ± 2	0 ± 0	1 ± 1	2 ± 4†	**<0.001**
Mem/Gas	0.77 ± 0.24	0.72 ± 0.19	0.75 ± 0.23	0. 93 ± 0.33	0.1
RBC/Mem	0.40 ± 0.12	0.43 ± 0.12	0.37 ± 0.10	0.39 ± 0.14	0.2
RBC/Gas	0.31 ± 0.09	0.31 ± 0.08	0.29 ± 0.10	0.36 ± 0.08	0.053
DEV mL	186 ± 42	203 ± 41	172 ± 38†	178 ± 40	**0.01**
Gas SNR	22 ± 7	23 ± 4	22 ± 8	20 ± 8	0.5
Mem SNR	22 ± 10	20 ± 7	27 ± 10†	15 ± 9‡	**<0.001**
RBC SNR	9 ± 4	8 ± 3	10 ± 5	5 ± 2†‡	**<0.001**

BMI=body mass index; DEV=dose equivalent volume; Dx days=days between COVID diagnosis and MRI; DL_{CO}=diffusion capacity of the lungs for carbon monoxide; FEV_1=forced expiratory volume in one second; FVC=forced vital capacity; Mem=interstitial membrane; RBC=red blood cells; RV=residual volume; SNR=signal-to-noise ratio; TLC=total lung capacity; VDP=ventilation defect percent; $\%_{pred}$=percent predicted.

*Using one-way ANOVA for parametric variables or Kruskal-Wallis tests for non-parametric variables.

[1]All participants at Site1 have all measurements; [2]Site2 n=22 for FEV_1, FVC, FEV_1/FVC, n=18 for TLC, n=17 for DL_{CO}; [3]Site3 n=10 for TLC, n=12 for DL_{CO}.

†Significantly different from Site1 ($p<0.05$); ‡Significantly different from Site2 ($p<0.05$

Table E4. Number of Participants in Each Cluster by Site.

	Cluster1 (n=24)	Cluster2 (n=22)	Cluster3 (n=19)	Cluster4 (n=8)
Site1 n (%)	15 (62)	6 (27)	8 (42)	3 (38)
Site2 n (%)	6 (25)	14 (64)	7 (37)	1 (12)
Site3 n (%)	3 (13)	2 (9)	4 (21)	4 (50)

Table E5. Interventions during Acute COVID Hospitalisation in Each Cluster.

	ALL (n=73)	**Cluster1 (n=24)**	**Cluster2 (n=22)**	**Cluster3 (n=19)**	**Cluster4 (n=8)**	**p-value***
COVID Hosp n (%)	18 (25)	3 (13)	7 (32)**	3 (16)	5 (63)	**0.03**
Hosp Days mean ± SD	13 ± 14	5 ± 1	9 ± 10	7 ± 3	28 ± 18	**0.03**
Hosp Days median [IQR]	6 [4,22]	5 [5,5]	5 [4,7]	8 [6,9]	24 [21,37]	0.06
Oxygen	11	3 (100)	2 (29)	1 (33)	5 (100)	**0.005**
ICU	8	0 (0)	3 (43)	1 (33)	4 (80)	**0.001**
Ventilation	3	0 (0)	0 (0)	0 (0)	3 (60)	**<0.001**
Type of Ventilation						
Invasive	3	-	-	-	3 (100)	-
Non-invasive	0	-	-	-	0 (0)	-
Medical Therapies						
Steroids	14 (78)	3 (100)	5 (71)	3 (100)	3 (60)	**0.045**
Monoclonal Antibodies	1 (6)	0 (0)	0 (0)	1 (33)	0 (0)	-
Antivirals	6 (33)	2 (67)	1 (14)	1 (33)	2 (40)	0.057
Antibiotics	4 (22)	1 (33)	2 (29)	0 (0)	1 (20)	0.08
Anticoagulants	1 (6)	0 (0)	1 (14)	0 (0)	0 (0)	0.05
Convalescent Plasma	1 (6)	0 (0)	1 (14)	0 (0)	0 (0)	0.05
Bronchodilators	1 (6)	0 (0)	1 (14)	0 (0)	0 (0)	0.05
Others	2 (11)	1 (33)	1 (14)	0 (0)	0 (0)	0.05

