UC Davis
UC Davis Previously Published Works

Title

Symptom Clusters Seen in Adult COVID-19 Recovery Clinic Care Seekers

Permalink

Journal

ISSN

Authors

Danesh, Valerie
Arroliga, Alejandro C
Bourgeois, James A
et al.

Publication Date

2023-02-01

DOI

Peer reviewed

Symptom Clusters Seen in Adult COVID-19 Recovery Clinic Care Seekers

Valerie Danesh, PhD[1,2] , Alejandro C. Arroliga, MD, MSc[3,4],
James A. Bourgeois, OD, MD[5,6], Leanne M. Boehm, PhD, RN[7,8],
Michael J. McNeal, MD, MSc[5,6], Andrew J. Widmer, MD, MPH[5,6], Tresa M. McNeal, MD[5,6],
and Shelli R. Kesler, PhD[2,9]

[1]Center for Applied Health Research, Baylor Scott & White Research Institute, 3434 Live Oak St, Dallas, TX, USA; [2]School of Nursing, University of Texas at Austin, Austin, TX, USA; [3]Baylor Scott & White Health, Dallas, TX, USA; [4]College of Medicine, Baylor College of Medicine, Houston, TX, USA; [5]Baylor Scott & White Health, Temple, TX, USA; [6]College of Medicine, Texas A&M University, College Station, TX, USA; [7]School of Nursing, Vanderbilt University, Nashville, TN, USA; [8]Critical Illness, Brain dysfunction, and Survivorship (CIBS) Center, Vanderbilt University Medical Center, Nashville, TN, USA; [9]Department of Diagnostic Medicine, Dell School of Medicine, University of Texas at Austin, Austin, TX, USA.

BACKGROUND: COVID-19 symptom reports describe varying levels of disease severity with differing periods of recovery and symptom trajectories. Thus, there are a multitude of disease and symptom characteristics clinicians must navigate and interpret to guide care.

OBJECTIVE: To find natural groups of patients with similar constellations of post-acute sequelae of COVID-19 (PASC) symptoms.

DESIGN: Cohort

SETTING: Outpatient COVID-19 recovery clinic with patient referrals from 160 primary care clinics serving 36 counties in Texas.

PATIENTS: Adult patients seeking COVID-19 recovery clinic care between November 15, 2020, and July 31, 2021, with laboratory-confirmed mild (not hospitalized), moderate (hospitalized), or severe (hospitalized with critical care) COVID-19.

MAIN MEASURES: Demographics, COVID illness onset, and duration of persistent PASC symptoms via semi-structured medical assessments.

KEY RESULTS: Four hundred forty-one patients (mean age 51.5 years; 295 [66.9%] women; 99 [22%] Hispanic, and 170 [38.5%] non-White, racial minority) met inclusion criteria. Using a k-medoids algorithm, we found that PASC symptoms cluster into two distinct groups: neuropsychiatric ($N = 186$) (e.g., subjective cognitive dysfunction) and pulmonary ($N = 255$) (e.g., dyspnea, cough). The neuropsychiatric cluster had significantly higher incidences of otolaryngologic ($X^2 = 14.3$, $p < 0.001$), gastrointestinal ($X^2 = 6.90$, $p = 0.009$), neurologic ($X^2 = 441$, $p < 0.001$), and psychiatric sequelae ($X^2 = 40.6$, $p < 0.001$) with more female ($X^2 = 5.44$, $p = 0.020$) and younger age ($t = 2.39$, $p = 0.017$) patients experiencing longer durations of PASC symptoms before seeking care ($t = 2.44$, $p = 0.015$). Patients in the pulmonary cluster were more often hospitalized for COVID-19 ($X^2 = 3.98$, $p = 0.046$) and had significantly higher comorbidity burden ($U = 20800$, $p = 0.019$) and pulmonary sequelae ($X^2 = 13.2$, $p < 0.001$).

CONCLUSIONS: Health services clinic data from a large integrated health system offers insights into the post-COVID symptoms associated with care seeking for sequelae that are not adequately managed by usual care pathways (self-management and primary care clinic visits). These findings can inform machine learning algorithms, primary care management, and selection of patients for earlier COVID-19 recovery referral.

TRIAL REGISTRATION: N/A

INTRODUCTION

Recovery clinics for multidisciplinary evaluation of post-acute COVID-19 patients in the context of long-term COVID-19 symptoms[1] not otherwise fully addressed in the primary care setting have been operating within several health systems in the USA and worldwide since the start of the pandemic.[2] Reports have cataloged more than 60 individual symptoms of post-acute sequelae of COVID-19 (PASC) including multiple neuropsychiatric symptoms, rash, thromboembolic conditions, dyspnea, chest pain, headache, diarrhea, joint pain, and fatigue.[3–6] Prevalence estimates of PASC have been classified into at least 10 organ systems (e.g., pulmonary, cardiovascular, neuropsychiatric, musculoskeletal, endocrine).[5] In epidemiologic studies of patients experiencing PASC, prevalence estimates of specific symptoms include dyspnea (5–40%), chest pain (10–30%), and difficulty in concentration (8–67%).[5–8] In addition, current evidence suggests that PASC persists for varying durations, with unknown resolution bounds.

