Open Forum Infectious Diseases

MAJOR ARTICLE

Identification of Distinct Long COVID Clinical Phenotypes Through Cluster Analysis of Self-Reported Symptoms

Grace Kenny,[1,2] Kathleen McCann,[2] Conor O'Brien,[3] Stefano Savinelli,[1,2] Willard Tinago,[1] Obada Yousif,[4] John S. Lambert,[1,3,5] Cathal O'Broin,[1,2] Eoin R. Feeney,[1,2] Eoghan De Barra,[6,7] Peter Doran,[3] and Patrick W. G. Mallon[1,2]; for the All-Ireland Infectious Diseases (AIID) Cohort Study Group

[1]Centre for Experimental Pathogen Host Research, University College Dublin, Dublin, Ireland, [2]Department of Infectious Diseases, St Vincent's University Hospital, Elm Park, Dublin, Ireland, [3]School of Medicine, University College Dublin, Belfield, Dublin, Ireland, [4]Endocrinology Department, Wexford General Hospital, Carricklawn, Wexford, Ireland, [5]Department of Infectious Diseases, Mater Misericordiae University Hospital, Dublin, Ireland, [6]Department of Infectious Diseases, Beaumont Hospital, Beaumont, Dublin, Ireland, and [7]Department of International Health and Tropical Medicine, Royal College of Surgeons in Ireland, Dublin, Ireland

Background. We aimed to describe the clinical presentation of individuals presenting with prolonged recovery from coronavirus disease 2019 (COVID-19), known as long COVID.

Methods. This was an analysis within a multicenter, prospective cohort study of individuals with a confirmed diagnosis of COVID-19 and persistent symptoms >4 weeks from onset of acute symptoms. We performed a multiple correspondence analysis (MCA) on the most common self-reported symptoms and hierarchical clustering on the results of the MCA to identify symptom clusters.

Results. Two hundred thirty-three individuals were included in the analysis; the median age of the cohort was 43 (interquartile range [IQR], 36–54) years, 74% were women, and 77.3% reported a mild initial illness. MCA and hierarchical clustering revealed 3 clusters. Cluster 1 had predominantly pain symptoms with a higher proportion of joint pain, myalgia, and headache; cluster 2 had a preponderance of cardiovascular symptoms with prominent chest pain, shortness of breath, and palpitations; and cluster 3 had significantly fewer symptoms than the other clusters (2 [IQR, 2–3] symptoms per individual in cluster 3 vs 6 [IQR, 5–7] and 4 [IQR, 3–5] in clusters 1 and 2, respectively; $P < .001$). Clusters 1 and 2 had greater functional impairment, demonstrated by significantly longer work absence, higher dyspnea scores, and lower scores in SF-36 domains of general health, physical functioning, and role limitation due to physical functioning and social functioning.

Conclusions. Clusters of symptoms are evident in long COVID patients that are associated with functional impairments and may point to distinct underlying pathophysiologic mechanisms of disease.

Keywords. long COVID; post–acute sequelae of SARS-CoV-2 infection; SARS-CoV-2.

Prolonged recovery from coronavirus disease (COVID-19), increasingly referred to as "long COVID," may occur in up to 37.7% of individuals presenting with COVID-19 [1]. However, clinical presentations vary considerably, and an accepted definition of long COVID currently does not exist. National Institute for Health and Care Excellence guidelines define symptoms after 12 weeks as "post–acute COVID syndrome" [2], while the Centers for Disease Control and Prevention defines symptoms >4 weeks post–COVID-19 diagnosis under the umbrella term "post-COVID conditions" [3]. More recently, the World Health Organization (WHO) has developed a clinical case definition for post–COVID-19 condition with fatigue, shortness of breath, and cognitive dysfunction noted as common symptoms [4].

Studies on long COVID to date are heterogeneous. Those derived from in-person assessments are often focused on symptomatic recovery in individuals who required hospitalization in the acute phase of COVID-19 [5–7], while fewer have described patients who had mild initial illness managed in the community. Those that do include individuals with an initial mild illness tend to involve data from online surveys, remote assessment, or analyses of healthcare databases, which lack objective data [1, 8, 9].

The clinical presentation of patients with long COVID, including symptoms, physical assessment, and diagnostic investigations, remains poorly characterized, and it is unclear if distinct phenotypes exist. Multiple correspondence analysis (MCA) and hierarchical clustering, an exploratory analytical technique that aims to identify homogenous groups of cases, have been used

Received 9 December 2021; editorial decision 26 January 2022; accepted 1 February 2022; published online 7 March 2022.

Correspondence: Grace Kenny, MB BCh BAO, Centre for Experimental Pathogen Host Research, University College Dublin, Belfield, Dublin 4, Ireland (grace.kenny1@ucd.ie).

Open Forum Infectious Diseases®2022

https://doi.org/10.1093/ofid/ofac060

to phenotype diverse conditions [10, 11], but this clustering approach has not yet been applied to long COVID.

This study aims to describe self-reported symptoms of individuals with a range of initial disease severities presenting to tertiary hospitals with long COVID, to identify underlying patterns in presentation using unsupervised cluster analysis, and to correlate these with objective measures of health-related quality of life.

METHODS

Study Design and Participants

The All-Ireland Infectious Diseases (AIID) Cohort Study is a prospective, multicenter study that recruits individuals attending clinical services with issues pertaining to infectious diseases, such as COVID-19, human immunodeficiency virus, or bone and joint infections. Individuals attending general infectious diseases or long COVID clinics in participating centers are routinely invited to participate in the AIID cohort and have clinical details collected at each assessment.

This analysis included AIID cohort participants who had a diagnosis of COVID-19 confirmed by positive polymerase chain reaction (PCR) for severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), with any symptoms persisting beyond 4 weeks from the date of onset of initial COVID-19 symptoms. When participants had >1 clinical assessment available, we only included symptoms recorded at the first visit, unless the first visit occurred <4 weeks from symptom onset.

Patient Consent Statement

Adult (≥18 years) subjects provided written informed consent for collection of data on demographics, clinical characteristics, and investigations undertaken as part of routine care. The study was approved in line with national and European regulations on health research by the St Vincent's Hospital Group Research Ethics Committee and the National Research Ethics Committee for COVID-19 in Ireland.