All values presented as number of participants n (%) unless otherwise stated, where the fraction in brackets represents the fraction of hospitalised patients within each cluster.
Steroids including dexamethasone, prednisone and/or not specified;
Monoclonal antibodies including tocilizumab;
Antivirals including remdesivir;
Antibiotics including piperacillin/tazobactam (Zosyn), clindamycin or not specified;
Anticoagulants including enoxaparin (Lovenox);
Bronchodilators including albuterol;
Others including potassium, sodium, sulfasalazine (disease-modifying anti-rheumatic drug), colchicine, primaquine (anti-malarial);
*Using one-way ANOVA for parametric variables, Kruskal-Wallis tests for non-parametric variables, or Fisher's exact test for categorial variables.
**COVID in-patient medical therapy information available for 6 participants.

Table E6. Number of Comorbidities by Participant in Each Cluster.

	ALL (n=73)	Cluster1 (n=24)	Cluster2 (n=22)	Cluster3 (n=19)	Cluster4 (n=8)	p-value*
Median	1	0.5	1	1	1.5	**0.04**
Frequencies						0.2
0	23	12	4	4	3	-
1	29	10	11	7	1	-
2	14	2	4	5	3	-
3	4	0	1	2	1	-
4	3	0	2	1	0	-

*Using one-way Kruskal-Wallis tests for non-parametric variables or Fisher's exact test for categorial variables.

Table E7. Pre-Existing Respiratory Disease and Respiratory Medications in Each Cluster.

	ALL (n=73)	Cluster1 (n=24)	Cluster2 (n=22)	Cluster3 (n=19)	Cluster4 (n=8)	p-value*
Asthma n (%)	20 (27)	5 (21)	8 (36)	6 (32)**	1 (13)	0.5
COPD n (%)	2 (3)	0 (0)	1 (5)	1 (5)	0 (0)	0.6
Respiratory Medications						
SABA	12 (55)	3 (6)	5 (56)	4 (57)	0	0.9
SAMA	0 (0)	0 (0)	0 (0)	0 (0)	0	-
SABA/SAMA	1 (5)	0 (0)	0 (0)	1 (14)***	0	0.6
LABA	1 (5)	1 (20)	0 (0)	0 (0)	0	0.3
LAMA	2 (9)	0 (0)	0 (0)	2 (29)***	0	0.2
LABA/LAMA	0 (5)	0 (0)	0 (0)	0 (0)	0	-
ICS	4 (18)	1 (20)	2 (22)	1 (14)	0	1.0
ICS/LABA	10 (45)	2 (40)	4 (44)***	4 (57)	0	0.8
ICS/LABA/LAMA	0 (0)	0 (0)	0 (0)	0 (0)	0	-
LTRA	1 (5)	0 (0)	1 (11)	0 (0)	0	1.0
Monoclonal Antibodies	1 (5)	0 (0)	0 (0)	1 (14)	0	0.6
OCS	1 (5)	0 (0)	0 (0)	0 (0)	1 (100)	0.05
Antibiotics	1 (5)	0 (0)	0 (0)	0 (0)	1 (100)	0.05

All values presented as number of participants n (%), where the fraction in brackets represents the fraction of patients with respiratory disease (asthma or COPD) within each cluster.

SABA=short-acting β-agonist; SAMA=short-acting muscarinic antagonist; LABA=long-acting β-agonist; LAMA=long-acting muscarinic antagonist; ICS=inhaled corticosteroid; LTRA=leukotriene receptor antagonist; OCS=oral corticosteroids.

*Using Fisher's exact test for categorical variables.

**Respiratory medications available for 5 participants with asthma.

***Including participant with COPD.

Table E8. Groupwise Comparisons for VDP.

ANOVA p=0.003	Never	Recovered	Cluster1	Cluster2	Cluster3	Cluster4
Negative	-	-	-	-	-	-
Recovered	1.0	-	-	-	-	-
Cluster1	1.0	1.0	-	-	-	-
Cluster2	0.09	**0.02**	**0.005**	-	-	-
Cluster3	1.0	1.0	1.0	0.9	-	-
Cluster4	0.9	0.6	0.2	1.0	1.0	-

VDP=ventilation defect percent derived from ^{129}Xe MRI.

Table E9. Groupwise Comparisons for Mem/Gas.