The myriad of COVID-19 symptom reports including varying criteria (e.g., mild, moderate, severe acute disease) at differing cross-section periods of recovery (e.g., 28 days, 60 days, 6 months) presents a multitude of

characteristics clinicians must navigate and interpret to guide care. Moreover, varying and heterogenous PASC symptoms affect multiple organ systems, and the rate of co-occurrence of these symptoms in individual patients remains unclear. Analyses to inform machine learning methods to inspect PASC symptom co-occurrence patterns could offer clinically relevant insights for clinicians by identifying and anticipating the patterns of symptoms which patients find problematic enough to seek care within a COVID recovery clinic (i.e., the symptoms that matter most to the patients).

A more systematic response to properly address PASC requires rigorous reporting, recognition, and research.[9] In this study, we used cluster analysis with an integrated health system multispecialty telemedicine-based consultation service for post-COVID patients. This consultation service was implemented to monitor and facilitate management of persistent PASC symptoms and support primary care physicians and advanced practice providers in providing ongoing care to these patients. Cluster analysis, a data-driven statistical learning approach that can discover unknown patterns within data, is most often used to partition data into groups that simultaneously have high intra-group similarity and low inter-group similarity. Specifically, we aimed to find natural groups of patients with similar constellations of symptoms spanning six medical specialty-referenced systems (i.e., psychiatric, neurological, cardiovascular, pulmonary, gastrointestinal [GI], otolaryngologic).

METHODS

Setting and Participants

Baylor Scott & White Health (BSWH) COVID-19 Recovery Clinic was established in November 2020 to support more than 160 primary care clinics and 50 hospitals serving rural, suburban, and metropolitan regions across Texas as part of the BSWH System. All referred patients had a history of laboratory-confirmed COVID-19 and active health insurance. Patients were referred by primary care clinicians for persistent PASC symptoms (e.g., beyond two weeks), including dyspnea, loss of taste or smell, and neuropsychiatric symptoms (e.g., cognitive dysfunction, brain fog), or other psychiatric symptoms (e.g., anxiety, depression, sleep disturbance).[10] Visits to the COVID-19 Recovery Clinic last approximately 1 h and include a standard intake evaluation and telemedicine examination by an internal medicine physician or advanced practice provider with substantial experience managing PASC. The COVID-19 Recovery Clinic team management approach for each patient includes concurrent same-day case history and care plan review with up to twenty specialists representing cardiology, gastroenterology, physical medicine and rehabilitation, pulmonary, infectious disease, geriatrics, hematology/oncology, otolaryngology, neurology, and psychiatry via secure electronic sharing to further define and refine an individualized care plan. Based on the initial visit, subsequent laboratory work and/or referral to specialists for follow-up care may be initiated and/or communicated for the primary care provider to consider.

Data Collection and Variables

A semi-structured interview guide with electronic health record review was used during the clinic visit to capture patient-reported symptoms and treatment history (Supplemental Table 1). Data from a 9-month period (November 2020–July 2021) were extracted from clinical narratives in the electronic health record to characterize patient comorbidities present at the time of the COVID-19 clinic visit, COVID-19 illness onset, and the time course of persistent PASC symptoms for 441 consecutive COVID-19 Recovery Clinic patients. A total of 57 distinct PASC symptoms were recorded from clinical documentation. PASC symptoms were subsequently classified by organ system for data analysis. A dichotomous variable for each organ system classification was included (e.g., cardiovascular sequelae: yes/no, psychiatric sequelae: yes/no). These dichotomous variables were used for clustering. Comorbidities were classified using the Charlson Comorbidity Index.[11] No sample size was calculated because all participants with COVID-19 Recovery Clinic visits during the 9-month study period were included. The study was approved by the BSWH Research Institute IRB (#021-145). This report adheres to STROBE/SQUIRE reporting standards.[12]

Data Analysis

Data were first examined visually using graphs and plots for outliers, errors, and normality of quantitative variables. Clustering was conducted using the following 13 features: psychiatric sequelae (1 = yes, 0 = no), neurologic sequelae (1 = yes, 0 = no), GI sequelae (1 = yes, 0 = no), otolaryngologic sequelae (1 = yes, 0 = no), cardiovascular sequelae (1 = yes, 0 = no), pulmonary sequelae (1 = yes, 0 = no), age (years), comorbidity burden measured using the Charlson Comorbidity Index (CCI), sex (1 = female, 0 = male), racial minority (non-White = 1, White = 0), hospitalized for COVID-19 (1 = yes, 0 = no), and Hispanic ethnicity (1 = yes, 0 = no).

