Predictive characteristics of prolonged symptoms and seroconversion in ambulatory patients recovering from SARS-CoV-2 infection

Vamsi P. Guntur MD, MSc, Brian Modena MD, MSc, Claudia Onofrei MD, MSc, Shu-Yi Liao MD, MPH, ScD, Pearlanne Zelarney MS, Jared J. Eddy MD, MPhil, MSc, Rebecca Keith MD, Rachel DeCosta NP, Irina Petrache, MD**, Nir Goldstein, MD**

Abstract

Introduction: With an increasing number of ambulatory visits for acute COVID-19 follow-up, we set out to characterize and identify clinical predictors of prolonged symptoms and antibody seroconversion. We hypothesized that patients who present with a high symptom burden are more likely to have prolonged post-acute sequelae of COVID-19 (PASC).

Methods: All adults with confirmed SARS-CoV-2 infection evaluated at a single ambulatory center between April–September 2020 were studied retrospectively using a logit model and ANOVA, and the importance of variables associated with prolonged symptoms and seroconversion was determined by machine learning methodology.

Results: The most common initial symptoms of 276 subjects were fatigue, dyspnea, cough, fever, and myalgia, with ~30% experiencing all five. Those with prolonged sequelae (>4 weeks; PASC) reported higher initial symptom burden compared to those without PASC (mean 8.2 vs. 3.3 symptoms, p < 0.0001). Anosmia (odds ratio, OR 23.0), myalgia (OR 12.8), and dyspnea (OR 10.8) had the highest predictive values for PASC. Neither lung function nor pre-existing lung disease correlated with PASC pulmonary symptoms (p = 0.17, p = 0.5, respectively). Natural post-COVID-19 seroconversion rate was 78%, with the male gender having higher- and corticosteroid treatment and elevated creatinine having lower seroconversion.

Conclusion: Ambulatory patients display a broad range of symptoms following acute COVID-19. A high initial symptom burden may predict PASC development. In unvaccinated, antibody seroconversion may be influenced by gender, corticosteroid use, and renal function.

Keywords: Ambulatory, SARS-CoV-2, Post-COVID-19, Post-acute sequelae of COVID-19 (PASC); lung disease; symptom burden; inhaled corticosteroids.

Introduction

Declared a pandemic in March 2020, COVID-19 led to unforeseen morbidity/mortality and remains a public health crisis. Initial efforts focused on management of severe illness, while comprehensive outpatient care for COVID-19 survivors was sporadic. An unmet need to manage convalescent COVID-19 outpatients remains, as most patients do not require hospitalization, but are at risk for developing post-acute sequelae of COVID-19 (PASC). To address this need, we developed a dedicated outpatient Center for Post-Covid Care and Recovery (CPCCR).

Persistence of symptoms for more than 4 weeks following acute COVID-19 (PASC, “long-COVID”) poses a significant healthcare burden. While vaccination efforts...
mitigate severe COVID-19, we expect that milder, breakthrough, and repeat infections due to waning immunity and development of variants will continue. Better information to predict who is at risk for PASC will allow for improved resource planning. Clinical predictors for seroconversion following natural infection may also inform targeted vaccination/booster efforts particularly in areas with diminished access or hesitancy to vaccination.

**Methods**

**Study design**

This retrospective cohort analysis of a prospective, observational study characterized all outpatient subjects >18 years old seen at National Jewish Health (NJH, Denver, CO), who tested positive for SARS-CoV-2 by Polymerase Chain Reaction (PCR) and completed a customized symptom questionnaire (CPCCR Worksheet, Appendix 1), between April 15–September 20, 2020. No subjects were excluded. Data were extracted from the electronic medical record (EMR) Allscripts Analytics™, using dataSCOUT™ software. The Biomedical Research Alliance of New York (BRANY) institutional review board (IRB) approved the study, removing informed consent requirement due to low risk and retrospective study design. Primary endpoints were symptom type, frequency, and duration. Secondary endpoints were clinical characteristics associated with symptom duration and post-infection antibody seroconversion. We analyzed the association of persistent pulmonary symptoms with respiratory dysfunction, assessed by pulmonary function testing. Symptoms were assessed during initial evaluation via CPCCR Worksheet (Appendix 1), and by longitudinal symptom tracking recorded at follow-up visits.