ANOVA p<0.001	Never	Recovered	Cluster1	Cluster2	Cluster3	Cluster4
Negative	-	-	-	-	-	-
Recovered	1.0	-	-	-	-	-
Cluster1	0.3	0.07	-	-	-	-
Cluster2	1.0	0.3	1.0	-	-	-
Cluster3	**<0.001**	**<0.001**	**<0.001**	**<0.001**	-	-
Cluster4	**<0.001**	**<0.001**	**<0.001**	**<0.001**	**<0.001**	-

Mem/Gas=ratio of interstitial membrane to gas signal derived from ^{129}Xe MRI.

Table E10. Groupwise Comparisons for RBC/Mem.

ANOVA p<0.001	Never	Recovered	Cluster1	Cluster2	Cluster3	Cluster4
Negative	-	-	-	-	-	-
Recovered	1.0	-	-	-	-	-
Cluster1	1.0	0.08	-	-	-	-
Cluster2	**<0.001**	**0.002**	**<0.001**	-	-	-
Cluster3	1.0	1.0	0.1	**0.03**	-	-
Cluster4	**<0.001**	**<0.001**	**<0.001**	0.08	**<0.001**	-

RBC/Mem=ratio of red blood cell to interstitial membrane signal derived from ^{129}Xe MRI.

Table E11. Groupwise Comparisons for RBC/Gas.

ANOVA p<0.001	Never	Recovered	Cluster1	Cluster2	Cluster3	Cluster4
Negative	-	-	-	-	-	-
Recovered	1.0	-	-	-	-	-
Cluster1	1.0	1.0	-	-	-	-
Cluster2	**<0.001**	**<0.001**	**<0.001**	-	-	-
Cluster3	0.5	**0.02**	0.052	**<0.001**	-	-
Cluster4	1.0	1.0	1.0	0.7	**0.02**	-

RBC/Gas=ratio of red blood cell to gas signal derived from ^{129}Xe MRI.

Table E12. Groupwise Comparisons for Age.

ANOVA p<0.001	Never	Recovered	Cluster1	Cluster2	Cluster3	Cluster4
Negative	-	-	-	-	-	-
Recovered	1.0	-	-	-	-	-
Cluster1	1.0	1.0	-	-	-	-
Cluster2	**<0.001**	**0.007**	**0.02**	-	-	-
Cluster3	0.4	1.0	1.0	1.0	-	-
Cluster4	0.4	1.0	1.0	1.0	1.0	-

Table E13. Groupwise Comparisons for BMI.

ANOVA p=0.004	Never	Recovered	Cluster1	Cluster2	Cluster3	Cluster4
Negative	-	-	-	-	-	-
Recovered	1.0	-	-	-	-	-
Cluster1	1.0	1.0	-	-	-	-
Cluster2	1.0	1.0	0.8	-	-	-
Cluster3	0.2	0.1	0.08	1.0	-	-
Cluster4	0.07	0.051	**0.04**	0.6	1.0	-

BMI=body mass index.

Table E14. Groupwise Comparisons for FEV_1.

ANOVA p=0.009	Never	Recovered	Cluster1	Cluster2	Cluster3	Cluster4
Negative	-	-	-	-	-	-
Recovered	1.0	-	-	-	-	-
Cluster1	1.0	0.9	-	-	-	-
Cluster2	1.0	1.0	1.0	-	-	-
Cluster3	1.0	1.0	1.0	1.0	-	-
Cluster4	**0.03**	**0.003**	0.2	0.2	0.06	-

FEV_1=forced expiratory volume in one second.

Table E15. Groupwise Comparisons for FVC.

ANOVA p<0.001	Never	Recovered	Cluster1	Cluster2	Cluster3	Cluster4
Negative	-	-	-	-	-	-
Recovered	1.0	-	-	-	-	-
Cluster1	1.0	1.0	-	-	-	-
Cluster2	1.0	1.0	1.0	-	-	-
Cluster3	1.0	1.0	1.0	1.0	-	-
Cluster4	**0.001**	**<0.001**	**0.001**	**0.02**	**0.007**	-

FVC=forced vital capacity.