Clinical Assessments

All participants had in-person evaluations with physical examination and diagnostic investigations as per local hospital standards of care, set by infectious diseases or respiratory specialist teams. The majority of participants were assessed using a standardized form, which incorporated the most frequent self-reported symptoms, and included disease severity, dyspnea score, orthostatic vital signs and electrocardiograms as part of routine clinical assessment, while further investigations were ordered at the discretion of the treating physician based on clinical presentation. Initial COVID-19 disease severity was graded as per the WHO scale [12]. In brief, mild disease included symptomatic participants without evidence of hypoxia; moderate disease included those with oxygen saturation >90% on room air; severe disease included individuals with oxygen saturation <90% on room air; and critical disease included those with either acute respiratory distress syndrome or sepsis. Time off work was recorded as the length of work absence at the most recent follow-up, with some not yet having returned work at this assessment.

Dyspnea was graded as per the Medical Research Council (MRC) dyspnea scale, a validated 5-point scale that assesses dyspnea-related exercise capacity ranging from no limitation (point 1) to significant limitation (point 5). This score has good intraobserver agreement and correlates well with other breathlessness scales and lung function measurements [13]. Patients completed a 36-Item Short-Form Survey (SF-36), a generic measure of health status and quality of life. It consists of 36 questions that evaluate an individual's perception of their performance in 8 domains. The SF-36 has been shown to be a reliable, valid, and sensitive measure of health status in a variety of clinical settings [14]. SF-36 scores in this cohort were compared to normative data for the Irish population [14].

To assess orthostatic vital signs, pulse, and blood pressure were measured 5 minutes after lying supine, then at 1 minute, 3 minutes, and 5 minutes after standing. A drop in blood pressure of ≥20 mm Hg systolic or ≥10 mm Hg diastolic was considered consistent with orthostatic hypotension. A sustained increase in heart rate of >30 beats per minute above supine heart rate, without a drop in blood pressure, was considered consistent with postural tachycardia.

Electrocardiographs (ECGs) and chest radiographs (CXRs) were reviewed by a clinician who evaluated the presence of abnormalities and judged if any changes present were clinically significant. For other cardiac investigations, findings were taken from specialist reports.

Laboratory tests were performed at each site, and routinely included measurement of full blood count, renal function, liver function, D-dimer, fibrinogen, high-sensitivity troponin, high-sensitivity C-reactive protein, creatinine kinase, ferritin, and interleukin 6.

Statistical Analysis

Continuous variables were summarized using median and interquartile range (IQR), and categorical variables as frequency and percentage. Two complementary statistical techniques were used to identify the self-reported symptom clusters. First, to reduce the dimensionality and eliminate redundancy across the 12 self-reported symptoms, MCA was performed. In short, MCA is a principal component analysis method that transforms categorical data into coordinates in multidimensional space (using χ^2 distance between coordinates so similar individuals lie closer together) and outputs several dimensional solutions. An optimal solution is selected using the smallest number of dimensions that account for the largest total explained variance resulting in a reduction in the number of variables needed to summarize the data. We performed the MCA on symptoms that were present in at least 10% of the subjects and collapsed lower-frequency symptoms (eg, diarrhea, nausea, and abdominal

pain) into a single variable (gastrointestinal) to maintain an overall appropriate number of variables for the sample size [15]. The optimal number of dimensions was selected using Scree plot visualization and Kaiser criterion.

Second, we used agglomerative hierarchical clustering on the MCA coordinates using Ward minimum-variance linkage methods. This method starts with each participant as its own cluster, combining the most "similar" participants based on closeness in Euclidean distance, continuing until the last 2 clusters merge into 1 cluster containing all participants. The number of clusters was selected by calculating the sum of the within-cluster inertia at each partition, and partitioning where there is the highest relative loss of inertia [16].

Quantitative and qualitative variables were compared between clusters using the Kruskal-Wallis test and χ^2 test, respectively. A multinomial logistic regression was performed to determine factors independently associated with cluster membership. For the multinomial logistic regression model, missing data was imputed using multiple imputation with predictive mean matching, using the mice package in R. Body mass index (BMI) data were missing for 11%, and disease severity and ethnicity for 2%. All statistical analysis was performed using R version 3.6.2 software.

RESULTS

Subject Characteristics

A total of 1451 subjects were recruited to the AIID cohort between 20 March 2020 and 9 April 2021. Of these, 1082 were individuals without COVID-19, or individuals recruited during the acute phase of COVID-19. Three hundred sixty-nine were individuals referred for a post–COVID-19 assessment, 98 of whom were not experiencing ongoing symptoms. Of those still experiencing symptoms, 29 did not have COVID-19 confirmed by a documented positive SARS-CoV-2 PCR and a further 9 were assessed <4 weeks from symptom onset and were excluded from the analysis. The remaining 233 subjects from 3 tertiary care hospitals in Dublin met the inclusion criteria and were included in the final analysis (Figure 1). Characteristics of the analyzed population are shown in Table 1. The median age of this cohort was 43 (IQR, 36–54) years, with a predominance of mild disease, and only 32% required hospital admission during the acute phase of their COVID-19 illness. The majority (66%) of participants were healthcare workers and 41% reported no history of comorbidity prior to diagnosis of COVID-19.

The median duration of symptoms at the time of first assessment was 18 (IQR, 10–29) weeks, and 161 subjects (69%) reported symptoms for at least 12 weeks since the initial COVID-19 diagnosis. The longest interval between symptom onset and initial assessment was 52 weeks.

Symptoms reported at the time of initial presentation to the long-COVID clinic are shown in Table 2. Fatigue was the most commonly reported symptom (81.9%), followed by shortness of breath (69%), chest pain (41%), palpitations (33.6%), and poor concentration (33.2%).