Clustering is a statistical technique that allows us to better understand the underlying structure of our data by determining subgroups that exist within the data that are not readily apparent. Clustering can therefore discover latent classes within a sample that may have important diagnostic characteristics for informing treatment planning. Clustering involves four main steps: (1) calculating the distance matrix, which defines the similarity and dissimilarity between data points; (2) choosing a clustering algorithm to apply to the distance matrix to determine the subgroups; (3) determining the number of clusters and then applying the chosen clustering algorithm; and (4) characterizing the resulting clusters to determine their clinical meaning.

Calculating the Distance Matrix. Because most features were categorical, we employed the Gower distance, which calculates the distance matrix for mixed numeric and categorical data by determining the best distance metric for each data level.[13] Specifically, we computed all pairwise dissimilarities (distances) between cases for the 13 features described above.

Choosing the Clustering Algorithm. We selected a *k*-medoids algorithm, also known as partitioning around medoids (PAM). *K*-medoids tend to be more robust to noise and outliers compared to *k*-means, akin to the median being more robust to outliers than the mean.[14]

Determining the Number of Clusters. To determine the optimal number of clusters, *k*, we used the average silhouette width, which is a measure of how similar an observation is to its own cluster compared to its closest neighboring cluster.[15] The silhouette width ranges from − 1 to 1, with higher values being optimal. We calculated the silhouette width for *k* = 2 to 10 for the PAM algorithm, and *k* = 2 was associated with the highest value. We then clustered the Gower distance matrix using PAM with *k* = 2. Since missing data were minimal, all cases were included, and any missing data were ignored when determining the medoid.

Characterizing the Clusters. Descriptive statistics including percentage, mean, standard deviation, median, and interquartile range were used to describe cluster characteristics, as appropriate. Characteristics including the features listed above (primary outcomes), as well as specific symptoms within each physiologic system that were not used in clustering (secondary outcomes), were compared between clusters using chi square test (X^2), *t* test (*t*), or Mann-Whitney test (*U*), as appropriate. To evaluate potential relationships among symptoms, exploratory, two-tailed Spearman correlations were conducted between symptoms that were significantly different between clusters, ignoring correlations within the same medical specialty system. We were particularly interested in the relationships that psychiatric symptoms had with other symptoms. Correlations were conducted within each cluster, separately. Alpha level for all analyses was set at $p < 0.05$. The *p* values for secondary and exploratory outcomes were corrected for multiple comparisons using the false discovery rate (FDR).[16] Data visualization and analyses were performed using the R Statistical Package version 4.1.2 (R Foundation, Vienna, Austria) including the "ggplot2" and "cluster" libraries.

RESULTS

This study includes 441 consecutive patients referred by primary care clinicians to COVID-19 Recovery Clinic services who completed a telehealth evaluation for persistent PASC symptoms. The mean (standard deviation) age was 51.5 (14.9) years, and there were 295 (66.9%) females and 145 (32.9%) males. In general, patients were seeking care for PASC symptoms approximately 2 months after initial COVID-19 symptom onset (median 58 days, interquartile range 61 days). Patients had up to 13 symptoms at COVID-19 onset and up to 16 symptomatic sequelae identified at the time of the COVID-19 Recovery Clinic visit. *K*-medoids revealed two distinct clusters of patients (Fig. 1). Cluster 1 was comprised of *n* = 186 patients and cluster 2 had *n* = 255 patients.

Cluster 1

Cluster 1 had significantly higher incidences of otolaryngologic ($X^2 = 14.3$, $p < 0.001$), GI ($X^2 = 6.90$, $p = 0.009$), neurologic ($X^2 = 441$, $p < 0.001$), and psychiatric sequelae ($X^2 = 40.6$, $p < 0.001$, Table 1). This cluster had more female patients ($X^2 = 5.38$, $p = 0.020$) and younger age ($t = 2.39$, $p = 0.017$), and reported a longer duration of PASC symptoms before seeking care ($t = 2.44$, $p = 0.015$) compared to cluster 2 (Table 1).

Individual otolaryngologic and GI symptoms were not significant after correction for multiple comparisons (Table 2). Significant individual neurological symptoms included brain fog ($X^2 = 154$, $p < 0.001$), confusion ($X^2 = 14.03$, $p < 0.001$), concentration difficulties (($X^2 = 60.31$, $p < 0.001$), headache ($X^2 = 144$, $p < 0.001$), memory loss ($X^2 = 58.66$, $p < 0.001$), paresthesia ($X^2 = 14.00$, $p < 0.001$), word finding difficulty ($X^2 = 21.29$, $p < 0.001$), and other ($X^2 = 24.24$, $p < 0.001$) (Table 2). Significant psychiatric symptoms included anxiety ($X^2 = 31.08$, $p < 0.001$, corrected), depressed mood ($X^2 = 8.29$, $p = 0.004$, corrected), insomnia ($X^2 = 13.61$, $p < 0.001$, corrected), and other ($X^2 = 6.93$, $p = 0.018$, corrected) (Table 2). Other psychiatric symptoms included substance abuse, survivor guilt, bizarre dreams, and panic attacks. Considering these characteristics, we qualitatively labeled cluster 1 the "Neuropsychiatric Cluster."