Table E16. **Groupwise Comparisons for RV.**

ANOVA p=0.008	Never	Recovered	Cluster1	Cluster2	Cluster3	Cluster4
Negative	-	-	-	-	-	-
Recovered	1.0	-	-	-	-	-
Cluster1	1.0	1.0	-	-	-	-
Cluster2	1.0	1.0	1.0	1.0	-	-
Cluster3	1.0	1.0	1.0	1.0	-	-
Cluster4	0.054	**0.02**	**0.002**	0.1	0.1	-

RV=residual volume.

Table E17. **Groupwise Comparisons for TLC.**

ANOVA p<0.001	Never	Recovered	Cluster1	Cluster2	Cluster3	Cluster4
Negative	-	-	-	-	-	-
Recovered	1.0	-	-	-	-	-
Cluster1	1.0	1.0	-	-	-	-
Cluster2	1.0	1.0	1.0	1.0	-	-
Cluster3	1.0	1.0	1.0	1.0	-	-
Cluster4	**<0.001**	**<0.001**	**<0.001**	**0.007**	**0.008**	-

TLC=total lung capacity.

Table E18. **Groupwise Comparisons for DL_{CO}.**

ANOVA p<0.001	Never	Recovered	Cluster1	Cluster2	Cluster3	Cluster4
Negative	-	-	-	-	-	-
Recovered	1.0	-	-	-	-	-
Cluster1	1.0	1.0	-	-	-	-
Cluster2	0.9	**0.050**	**0.04**	-	-	-
Cluster3	1.0	1.0	1.0	0.1	-	-
Cluster4	**<0.001**	**0.001**	**<0.001**	**<0.001**	**<0.001**	-

DL_{CO}=diffusion capacity of the lungs for carbon monoxide.

Table E19. **Groupwise Comparisons for SGRQ.**

ANOVA p<0.001	Never	Recovered	Cluster1	Cluster2	Cluster3	Cluster4
Negative	-	-	-	-	-	-
Recovered	1.0	-	-	-	-	-
Cluster1	**<0.001**	**<0.001**	-	-	-	-
Cluster2	**<0.001**	**<0.001**	1.0	-	-	-
Cluster3	**<0.001**	**<0.001**	1.0	1.0	-	-
Cluster4	**<0.001**	**<0.001**	**0.03**	0.3	0.3	-

SGRQ=St. George's Respiratory Questionnaire.

Figure E1. ^{129}Xe MRI Measurements by Group.

Plots by group for MRI (A) ventilation defect percent (VDP), (B) interstitial membrane (Mem)/Gas ratio or membrane uptake, (C) red blood cell (RBC)/Mem ratio or RBC transfer, and (D) RBC/Gas ratio, highlight spread and therefore heterogeneity of MRI measurements in the total long COVID group. †Significantly different from COVID-negative ($p<0.05$)