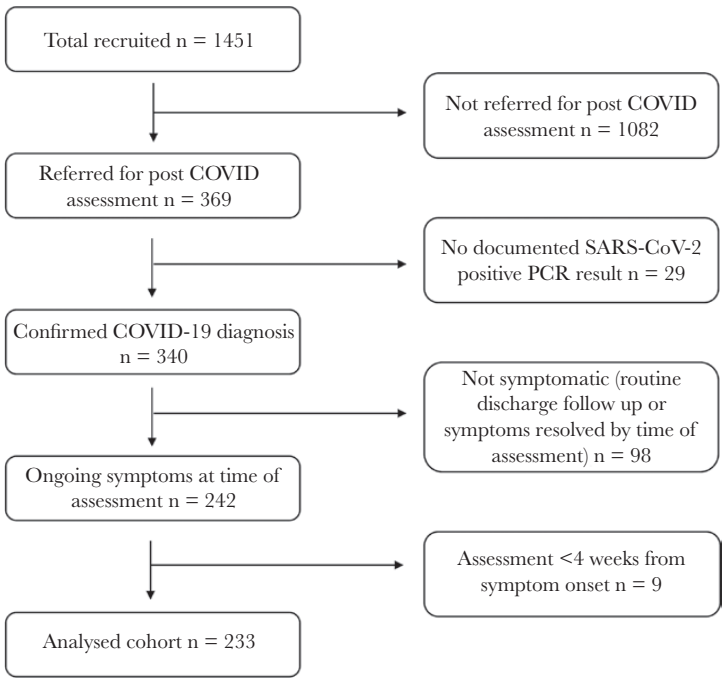

Figure 1. Study cohort flowchart. Abbreviations: COVID-19, coronavirus disease 2019; PCR, polymerase chain reaction; SARS-CoV-2, severe acute respiratory syndrome coronavirus 2.

Table 1. Patient Demographics and Demographic Differences Between Clusters

Characteristic	Total (N = 233)	Cluster 1 (n = 37)	Cluster 2 (n = 87)	Cluster 3 (n = 108)	*P* Value[a]
Age, y, median (IQR)	43 (36–54)	47 (40–55)	42 (35–52)	44 (32–55)	.48
Female sex	173 (74)	32 (86.5)	70 (80.4)	70 (64.8)	<.01
White ethnicity	184 (80)	28 (83.9)	73 (77.8)	83 (79)	.62
BMI kg/m^2, median (IQR)	27.41 (23.75–32.28)	29.5 (25.1–33.7)	27.7 (23.6–33.1)	26.5 (23.5–30.2)	.09
Mild disease[b]	180 (77.3)	29 (78)	76 (87)	75 (69)	.07
Hospitalized during initial COVID-19 illness	75 (32)	26 (70.3)	65 (74.7)	67 (62)	.16
Admitted to ICU	12 (5)	2 (1)	2 (1)	8 (3)	.28
Time from onset of acute COVID-19 symptoms, wk, median (IQR)	18 (10–29)	22.7 (10–37.9)	19.9 (13–27.8)	15.5 (9–27.2)	.01
Healthcare worker	155 (66)	31 (83.8)	63 (72)	60 (55)	.002
Length of work absence, wk, median (IQR)	7 (4–15)	10 (7.5–24)	12 (6–24)	6 (2–12)	<.01
MRC score, median (IQR)	2 (1–3)	2 (2–3)	3 (2–3)	1 (1–2)	<.01
ED attendance, No. (%) of patients attending	86 (37)	15 (45)	43 (49)	28 (26)	<.01

Data are presented as No. (%) unless otherwise specified.

Abbreviations: BMI, body mass index; COVID-19, coronavirus disease 2019; ED, emergency department; ICU, intensive care unit; IQR, interquartile range; MRC, Medical Research Council.

[a] *P* value refers to differences between clusters by χ^2 or Kruskal-Wallis test for qualitative and quantitative data, respectively.

[b] Graded by World Health Organization severity [12].

Cluster Analysis

Of the 233 participants, 1 participant did not have data on self-reported symptoms and was excluded from further analyses. MCA and hierarchical clustering performed on self-reported symptoms revealed 3 clusters (Figure 2A). A heat map of symptoms experienced by subjects in each cluster with the hierarchical clustering dendrogram is shown in Figure 2B.

Table 2. Proportion of Patients Experiencing Symptoms

Symptom	No. (%) of Patients Reporting Symptom
Fatigue	190/231 (81.9)
Respiratory	
Shortness of breath	160/231 (69)
Cough	37/231 (15.9)
Cardiovascular	
Chest pain	96/232 (41)
Palpitations	78/231 (33.6)
Neurological	
Poor concentration	82/232 (35.3)
Headache	48/232 (20.7)
Dizziness	28/227 (12.1)
Anosmia/hyposmia	27/231 (11.6)
Gastrointestinal	
Nausea/vomiting	13/231 (5.6)
Abdominal pain	11/229 (4.7)
Diarrhea	6/231 (2.6)
Musculoskeletal	
Joint pain	53/230 (22.8)
Myalgia	36/232 (15.5)
Other	
Rash	11/232 (4.7)
Sore throat	13/230 (5.6)
Fever	9/226 (4)
Coryza	2/228 (0.9)
Conjunctivitis	2/231 (0.9)

Percentage reflects the proportion of individuals experiencing symptoms within those with complete information for each symptom. For analysis, all gastrointestinal symptoms were collapsed into a single variable; other symptoms present in <10% of individuals were excluded from analysis.

Cluster 1 was the smallest cluster, comprising 37 participants. This cluster reported predominantly musculoskeletal and pain-related symptoms, with joint pain and myalgia being the most characteristic symptoms of this cluster, but also having the highest proportion of headache, dizziness, gastrointestinal symptoms, and cough.

Cluster 2 contained 87 participants and was dominated by cardiorespiratory symptoms, with the presence of chest pain, shortness of breath, or palpitations being the symptoms that best characterized this cluster, followed by fatigue and poor concentration.

Cluster 3 was the largest cluster, comprising 108 participants. Compared to clusters 1 and 2, this cluster was characterized by a significantly lower number of reported symptoms per individual (2 [IQR, 2–3] symptoms per individual in cluster 3 vs 6 [IQR, 5–7] and 4 [IQR, 3–5] in clusters 1 and 2, respectively; $P < .001$). The proportion of individuals experiencing symptoms in each cluster is shown in Supplementary Table 1.

The difference in demographics between clusters is shown in Table 1. There were significantly more women in clusters 1 and 2 as well as significantly more healthcare workers. Cluster 3 had the shortest time from symptom onset to initial review. Although a higher proportion of participants in cluster 3 required hospitalization (38% compared to 29.7% in cluster 1 and 25.3% in cluster 2), this difference was not statistically significant ($P = .16$).