Within cluster 1, depressed mood was correlated with memory loss, but this was not significant after correction for multiple comparisons. No other correlations were observed between psychiatric and other symptoms (Fig. 2).

Cluster 2

Cluster 2 patients were more often hospitalized for COVID-19 ($X^2 = 3.98$, $p = 0.046$) and had significantly higher comorbidity burden ($U = 20800$, $p = 0.019$) and pulmonary sequelae ($X^2 = 13.2$, $p < 0.001$) (Table 1). The cluster 2 difference in individual pulmonary symptoms included cough ($X^2 = 12.37$, $p < 0.001$, corrected) and dyspnea ($X^2 = 20.39$, $p < 0.001$, corrected) (Table 2). Cardiovascular sequelae were not significant after correction for multiple corrections (Table 2) in either cluster. Considering these characteristics, we qualitatively labeled cluster 2 the "Pulmonary Cluster."
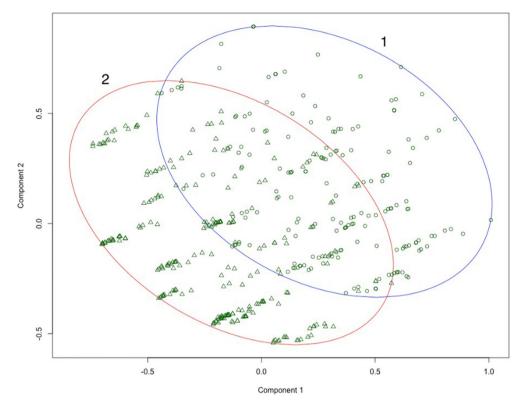
Figure 1 Bivariate cluster plot. A 2-dimensional representation of the cluster solution using principal components analysis. These components explain 31.14% of the point variability.

Cluster 2 differed only in terms of pulmonary symptoms, so no correlations were conducted.

DISCUSSION

In this study, we evaluated symptom clusters among adults with laboratory-confirmed COVID-19 with persistent PASC seeking care through COVID-19 Recovery Clinic evaluations. In our sample of adults with mild (not hospitalized), moderate (hospitalized), and severe COVID-19 (hospitalized with critical care), we found that symptoms cluster into two distinct groups: (1) neuropsychiatric (e.g., subjective cognitive dys-function, dizziness, memory problems) and (2) pulmonary (e.g., dyspnea, cough). While other epidemiologic reports and cluster analyses have contributed to discerning that PASC is comprised of symptoms referable to multiple organ systems with varied time courses, our approach characterizes symptom patterns prompting patients to seek and receive specialized COVID-19 recovery care.

Previous reports suggest that PASC disproportionately affects patients in middle adulthood, Caucasians, and females, with similar clusters of pulmonary and multisystem presenta-tions. Generalizability has been limited by reliance on conve-nience samples that were disproportionately female and not ethnically diverse, and which under-represented older-age

Table 1 Primary Cluster Characteristics

Variables*	Cluster 1 (N = 186)	Cluster 2 (N = 255)	Statistic	p value
Female	136 (73%)	159 (63%)	$X^2 = 5.38$	0.020
Age	Mean = 50 ± 13	Mean = 53 ± 16	$t = 2.39$	0.017
Racial minority	71 (39%)	99 (39%)	$X^2 = 0.008$	0.928
Hispanic ethnicity	43 (23%)	56 (22%)	$X^2 = 0.057$	0.812
Otolaryngologic sequelae	69 (37%)	53 (21%)	$X^2 = 14.3$	< 0.001
Cardiovascular sequelae	48 (26%)	49 (19%)	$X^2 = 2.72$	0.100
Pulmonary sequelae	114 (61%)	197 (77%)	$X^2 = 13.2$	< 0.001
Gastrointestinal sequelae	39 (21%)	30 (12%)	$X^2 = 6.90$	0.009
Neurologic sequelae	186 (100%)	0 (0.0%)	$X^2 = 441$	< 0.001
Psychiatric sequelae	75 (40%)	35 (14%)	$X^2 = 40.6$	< 0.001
Symptom duration (months)	Median = 63, IQR = 75	Mean = 50, IQR = 53	$U = 25254$	0.005
CCI	Median = 0, IQR = 1.75	Median = 1, IQR = 2	$U = 20800$	0.019
Hospitalized	43 (23%)	81 (32%)	$X^2 = 3.98$	0.046

CCI Charlson Comorbidity Index, IQR interquartile range
*These variables were included in the clustering algorithm