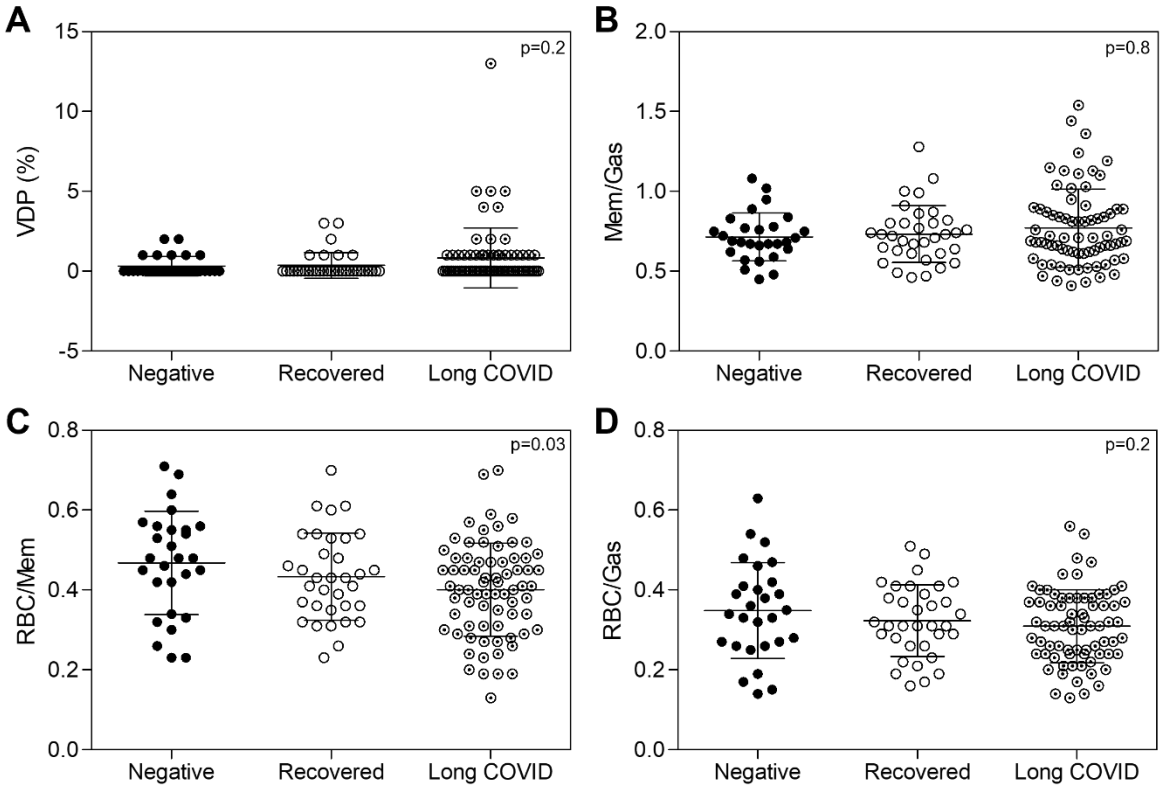


Figure E2. ^{129}Xe MRI and CT Across Long COVID Clusters.

^{129}Xe MR gas, membrane (Mem) and red blood cell (RBC) and computed tomography (CT) images for representative participants in each cluster as well as COVID-negative and recovered participants. All images shown from Site1 with a consistent CT protocol. Cluster1 participant had normal MRI findings and no CT abnormalities. Cluster2 participant had reduced MRI RBC/Mem despite no CT abnormalities. Cluster3 participant exhibited mildly elevated Mem/Gas with some reticulation on CT, whereas Cluster4 participant exhibited markedly elevated MRI Mem/Gas with ground glass, reticulation and honeycombing on CT.

D=defect (no signal); L=low-intensity signal; H=high-intensity signal.