In unadjusted multinomial logistic regression, female sex, being a healthcare worker, and longer duration of symptoms were significantly associated with membership of cluster 1, while female sex, being a healthcare worker, and mild initial severity of initial COVID-19 were significantly associated with

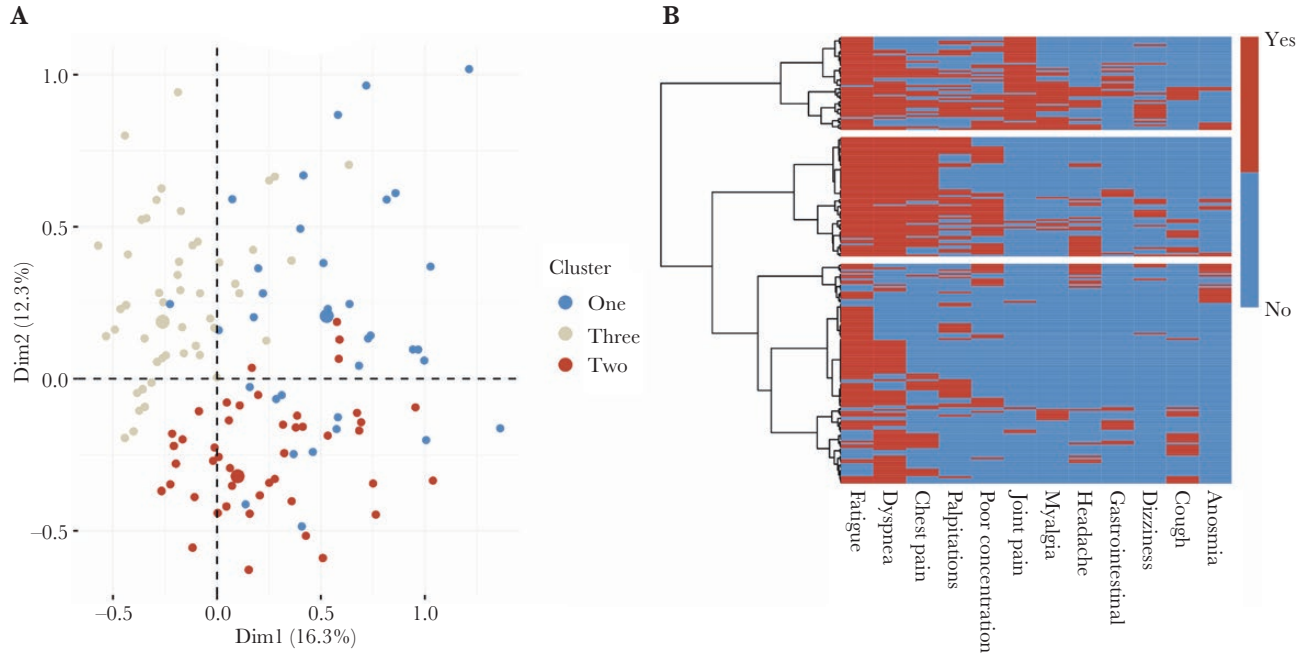

Figure 2. Multiple correspondence analysis (MCA) and hierarchical clustering of symptoms in individuals presenting with prolonged recovery from coronavirus disease 2019 (long COVID). *A*, MCA factor map showing individual coordinates used to generate the dendrogram. *B*, Heatmap showing the symptoms present in individuals within clusters. Compared to cluster 3 (bottom bar), cluster 1 (top bar) demonstrates musculoskeletal and pain symptoms, and cluster 2 (middle bar) shows cardiorespiratory symptoms. Abbreviations: Dim1, dimension 1; Dim2, dimension 2.

membership to cluster 2. Age and ethnicity were not associated with cluster membership. In adjusted analyses including age, sex, employment as a healthcare worker, disease severity and BMI, being female, and employment as a healthcare worker remained significantly associated with cluster 1, while mild initial disease severity and higher BMI were associated with cluster 2.

Functional Outcomes Between Clusters

Cluster 1 and cluster 2 were associated with greater functional impact than cluster 3. Median MRC scores were higher, with a median MRC score of 2 (IQR, 2–3) in cluster 1, and 3 (IQR, 2–3) in cluster 2, whereas cluster 3 had a normal median MRC score of 1 (IQR, 1–2). There was greater healthcare utilization with 45% and 49% of individuals in clusters 1 and 2, respectively, attending an emergency department compared to 26% in cluster 3 ($P < .01$). Clusters 1 and 2 missed a median of 10 (IQR, 7.5–24) weeks and 12 (IQR, 6–24) weeks of employment, respectively, compared to 6 (IQR, 2–12) weeks in cluster 1 ($P < .01$). SF-36 scores were available for 153 patients. Within the whole cohort, mean scores across all domains were lower than that of the general population, with lower scores representing worse health-related quality of life. Mean scores in this cohort compared to the general population were 51.19 vs 83.20 in physical functioning, 16.33 vs 80.5 in role limitation due to physical functioning, 63.38 vs 77.84 in emotional health, 48.56 vs 83.22 in role limitation due to emotional health, 48.84 vs 84.08 in social functioning, 46.02 vs 77.57 in pain, and 53.69 vs 73.82 in general health. Compared to cluster 3, clusters 1 and 2 had significantly lower scores in the following domains: physical functioning, role limitation due to physical functioning, social functioning, and general health, with no significant differences between clusters 1 and 2. There were no between-group differences in role limitation due emotional functioning, energy, or pain domains (Figure 3).

Clinical Assessments and Investigations

Resting heart rate and blood pressure were captured in 178 participants, and 129 participants had orthostatic heart rate and blood pressure assessed. Overall, 9% of participants had a resting tachycardia (>100 beats per minute). Cluster 2 had the highest proportion of tachycardic patients (13% compared to 6% in cluster 1 and 7% in cluster 3), but this difference was not statistically significant ($P = .3$). Of those with orthostatic heart rate assessment, 5 met the criteria for postural tachycardia (4 in cluster 2 and 1 in cluster 3), and 8 met the criteria for orthostatic hypotension (5 in cluster 2 and 3 in cluster 1). Cluster 2 also had the largest proportion of participants who were tachycardic at 5 minutes of standing (26% vs 16% in cluster 1 and 17% in cluster 3), but again this difference was not statistically significant ($P = .42$).

There were no between-group differences in recorded laboratory values, and the median values for laboratory results within each cluster were within the normal range for each test (Supplementary Table 2). CXR data were available for 175 participants, 149 (85%) of whom reported no significant abnormality.