Table 2 Secondary Cluster Characteristics

Variable*	Cluster 1 (N = 186)	Cluster 2 (N = 255)	Chi-squared (X^2)	p	p FDR corrected
Otolaryngologic					
Ageusia	18 (9.7%)	20 (7.8%)	0.60	0.498	0.543
Anosmia	30 (16%)	32 (13%)	1.14	0.285	0.428
Dizziness	4 (8.6%)	11 (4.3%)	3.44	0.064	0.216
Dysphagia	4 (2.2%)	0 (0.0%)	5.53	0.019	0.114
Mouth sores	1 (0.54%)	0 (0.0%)	1.37	0.241	0.413
Otalgia	1 (0.54%)	3 (1.2%)	0.488	0.485	0.543
Rhinorrhea	1 (0.54%)	1 (0.39%)	0.050	0.822	0.822
Sneezing	0 (0.0%)	1 (0.39%)	0.731	0.393	0.524
Sore throat	7 (3.8%)	3 (1.2%)	3.25	0.072	0.216
Tinnitus	5 (2.7%)	0 (0.0%)	6.93	0.008	0.096
Vision disturbance	6 (3.2%)	3 (1.2%)	2.26	0.133	0.300
Other	14 (7.5%)	11 (4.3%)	2.08	0.150	0.300
Gastrointestinal					
Anorexia	10 (5.4%)	8 (3.1%)	1.38	0.241	0.626
Diarrhea	12 (6.5%)	14 (5.5%)	0.179	0.672	0.896
Gastroesophageal reflux symptoms	3 (1.6%)	4 (1.6%)	0.001	0.971	0.971
Nausea	15 (8.1%)	8 (3.1%)	5.28	0.022	0.176
Vomiting	4 (2.2%)	1 (0.39%)	2.97	0.085	0.340
Abdominal pain	1 (0.54%)	4 (1.6%)	1.02	0.313	0.626
Weight loss	4 (2.2%)	6 (2.4%)	0.020	0.888	0.971
Other	5 (2.7%)	4 (1.6%)	0.674	0.412	0.659
Neurological					
Expressive aphasia	1 (0.54%)	0 (0.0%)	1.37	0.241	0.241
Bell's palsy	3 (1.6%)	0 (0.0%)	4.14	0.042	0.063
Brain fog	89 (48%)	0 (0.0%)	153.87	< 0.001	< 0.001
Confusion	10 (5.4%)	0 (0.0%)	14.03	< 0.001	< 0.001
Impaired concentration	40 (21%)	0 (0.0%)	60.31	< 0.001	< 0.001
Headache	85 (46%)	0 (0.0%)	144.36	< 0.001	< 0.001
Hemiplegia	2 (1.1%)	0 (0.0%)	2.75	0.097	0.121
Memory loss	39 (21%)	0 (0.0%)	58.66	< 0.001	< 0.001
Neuropathy	2 (1.1%)	0 (0.0%)	2.75	0.097	0.121
Facial pain	1 (0.54%)	0 (0.0%)	1.37	0.241	0.240
Neuropathic pain	6 (3.2%)	0 (0.0%)	8.34	0.004	0.007
Unilateral paralysis	1 (0.54%)	0 (0.0%)	1.37	0.241	0.241
Paresthesia	10 (5.4%)	0 (0.0%)	14.003	< 0.001	< 0.001
Word finding difficulty	15 (8.1%)	0 (0.0%)	21.29	< 0.001	< 0.001
Other	17 (9.1%)	0 (0.0%)	24.24	< 0.001	< 0.001
Psychiatric					
Anxiety	40 (22%)	11 (4.3%)	31.08	< 0.001	< 0.001
Apathy	3 (1.6%)	1 (0.39%)	1.78	0.182	0.205
Bereavement	3 (1.6%)	1 (0.39%)	1.78	0.182	0.205
Depressed mood	18 (9.7%)	8 (3.1%)	8.29	0.004	0.012
Insomnia	43 (21%)	26 (10%)	13.61	< 0.001	0.001
Irritability	2 (1.1%)	0 (0.0%)	2.75	0.097	0.146
Mood lability	5 (2.7%)	1 (0.39%)	4.23	0.040	0.072
Posttraumatic stress disorder	1 (0.54%)	2 (0.78%)	0.097	0.756	0.756
Other	5 (2.7%)	0 (0.0%)	6.93	0.008	0.018
Pulmonary					
Bronchitis	0 (0.0%)	1 (0.39%)	10.731	0.393	0.822
Congestion	9 (4.8%)	9 (3.5%)	0.471	0.493	0.822
Cough	34 (18%)	85 (33%)	12.37	< 0.001	< 0.001
Dyspnea	96 (52%)	185 (73%)	20.39	< 0.001	< 0.001
Suspected obstructive sleep apnea	1 (0.54%)	1 (0.39%)	0.050	0.822	0.822
Pneumonia	6, 3.2%	7 (2.7%)	0.087	0.768	0.822
Wheezing	6 (3.2%)	11 (4.3%)	0.343	0.558	0.822
Other	12 (6.5%)	18 (7.1%)	0.062	0.803	0.822
Cardiovascular					
Chest pain	27 (15%)	50 (20%)	1.94	0.164	0.410
Hypotension	1 (0.54%)	2 (0.78%)	0.097	0.756	0.880
Hypertension	4 (2.1%)	5 (2.0%)	0.019	0.880	0.880
Palpitations	40 (22%)	36 (14%)	4.12	0.042	0.210
Other	8 (4.3%)	13 (5.1%)	0.151	0.698	0.880