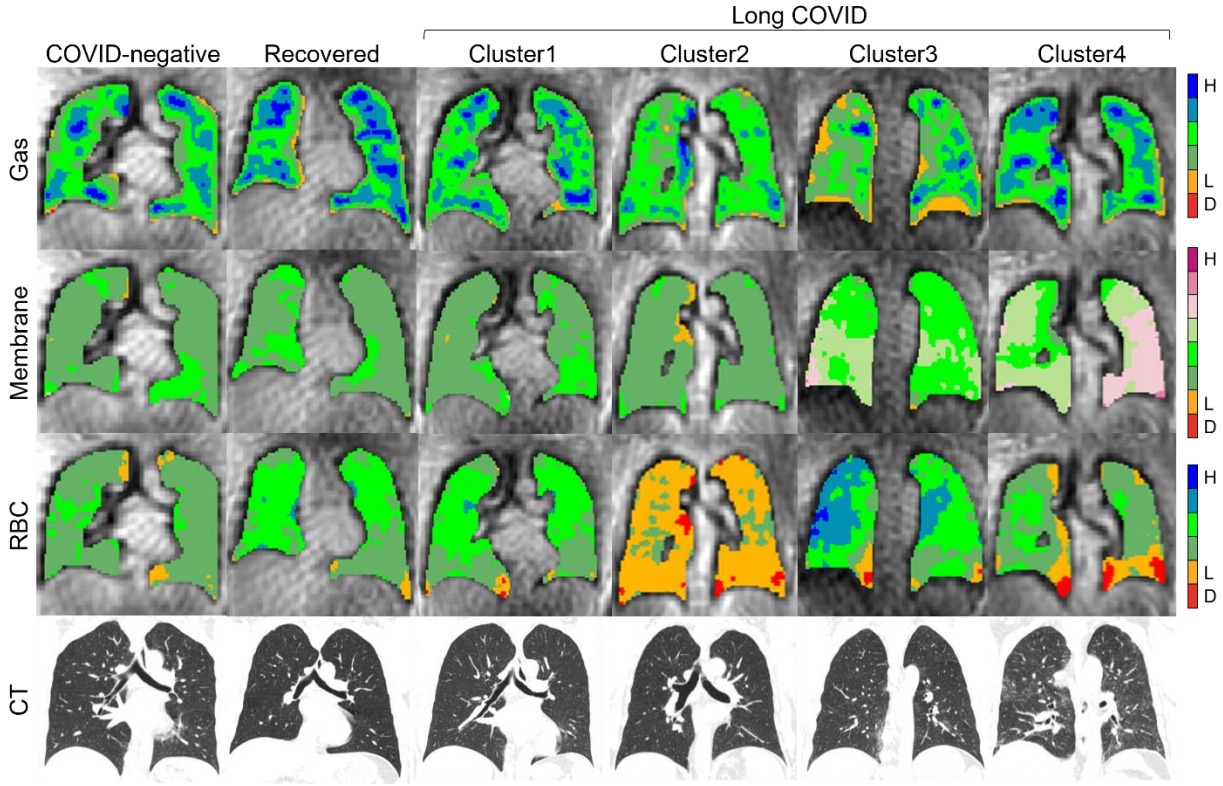
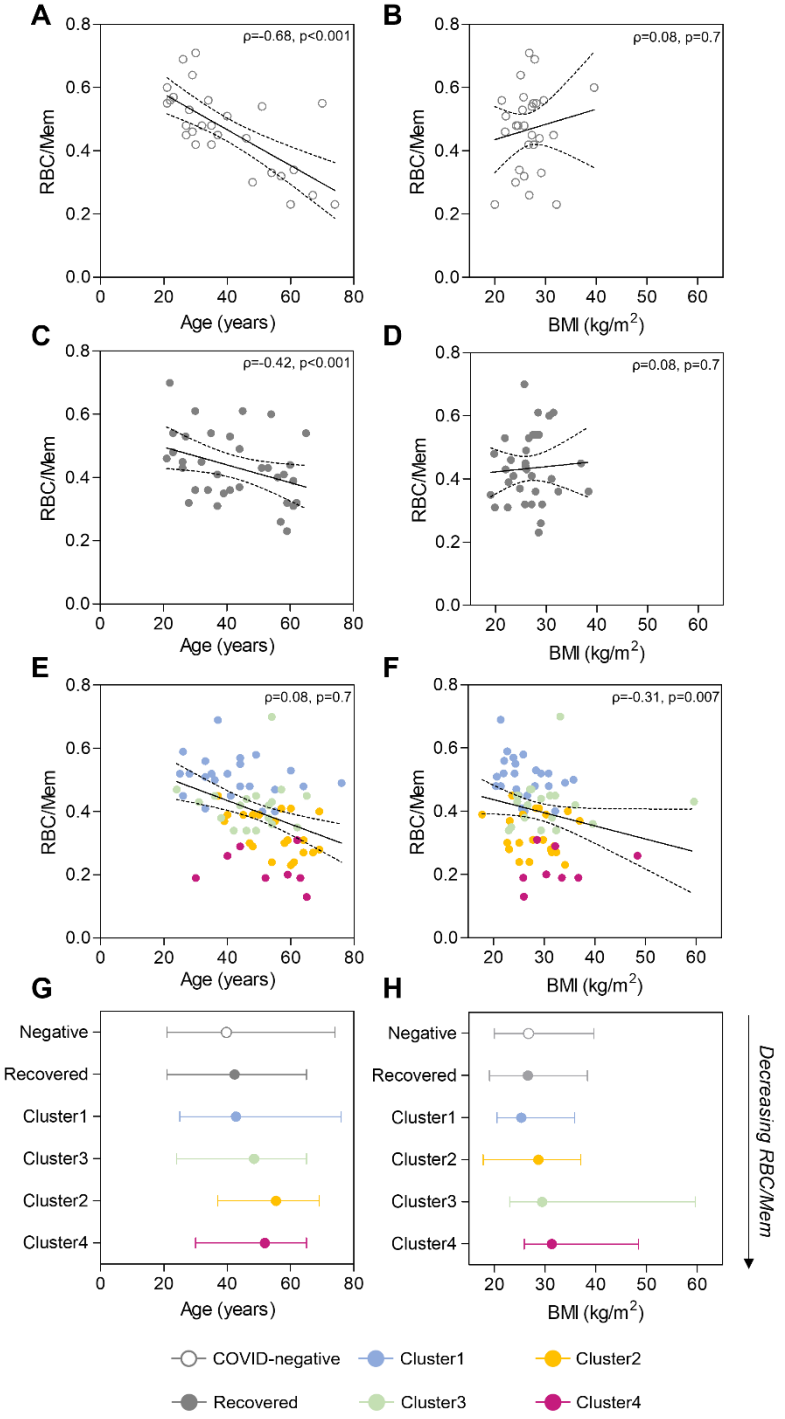