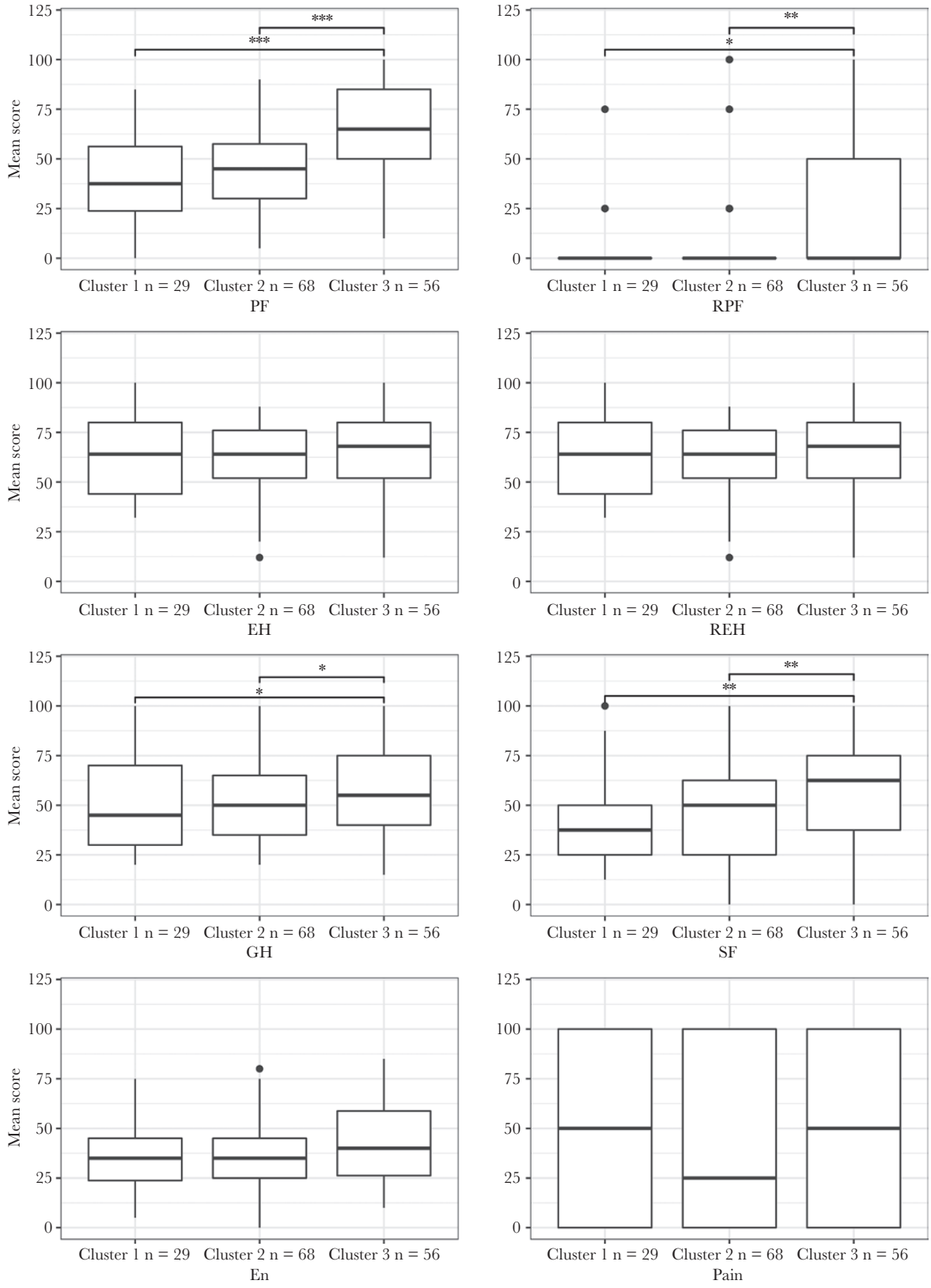

Figure 3. 36-Item Short-Form Survey (SF-36) domain differences between clusters. Boxplots showing SF-36 scores between clusters. Center, median; box limits, first and third quartiles; whiskers, 1.5 × interquartile range; black dots, outliers. Significance determined by Wilcoxon rank-sum test: *$P < .05$; **$P < .01$; ***$P < .001$. Abbreviations: EH, emotional health; En, energy; GH, general health; PF, physical functioning; REH, role limitations due to emotional health; RPF, role limitations due to physical functioning; SF, social functioning.

Of the 26 abnormal CXRs, 20 were dated within 3 weeks of onset of symptoms, with only 6 (3%) taken in the postacute period; 4 of these were from participants in cluster 3, and 2 from cluster 1. The most common abnormality was bilateral infiltrates or consolidation in 20 (77%), with unilateral infiltrates or consolidation in the remaining 6 (23%). The majority (73%) of these individuals had moderate or severe initial disease.

ECG data were available for 90 participants. In addition to the sinus tachycardia described above, only 1 clinically significant abnormality (q waves in lead 3 and augmented vector foot [aVF]) was reported. Echocardiography was performed in 56 subjects: 9 from cluster 1, 33 from cluster 2, and 14 from cluster 3. Overall, 36 (64%) reported no abnormality. Most abnormalities were reported in individuals from cluster 2, including findings consistent with pericarditis (n = 5), diastolic dysfunction (n = 4), left ventricular (LV) dysfunction, (n = 1, ejection fraction 40%), pericardial effusion (n = 2), right ventricular dysfunction (n = 1), and aortic papillary fibroelastoma (n = 1). In cluster 1, reported abnormalities included pericardial effusion (n = 2) and LV dysfunction (n = 1, ejection fraction 50%). Abnormalities reported in cluster 3 included LV dysfunction (n = 2, ejection fraction 35% and 40%), regional wall motion abnormality (n = 2), and diastolic dysfunction (n = 1). Additionally, 23 cardiac magnetic resonance imaging (MRI) scans were captured, with most (74%) performed in subjects in cluster 2. Only 4 MRIs reported abnormalities, 3 with findings consistent with myocarditis and 1 with pericarditis, all of which were from subjects in cluster 2. Of those with abnormal cardiac MRI findings, all had normal ECG and troponin. Two had small pericardial effusions, 1 had findings suggestive of pericarditis, and the remaining echo was normal.