FDR false discovery rate correction for multiple comparisons
*These variables were not included in the clustering algorithm but were examined to further characterize the identified cluster

individuals.[17] Our sample of COVID-19 laboratory test confirmed patient presentations offers diversity in age, sex, and severity of illness of COVID-19 not present in cluster analyses in other US and non-US samples to date.[5,6,18] Results of natural groups of patients with similar presentations are presented in the context of a multitude of PASC phenotypes and differing courses of recovery. These natural groups inform primary care clinicians of the patients who seek COVID-19 Recovery Clinic follow-up for PASC.

The neuropsychiatric cluster and pulmonary cluster represent two distinct symptom groups with maximum intra-group similarity and maximum inter-group dissimilarity. The
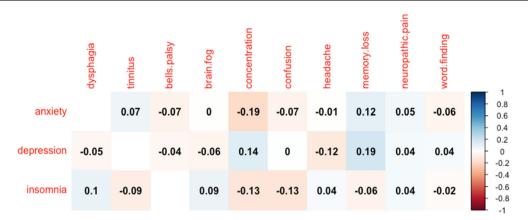
Figure 2 Correlation heatmap for cluster 1. Pairwise correlation coefficients shown for psychiatric symptoms and neurologic symptoms. Psychiatric symptoms were not significantly associated with other cluster 1 symptoms. Color map represents correlation coefficients, which are also displayed within the heatmap grids.

emergence of two symptom clusters from a sample of 441 patient assessments supports emerging evidence that PASC may be more than one syndrome.[9,18] Anaya demonstrated two clusters,[6] and our increased sample size and diversity offers more insight by building on these characteristics. Other analyses apply multiple logistic regression in pre-defined organ system clusters,[7] which have linear constraints in separating sets of clusters.

The neuropsychiatric cluster had a higher prevalence of neurological, psychiatric, and otolaryngological-related concerns and a longer duration of symptoms prior to seeking COVID-19 Recovery Clinic care, and these patients were more often younger and female. Within this group, 100% endorsed at least one neurological symptom and were significantly more likely to have neurological symptoms compared to cluster 2 patients. Importantly, anxiety, depressed mood, and insomnia were not significantly associated with other cluster 1 symptoms (e.g., headache, memory loss, concentration problems). These results suggest that mood symptoms alone are an insufficient sole explanation for PASC-related neurological symptoms. Women more often endorse symptoms of mood disturbance compared to men, and there were statistically more women in cluster 1. However, this represented only a 10% difference, so does not likely reflect a true gender effect. Patients in cluster 1 were significantly younger, on average, compared to patients in cluster 2. Previous studies, including our own, have indicated a relationship between younger age and increased PASC-related neuropsychiatric symptoms.[19,20] However, the age difference between the clusters was only 3 years, on average, which is not likely clinically relevant. Further research is required to determine the socio-demographic and physiologic factors that contribute to PASC symptom clusters.

COVID-19 Recovery Clinic care seekers in cluster 2 (pulmonary) are representative of higher comorbidity burdens and seeking care for pulmonary symptoms persisting more than a month after initial illness. Pulmonary sequelae clusters have been previously discerned from self-reported symptom report surveys but are constrained by overlapped distribution across multiple clusters with pulmonary symptoms highlighted in four of six time series clusters.[21] Other cluster analysis approaches have employed clustering self-reported symptoms pre-defined by organ system to develop a pulmonary cluster. In contrast to these approaches, our analysis relies on expert clinician interviews and assessments to classify PASC, and presents a distinct cluster set of persistent pulmonary symptoms.

Cluster 2 patients did not endorse any neurologic symptoms suggesting that this sample of patients was best separated by neurologic symptoms. It is surprising that patients with higher comorbidity and greater pulmonary symptom burden would not have neurologic symptoms. It is possible that these symptoms had yet to develop in this subgroup given that they sought care earlier than cluster 1 patients. Neurologic symptoms may also develop more insidiously than pulmonary symptoms during post-COVID-19 recovery.[22] It is also possible that cluster 2 patients were not reporting their neurologic symptoms due to the prominence of pulmonary symptoms which may have been more urgent at the time of consultation. In a review of 197 brain autopsies of patients who succumbed to SARS-CoV-2 illness, the most common finding was hypoxic injury followed by ischemia, hemorrhage, and reactive astrocytosis and gliosis.[23] These are findings typical of systemic inflammation. This suggests coexistence of neurologic involvement related to hypoxia from pulmonary involvement and systemic inflammatory factors to be present in many patients even if they do not report neurologic symptoms. Patients with significant respiratory distress may focus on these symptoms primarily and not report on less dramatically impairing struggles with cognitive performance and mood, which may lead to underreporting of CNS symptoms dependent on the depth of clinician questioning. Neurologic symptoms may then become more distressing as patients resume normal activities. Further research is required to determine the timeline of symptom development in PASC subgroups.