Figure E3. Age and BMI Relationships with MRI RBC/Mem.
^{129}Xe MRI red blood cell to interstitial membrane (RBC/Mem) ratio was significantly correlated with age for (A) COVID-negative, (C) fully recovered, and (E) long COVID groups. RBC/Mem was not significantly correlated with body mass index (BMI) in COVID-negative (B) nor recovered (D) groups. RBC/Mem and BMI were significantly but weakly correlated for all long COVID participants (F). Long COVID clusters shown with coloured points. (D, H) Points show mean age or BMI with range, for groups arranged from top to bottom with decreasing RBC/Mem, demonstrating overlapping age and BMI ranges across all groups.



REFERENCES

1. He M, Robertson SH, Kaushik SS, et al. Dose and pulse sequence considerations for hyperpolarized (129)Xe ventilation MRI. *Magn Reson Imaging* 2015: 33(7): 877-885.

2. Niedbalski PJ, Hall CS, Castro M, et al. Protocols for multi-site trials using hyperpolarized (129) Xe MRI for imaging of ventilation, alveolar-airspace size, and gas exchange: A position paper from the (129) Xe MRI clinical trials consortium. *Magn Reson Med* 2021: 86(6): 2966-2986.

3. Kaushik SS, Robertson SH, Freeman MS, et al. Single-breath clinical imaging of hyperpolarized (129)Xe in the airspaces, barrier, and red blood cells using an interleaved 3D radial 1-point Dixon acquisition. *Magn Reson Med* 2016: 75(4): 1434-1443.

4. Niedbalski PJ, Willmering MM, Thomen RP, et al. A single-breath-hold protocol for hyperpolarized (129) Xe ventilation and gas exchange imaging. *NMR Biomed* 2023: e4923.

5. Niedbalski PJ, Lu J, Hall CS, et al. Utilizing flip angle/TR equivalence to reduce breath hold duration in hyperpolarized (129) Xe 1-point Dixon gas exchange imaging. *Magn Reson Med* 2022: 87(3): 1490-1499.

6. Ruppert K, Amzajerdian F, Hamedani H, et al. Assessment of flip angle-TR equivalence for standardized dissolved-phase imaging of the lung with hyperpolarized 129Xe MRI. *Magn Reson Med* 2019: 81(3): 1784-1794.

7. Leewiwatwong S, Lu J, Dummer I, et al. Combining neural networks and image synthesis to enable automatic thoracic cavity segmentation of hyperpolarized (129)Xe MRI without proton scans. *Magn Reson Imaging* 2023.

8. Wang Z, Robertson SH, Wang J, et al. Quantitative analysis of hyperpolarized (129) Xe gas transfer MRI. *Med Phys* 2017: 44(6): 2415-2428.

9. Davies DL, Bouldin DW. A cluster separation measure. *IEEE Trans Pattern Anal Mach Intell* 1979(2): 224-227.

10. Rousseeuw PJ. Silhouettes: A graphical aid to the interpretation and validation of cluster analysis. *J Comp Appl Math* 1987: 20: 53-65.

11. Dunn JC. Well-separated clusters and optimal fuzzy partitions. *Journal of Cybernetics* 1974: 4(1): 95-104.

12. Plummer JW, Willmering MM, Cleveland ZI, et al. Childhood to adulthood: Accounting for age dependence in healthy-reference distributions in (129) Xe gas-exchange MRI. *Magn Reson Med* 2023: 89(3): 1117-1133.

13. Mummy D, Swaminathan A, Bier E, et al. Hyperpolarized 129Xe MRI and spectroscopy in healthy control subjects reveals age-related changes in measurements of pulmonary gas exchange. *Proc Intl Soc Magn Reson Med* 2022: 30(1168).