Differences in Individuals Requiring Hospital Admission During Acute Illness

Individuals managed as outpatients during their acute illness had significantly more long COVID symptoms than those admitted (median, 4 [IQR, 2.25–5] in outpatients vs 3 [IQR, 2–5] in admitted patients; $P = .03$), and all symptoms except fatigue, dyspnea, and gastrointestinal symptoms were experienced by a greater proportion of outpatients than hospitalized individuals (data not shown). Those who had been admitted in the acute phase of their illness had significantly higher dyspnea scores (median MRC, 2 [IQR, 2–3] vs 2 [IQR, 1–3]; $P = .04$). Similarly, those admitted to the intensive care unit (ICU) had fewer long COVID symptoms than those not admitted to the ICU (2 [IQR, 2–4] vs 3 [IQR, 3–5]), but higher median dyspnea scores (3 [IQR, 2–3.5] vs 2 [IQR, 1–3]), although these differences did not reach statistical significance ($P = .1$ and $P = .13$, respectively).

DISCUSSION

This is the first study to utilize hierarchical clustering to identify distinct patterns of symptoms in individuals presenting with long COVID and to associate these with objective measures of disability. The results suggest 3 broad clinical phenotypes, 2 of which (musculoskeletal/pain [cluster 1] and cardiorespiratory [cluster 2]) had significantly higher reported symptoms associated with greater functional impairment. Such subclassification may help prioritize patients for assessment and treatment.

Cluster 1 had the highest median number of reported symptoms, and a predominance of pain symptoms such as myalgia, joint pain, and headache. Although there were no consistent objective clinical findings in this cluster, individuals within it reported significant disability, with abnormal median MRC scores and high rates of emergency department attendance. Similarities have been drawn between long COVID and both postinfectious myalgic encephalomyelitis (ME) [17] and fibromyalgia [18], conditions with a female preponderance, often with normal investigations but a profound impact on quality of life. Participants in this cluster share features of both these syndromes, where both fatigue and pain in the absence of inflammation are characteristic, but key symptoms that differentiate ME and fibromyalgia, such as postexertional symptom exacerbation [19], were not assessed in this study.

Cluster 2 reported predominantly cardiorespiratory symptoms of chest pain, palpitations, and shortness of breath. These symptoms were corroborated by objective clinical and diagnostic findings, with the highest rate of tachycardia (both resting and orthostatic) and all those with findings of pericarditis or myocarditis on imaging were contained within cluster 2. Despite 87% of participants in cluster 2 reporting a mild initial COVID-19 illness, symptoms reported as part of long COVID were severe, with a median MRC score of 3 and almost half seeking treatment at an emergency department for their symptoms. Myocarditis has been well described in convalescent individuals with an initial mild illness, with a prevalence ranging from 2.1% to 37% [20, 21], and other abnormal cardiac MRI findings reported in up to 78% [22]. However, the absence of objective inflammation on laboratory tests, normal ECGs, and normal or absent cardiac imaging in many of this group is notable. Of the 17 cardiac MRIs captured in this group, only 23% were abnormal, but the small sample size precludes further interpretation of this finding.

Cluster 3 reported the lowest median number of symptoms and the least associated disability of the 3 clusters, despite having the highest proportion of severe initial COVID-19 disease and undergoing review earlier after onset of COVID-19 symptoms. Fatigue and shortness of breath were the most commonly reported symptoms, but median MRC score was normal, indicating mild dyspnea-related impairment, and this cluster had the lowest proportion of emergency department attendance. Fatigue is well recognized postinfection, persisting in 27% of those with atypical infections managed in the community at 3 months postdiagnosis [23], so it is plausible that this cluster represents individuals within the normal realm of recovery

from an atypical viral illness, which may have useful prognostic consequences for individuals within this cluster.

Risk factors for illness-related absenteeism or loss of productivity associated with long COVID are not well defined. Our analysis points to reduced productivity in individuals within clusters 1 and 2 compared to cluster 3, with significantly longer work absences and greater subjective impairment in ability to work and socialize as measured by SF-36. Of note, although SF-36 scores within the emotional health domain for the whole cohort were lower than population means, there was no difference between clusters, suggesting that physical symptoms were the primary mediator of differences in productivity loss between clusters. While there are few estimates of the socioeconomic impact of long COVID, our data are consistent with other studies; a patient-led survey found that 45% of individuals with long COVID had reduced their workload [9], and a United Kingdom–based population study showed that long COVID was associated with a negative impact on household finances [24]. This analysis shows that a greater economic impact is associated with certain clinical phenotypes and could help guide strategies for more cost-efficient care pathways.

Characterizing long COVID as a single syndrome, rather than classifying based on clinical presentation, may explain some of the inconsistencies apparent in the reported evidence to date. For example, studies do not consistently find a relationship between initial disease severity and subsequent functional impairment [25–27], and contrasting mechanisms for post–COVID-19 dyspnea ranging from pulmonary injury to impaired systemic oxygen extraction have been described [28–31]. Observed tachycardias have been attributed to autonomic dysfunction [32, 33], but standard testing has shown limited benefit [34]. Other mechanisms have been proposed, such as endocrinopathies [35, 36] or alterations in sensory processing and central sensitization [37], which is seen in both fibromyalgia and ME [38, 39], but further research is needed. Other studies have used different analytic approaches to identify clusters of long COVID symptoms. Two analyses, one based on the results of an online survey in Italy and Austria, and another using data from the COVID Symptom Study app, both revealed a multiorgan phenotype similar to cluster 1 in this study [40, 41]. A study using the REal-time Assessment of Community Transmission-2 (REACT-2) cohort from England found 2 clusters, 1 of which, a "respiratory cluster" with a high prevalence of shortness of breath, chest tightness, and chest pain, bears resemblance to cluster 2 in this study [1]. Classifying long COVID based on different clinical presentations, as in these analyses, is a vital first step in understanding potentially distinct underlying pathophysiologic mechanisms, in order to identify appropriate therapeutic interventions.