Understanding PASC is important to guiding investigations into underlying pathophysiology tied to sites of inflammatory pathways and/or evolving tissue damage,[24] particularly in the

context of myocardial inflammation and neurological sequelae.[25] Mechanisms of cognitive impairment attributed to PASC could be tied to infectious, toxic, vascular, and/or metabolic pathways.[25] A strength of this study is that patients seeking care at different intervals after initial onset are included. This reflects important pragmatic pattern-based insights, to offer overarching symptom clusters that are not tied to a traditional 3- or 6-month follow-up endpoint in epidemiologic research designs. The data also reflects direct patient-provider interactions and assessments conducted via telemedicine and does not rely on survey completion by a convenience sample. Furthermore, data collected were elicited by real-time reporting and recording of persistent sequelae at the time of the clinic visit, which minimizes retrospective recall bias present in other studies. Consequently, our approach represents a high fidelity of data collection with more homogenous groups compared with other cluster analysis reports including diagnostic/antibody negative or untested self-report participants.[5]

Limitations

There are limitations inherent to the design and data drawn from a single integrated health system. First, we may be underpowered and not be able to discern whether there are additional clusters or other differences between clusters in persistent sequelae (e.g., differences associated with self-reported cognitive dysfunction/brain fog, patients reporting that they are unable to return to work due to PASC symptoms). Future research to perform cross-validation with independent data from other COVID-19 recovery clinics is needed. Second, given that "other" neuropsychiatric sequelae were among the largest effects, the classification of persistent sequelae might not be sufficiently consistent. Persistent sequelae were drawn from patient responses to standard-of-care telehealth-enabled assessments performed by expert clinicians using a semi-structured interview guide, and standardized cognitive assessments and inventories were not performed. However, the interactive standard-of-care comprehensive interviews allowed for expertise-informed symptom assessment and were not confined to a self-report checklist of symptoms, which contributes to a more nuanced analysis of symptoms in this analysis. Third, we are unable to predict which sequelae will be persistent after severe illness or delineate differences related to vaccination status. However, an advantage of this dataset is that it represents the full continuum of care seekers and is not restricted to vaccination status or severe disease. Some hypotheses regarding persistent COVID-19 morbidity include speculation that brain injury could explain some neuropsychiatric sequelae,[26] so neuroimaging and other diagnostic evaluation may be needed as a pathway to advancing knowledge around underlying pathophysiology to further understand phenotypes associated with symptom clusters. However, routine clinical neuroimaging may be insufficiently sensitive to assess for subtle evidence of prolonged COVID-19 symptoms.

CONCLUSION

Health services clinic data from a large integrated health system offers insights into the PASC symptoms associated with care seeking for persistent sequelae that are not adequately managed by usual care pathways (self-management at home and/or primary care clinic visits). Insights into the epidemiologic prevalence of sequelae are important for establishing the characteristics of the illness caused by SARS-CoV-2 virus. Of critical importance to clinicians and administrators are insights into the prevalence of persistent symptoms that are not only present, but which persist to such a degree as to prompt patients to access clinical services for COVID-19 recovery and rehabilitation. Globally endorsed definitions of COVID-associated illness have been developed with an explicit intent of evolving as epidemiologic and health services research insights advance knowledge on prevalence, risk factors, and outcomes.[4] Pairing comprehensive epidemiologic data with real-world care-seeking behaviors of patients experiencing PASC delivers a balanced view for policymakers on incidence/prevalence (epidemiologic data) and healthcare utilization (health services clinic data) to inform resource allocation, tailor recovery and rehabilitation program design and implementation, and inform clinical trial designs and prioritizations. Epidemiologic research on prevalence, risk factors, and outcomes is important to public health and health services administration, as are parallel investigations cultivating machine learning–enabled pragmatic insights discerned from COVID-19 Recovery Clinic care seekers to highlight the clusters of symptoms that matter most to patients.