Our study has limitations. As the AIID cohort records data collected through routine clinical assessments, differences in assessment between centers and symptom-guided referral for investigations resulted in data gaps. While the standardized form used was based off those symptoms most commonly reported in these centers, they could not capture all the symptoms experienced by individuals with long COVID, and the clustering process may have been altered by the inclusion of other variables. Additionally, as objective data were incomplete, we were not able to include these in the clustering process. This study did not validate these clusters in a larger cohort and lacked a control group with which to compare symptom prevalence, and cross-sectional design precluded valuable information on outcome or time to resolution of symptoms. Health service occupational health departments were a major source of referrals for the post–COVID-19 clinics, particularly early after their establishment, which may have resulted in an overrepresentation of women and healthcare workers in our analyzed population. Despite these limitations, this study is large, includes individuals from different referral sources and initial disease severity, and is representative of the diversity presenting in clinical practice and all individuals received an in-person assessment. Using a novel analytical approach on this well-phenotyped cohort has provided new insights that can help classify long COVID.

In summary, this study demonstrates distinct clusters of symptoms in patients with long COVID. Further work is needed to determine the underlying pathophysiology behind these phenotypes. Functional outcome measures demonstrated substantial impact on quality of life and productivity associated with this condition, and underscore the urgency to improve diagnosis and treatment.

Supplementary Data

Supplementary materials are available at *Open Forum Infectious Diseases* online. Consisting of data provided by the authors to benefit the reader, the posted materials are not copyedited and are the sole responsibility of the authors, so questions or comments should be addressed to the corresponding author.

Notes

Author contributions. G. K. conceived the project, coordinated clinical recruitment and clinical data collection, conducted the data analysis, and wrote the manuscript. K. M. conceived the project, coordinated clinical recruitment and clinical data collection, and contributed to writing the manuscript. C. O'Brien coordinated clinical data collection and data interpretation. S. S. conceived and supervised the project and coordinated clinical recruitment and clinical data collection. W. T. supervised data analysis, contributed to interpretation of the results, and contributed to writing the manuscript. C. O'Broin, O. Y., J. S. L., E. F., and E. D. B. coordinated clinical recruitment, clinical data collection, and project administration. P. D. supervised the project and contributed to project administration. P. W. G. M. conceived the project, verified the underlying data, provided input into data analysis and interpretation, and contributed to writing the manuscript.

Acknowledgments. The authors wish to thank all study participants and their families for their participation and support in the conduct of the AIID Cohort Study.

Financial support. This work was supported by an unrestricted grant from Smurfit Kappa.

Potential conflicts of interest. G. K. was funded through a fellowship from the Embassy of the United States in Ireland during the study. E. F. has received consulting fees from Gilead, ViiV, and Vidacare Ireland, and

has been awarded a grant from Science Foundation Ireland, outside the submitted work. E. D. B. has received consulting fees from Sanofi Pasteur and honoraria/travel grant from Pfizer, and is secretary of the Infectious Diseases Society of Ireland. P. W. G. M. has received honoraria and/or travel grants from Gilead Sciences, MSD, Bristol-Myers Squibb, and ViiV Healthcare, and has been awarded grants by Science Foundation Ireland, outside the submitted work. All other authors report no potential conflicts of interest.

All authors have submitted the ICMJE Form for Disclosure of Potential Conflicts of Interest. Conflicts that the editors consider relevant to the content of the manuscript have been disclosed.

All-Ireland Infectious Diseases (AIID) Cohort Study Investigators. Mater Misericordiae University Hospital: A. Cotter, E. Muldoon, G. Sheehan, T. McGinty, J. S. Lambert, S. Green, K. Leamy. St Vincent's University Hospital: G. Kenny, K. McCann, R. McCann, C. O'Broin, S. Waqas, S. Savinelli, E. Feeney, P. W. G. Mallon. Centre for Experimental Pathogen Host Research: A. Garcia Leon, S. Miles, D. Alalwan, R. Negi. Beaumont Hospital: E. de Barra, S. McConkey, K. Hurley, I. Sulaiman. University College Cork: M. Horgan, C. Sadlier, J. Eustace. University College Dublin: C. Kelly, T. Bracken. Sligo University Hospital: B. Whelan. Our Lady of Lourdes Hospital: J. Low. Wexford General Hospital: O. Yousif. University Hospital Galway: B. McNicholas. St Luke's Hospital Kilkenny: G. Courtney. Children's Health Ireland: P. Gavin.

References

1. Whitaker M, Elliott J, Chadeau-Hyam M, et al. Persistent symptoms following SARS-CoV-2 infection in a random community sample of 508,707 people. *medRxiv* [Preprint]. Posted online 3 July **2021**. doi:10.1101/2021.06.28.21259452.
2. National Institute for Health and Care Excellence. COVID-19 rapid guideline: managing the long-term effects of COVID-19. **2020**. https://www.nice.org.uk/guidance/ng188. Accessed 23 August 2021.
3. Centers for Disease Control and Prevention. Post-COVID conditions: information for healthcare providers. **2021**. https://www.cdc.gov/coronavirus/2019-ncov/hcp/clinical-care/post-covid-conditions.html?CDC_AA_refVal=https%3A%2F%2Fwww.cdc.gov%2Fcoronavirus%2F2019-ncov%2Fhcp%2Fclinical-care%2Flate-sequelae.html. Accessed 23 August 2021.
4. World Health Organization. A clinical case definition of post COVID-19 condition by a Delphi consensus. **2021**. https://apps.who.int/iris/handle/10665/345824. Accessed 1 December 2021.
5. Tleyjeh IM, Saddik B, AlSwaidan N, et al. Prevalence and predictors of post-acute COVID-19 syndrome (PACS) after hospital discharge: a cohort study with 4 months median follow-up. *PLoS One* **2021**; 16:e0260568.
6. Asadi-Pooya AA, Akbari A, Emami A, et al. Risk factors associated with long COVID syndrome: a retrospective study. Iran J Med Sci **2021**; 46:428–36.
7. Evans RA, McAuley H, 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**; 9:1275–87.
8. Chevinsky JR, Tao G, Lavery AM, et al. Late conditions diagnosed 1-4 months following an initial COVID-19 encounter: a matched cohort study using inpatient and outpatient administrative data—United States, March 1–June 30, 2020. Clin Infect Dis **2021**; 73(Suppl 1):S5–16.
9. Davis HE, Assaf GS, McCorkell L, et al. Characterizing long COVID in an international cohort: 7 months of symptoms and their impact. EClinicalMedicine **2021**; 38:101019.
10. Bailly S, Destors M, Grillet Y, et al. Obstructive sleep apnea: a cluster analysis at time of diagnosis. PLoS One **2016**; 11:e0157318.
11. Sciascia S, Radin M, Cecchi I, et al. Identifying phenotypes of patients with antiphospholipid antibodies: results from a cluster analysis in a large cohort of patients. Rheumatology **2021**; 60:1106–13.
12. World Health Organization. COVID-19 clinical management living guidance. **2021**. https://www.who.int/publications/i/item/WHO-2019-nCoV-clinical-2021-2. Accessed 13 February 2022.
13. Stenton C. The MRC breathlessness scale. Occup Med **2008**; 58:226–7.
14. Blake C, Codd MB, O'Meara YM. The Short Form 36 (SF-36) Health Survey: normative data for the Irish population. Ir J Med Sci **2000**; 169:195–200.
15. Di Franco G. Multiple correspondence analysis: one only or several techniques? Qual Quant **2016**; 50:1299–315.
16. Husson F, Josse J, Le S, Mazet J. Package 'FactoMineR.' **2020**. https://cran.rediris.es/web/packages/FactoMineR/FactoMineR.pdf. Accessed 23 August 2021.
17. Wong TL, Weitzer DJ. Long COVID and myalgic encephalomyelitis/chronic fatigue syndrome (ME/CFS)—a systemic review and comparison of clinical presentation and symptomatology. Medicina (Kaunas) **2021**; 57:418.
18. Ursini F, Ciaffi J, Mancarella L, et al. Fibromyalgia: a new facet of the post-COVID-19 syndrome spectrum? Results from a web-based survey. RMD Open **2021**; 7:e001735.
19. Carruthers BM, van de Sande MI, De Meirleir KL, et al. Myalgic encephalomyelitis: international consensus criteria. J Intern Med **2011**; 270:327–38.
20. Rajpal S, Tong MS, Borchers J, et al. Cardiovascular magnetic resonance findings in competitive athletes recovering from COVID-19 infection. JAMA Cardiol **2021**; 6:116–8.
21. Eiros R, Barreiro-Perez M, Martin-Garcia A, et al. Pericarditis and myocarditis long after SARS-CoV-2 infection: a cross-sectional descriptive study in healthcare workers. *medRxiv* [Preprint]. Posted online 14 July **2020**. doi:10.1101/2020.07.12.20151316.
22. Puntmann VO, Carerj ML, Wieters I, et al. Outcomes of cardiovascular magnetic resonance imaging in patients recently recovered from coronavirus disease 2019 (COVID-19). *JAMA Cardiol* **2020**; 5:1265.
23. Hickie I, Davenport T, Wakefield D, et al. Post-infective and chronic fatigue syndromes precipitated by viral and non-viral pathogens: prospective cohort study. BMJ **2006**; 333:575.
24. Mayhew M, Kerai G, Ainslie D. Coronavirus and the social impacts of "long COVID" on people's lives in Great Britain: 7 April to 13 June 2021. **2021**. https://www.ons.gov.uk/peoplepopulationandcommunity/healthandsocialcare/conditionsanddiseases/articles/coronavirusandthesocialimpactsoflongcovidonpeopleslivesingreatbritain/7aprilto13june2021. Accessed 22 September 2021.
25. Huang C, Huang L, Wang Y, et al. 6-month consequences of COVID-19 in patients discharged from hospital: a cohort study. Lancet **2021**; 397:220–32.
26. Jacobson KB, Rao M, Bonilla H, et al. Patients with uncomplicated COVID-19 have long-term persistent symptoms and functional impairment similar to patients with severe COVID-19: a cautionary tale during a global pandemic. Clin Infect Dis **2021**; 73:e826–9.
27. Townsend L, Dowds J, O'Brien K, et al. Persistent poor health post-COVID-19 is not associated with respiratory complications or initial disease severity. Ann Am Thorac Soc **2021**; 18:997–1003.
28. Cortés-Telles A, López-Romero S, Figueroa-Hurtado E, et al. Pulmonary function and functional capacity in COVID-19 survivors with persistent dyspnoea. Respir Physiol Neurobiol **2021**; 288:103644.
29. Singh I, Joseph P, Heerdt PM, et al. Persistent exertional intolerance after COVID-19. Chest **2021**; 161:54–63.
30. Baratto C, Caravita S, Faini A, et al. Impact of COVID-19 on exercise pathophysiology: a combined cardiopulmonary and echocardiographic exercise study. J Appl Physiol **2021**; 130:1470–8.
31. Rinaldo RF, Mondoni M, Parazzini EM, et al. Deconditioning as main mechanism of impaired exercise response in COVID-19 survivors. Eur Respir J **2021**; 58:2100870.
32. Dani M, Dirksen A, Taraborrelli P, et al. Autonomic dysfunction in "long COVID": rationale, physiology and management strategies. Clin Med **2021**; 21:e63–7.
33. Radin JM, Quer G, Ramos E, et al. Assessment of prolonged physiological and behavioral changes associated with COVID-19 infection. JAMA Netw Open **2021**; 4:e2115959.
34. Townsend L, Moloney D, Finucane C, et al. Fatigue following COVID-19 infection is not associated with autonomic dysfunction. PLoS One **2021**; 16:e0247280.
35. Bansal R, Gubbi S, Koch CA. COVID-19 and chronic fatigue syndrome: an endocrine perspective. J Clin Transl Endocrinol **2022**; 27:100284.
36. Montefusco L, Ben Nasr M, D'Addio F, et al. Acute and long-term disruption of glycometabolic control after SARS-CoV-2 infection. Nat Metab **2021**; 3:775–85.
37. Mondelli V, Pariante CM. What can neuroimmunology teach us about the symptoms of long-COVID? Oxf Open Immunol **2021**; 2:iqab004.
38. Sluka KA, Clauw DJ. Neurobiology of fibromyalgia and chronic widespread pain. Neuroscience **2016**; 338:114–29.
39. Light AR, White AT, Hughen RW, Light KC. Moderate exercise increases expression for sensory, adrenergic, and immune genes in chronic fatigue syndrome patients but not in normal subjects. J Pain **2009**; 10:1099–112.
40. Sahanic S, Tymoszuk P, Ausserhofer D, et al. Phenotyping of acute and persistent COVID-19 features in the outpatient setting: exploratory analysis of an international cross-sectional online survey [manuscript published online ahead of print 26 November 2021]. Clin Infect Dis **2021**. doi:10.1093/cid/ciab978.
41. Sudre CH, Murray B, Varsavsky T, et al. Attributes and predictors of long COVID. Nat Med **2021**; 27:626–31.