Corresponding Author: *Valerie Danesh, PhD; Center for Applied Health Research, Baylor Scott & White Research Institute, 3434 Live Oak St, Dallas, TX 75204, USA (e-mail: Valerie.Danesh@BSWHealth. org).*

Declarations:

REFERENCES

1. **Danesh V, Boehm LM, Eaton TL, et al.** Characteristics of post-ICU and post-COVID recovery clinics in 29 US health systems. *Critical Care Explorations.* 2022;4(3):e0658-e0668.
2. **Greenhalgh T, Knight M, A'Court C, Buxton M, Husain L.** Management of post-acute covid-19 in primary care. *BMJ.* 2020;370:m3026.
3. **Nalbandian A, Sehgal K, Gupta A, et al.** Post-acute COVID-19 syndrome. *Nat Med.* 2021.
4. **Soriano JB, Murthy S, Marshall JC, Relan P, Diaz JV**, WHO Clinical Case Definition Working Group on Post-COVID-19. A clinical case definition of post-COVID-19 condition by a Delphi consensus. *Lancet Infect Dis.* 2021.

5. **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.

6. **Anaya J-M, Rojas M, Salinas ML, et al.** Post-COVID syndrome: a case series and comprehensive review. *Autoimmun Rev.* 2021;20(11):102947.

7. **Wong-Chew RM, Rodriguez Cabrera EX, Rodriguez Valdez CA, et al.** Symptom cluster analysis of long COVID-19 in patients discharged from the Temporary COVID-19 Hospital in Mexico City. *Ther Adv Infect Dis.* 2022;9:20499361211069264.

8. **Crook H, Raza S, Nowell J, Young M, Edison P.** Long covid-mechanisms, risk factors, and management. *BMJ.* 2021;374:n1648.

9. **Alwan NA.** The road to addressing long Covid. *Science.* 2021;373(6554):491-493.

10. **Danesh V, Arroliga AC, Bourgeois JA, Widmer AJ, McNeal MJ, McNeal TM.** Post-acute sequelae of COVID-19 in adults referred to COVID recovery clinic services in an integrated health system in Texas. *Proc (Bayl Univ Med Cent).* 2021;34(6):645-648.

11. **Charlson ME, Pompei P, Ales KL, MacKenzie CR.** A new method of classifying prognostic comorbidity in longitudinal studies: development and validation. *Journal of Chronic Diseases.* 1987;40(5):373-383.

12. **Connell L, MacDonald R, McBride T, et al.** Observational studies: getting clear about transparency. *Plos Medicine.* 2014;11(8).

13. **Gower JC.** A general coefficient of similarity and some of its properties. *Biometrics.* 1971;27(4):857-971.

14. **Soni KG, Patel A.** Comparative analysis of K-means and K-medoids algorithm on IRIS data. *Int J Computa Intel Res.* 2017;13(5):899-906.

15. **Kaufman L, Rousseeuw PJ.** *Finding Groups in Data: an Introduction to Cluster Analysis.* Vol 344: John Wiley & Sons; 2009.

16. **Benjamini Y, Yekutieli D.** The control of the false discovery rate in multiple testing under dependency. *Annals Stat.* 2001;29(4):1165-1188.

17. **Sudre CH, Murray B,** Varsavsky T, et al. Attributes and predictors of long COVID. *Nat Med.* 2021.

18. **Caspersen IH, Magnus P, Trogstad L.** Excess risk and clusters of symptoms after COVID-19 in a large Norwegian cohort. In. medRxiv: Cold Spring Harbor Laboratory; 2021.

19. **Henneghan AM, Lewis KA, Gill E, Kesler SR.** Cognitive impairment in non-critical, mild-to-moderate COVID-19 survivors. *Front Psychol.* 2022;13:770459.

20. **Varma P, Junge M, Meaklim H, Jackson ML.** Younger people are more vulnerable to stress, anxiety and depression during COVID-19 pandemic: a global cross-sectional survey. *Prog Neuropsychopharmacol Biol Psychiatry.* 2021;109:110236.

21. **Sudre CH, Lee KA, Lochlainn MN, et al.** Symptom clusters in COVID-19: a potential clinical prediction tool from the COVID Symptom Study app. *Sci Adv.* 2021;7(12).

22. **Uversky VN, Elrashdy F, Aljadawi A, Ali SM, Khan RH, Redwan EM.** Severe acute respiratory syndrome coronavirus 2 infection reaches the human nervous system: How? *J Neurosci Res.* 2021;99(3):750-777.

23. **Maiese A, Manetti AC, Bosetti C, et al.** SARS-CoV-2 and the brain: a review of the current knowledge on neuropathology in COVID-19. *Brain Pathol.* 2021;31(6):e13013.

24. **Huang Y, Pinto MD, Borelli JL, et al.** COVID Symptoms, Symptom Clusters, and Predictors for Becoming a Long-Hauler: Looking for Clarity in the Haze of the Pandemic. In. medRxiv: Cold Spring Harbor Laboratory; 2021.

25. **Sieracka J, Sieracki P, Kozera G, et al.** COVID-19 - neuropathological point of view, pathobiology, and dilemmas after the first year of the pandemic struggle. *Folia Neuropathol.* 2021;59(1):1-16.

26. **Douaud G, Lee S, Alfaro-Almagro F, et al.** SARS-CoV-2 is associated with changes in brain structure in UK Biobank. *Nature.* 2022.