**CPCCR design**

A multidisciplinary ambulatory program, CPCCR was established in March 2020 to provide post-acute COVID-19 care. Following an initial visit staffed by a pulmonologist, multi-disciplinary consultations are provided based on appropriate clinical scenarios (Appendix 1). Ancillary services include Pulmonary Rehabilitation, Physical/Occupational Therapy, and Speech/Language Therapy. During the first 6 months, CPCCR visits were categorized into 2 phases based on the level of infection control needed:

CPCCR Phase 1, potentially infectious patients, i.e., prior to clearing a 2 PCR test-based strategy (used early in the pandemic) or with persistent PCR-positivity (6) were examined by a limited number of providers donning CDC-recommended personal protective equipment (PPE). Patients’ visits were conducted in negative pressure rooms, with 30 minutes between patient visits, based on 99.9% airborne contaminant removal efficiency for rooms with 15 hourly air exchanges. Diagnostic testing included SARS-CoV-2 PCR nasopharyngeal swabs; electrocardiogram (ECG); portable chest radiographs (CXR); and SARS-CoV-2 serology, complete blood counts (CBC), metabolic profiles (BMP/CMP), D-dimer (DD), C-reactive protein (CRP), troponin I (Trop I), brain natriuretic peptide (BNP), ferritin, and lactate dehydrogenase (LDH).

CPCCR Phase 2, low/non-infectious patients, i.e. demonstrating 2 sequential-negative PCR tests (before July 17, 2020) or who met CDC guidelines of time and symptom-based criteria of low infectivity (after July 17, 2020) were examined by any appropriate provider donning surgical masks. Visits were conducted in usual clinical areas and testing was ordered as clinically appropriate (Appendix 2).

**Statistical analysis**

EMR-extracted data were compiled using dataScout™ and data wrangling was performed using tidyR package. Overdispersed data R package was used to create a logit model. Between-group comparisons of continuous variables were performed by pooled t-test or analysis of variance (ANOVA) for two-group or multi-group comparisons using JMP software.

Important variables associated with seroconversion were determined using machine learning methods. Elastic net was applied for variable selection to address over-regularization by balancing between LASSO and ridge penalty, whereby Alpha = 1 indicates full LASSO model, and Alpha = 0 indicates full ridge model. The optimal Alpha level was determined by using cross-validation. Subjects were randomly divided into an 80:20 ratio for training and testing sets,
respectively. Alpha coefficients were first obtained using the training set, and the area under the receiver operating characteristic curve (AUC) was calculated using the test set. Acceptable discrimination was defined as AUC > 0.70. The importance of the selected variable was estimated using the variable importance plots (VIP). Analysis was performed using the R package, glmnet. Nine candidate variables included OCS use, levels of circulating WBC, lymphocytes, platelets, hemoglobin, creatinine, glucose, age, and sex.

A similar approach was applied to identify important symptom variables associated with prolonged (≥4 weeks) symptoms, of which 14 were included in the machine learning model: abdominal pain, anosmia, chest discomfort, cough, diarrhea, dysgeusia, fatigue, fever, headache, myalgia, nasal congestion, nausea, dyspnea, and vomiting.

Results

Subject Characteristics

Of 276 subjects evaluated in CPCCR during the 6-month study period and included in analysis (Table 1), none were excluded. The majority were female (65%) and non-Hispanic white (59%). Smoking status was available on 213 subjects, of whom 2.8% were active and 32% were former smokers. Prior health records were available in 161 subjects (58%), showing that 70% had lung disease (mostly asthma, followed by interstitial lung disease and COPD) (Table 1). Most (153 subjects) had normal PFTs prior to COVID-19 diagnosis: pre-bronchodilator (pre-BDR) forced expiratory volume in 1 second mean %predicted (FEV₁, 89.7%), forced vital capacity mean %predicted, (FVC; 91.3%), total lung capacity mean %predicted (TLC, 98.1%) and diffusion capacity mean %predicted (DLCO, 103.4%).

Clinical Symptoms

Most (272) subjects completed the CPCCR worksheet, reporting abdominal pain, dyspnea, anosmia, blood clots, chest discomfort, cough, diarrhea, dysgeusia, fatigue, and fever (Appendix 1). The most reported symptoms were fatigue, dyspnea, cough, fever, and myalgia (Figure 1). Less common symptoms included headache, nasal, or GI symptoms. Six subjects reported blood clots.

Subjects averaged ~7 symptoms (mean = 6.94; SD = 2.88) at presentation; half reported ≥4 symptoms; one-third reported all 5 of the most common symptoms (fatigue, dyspnea, cough, fever, and myalgia); and only a minority reported a single symptom (Figure 2).

Table 1. Demographics and Clinical Characteristics of Subjects Seen in Respiratory Recovery Clinic (April 15, 2020 to September 19, 2020)

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>COVID (+) Subjects</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total number of subjects</td>
<td>276</td>
</tr>
<tr>
<td>Health records available</td>
<td>161</td>
</tr>
<tr>
<td>Age (years), mean + SD</td>
<td>53 + 16</td>
</tr>
<tr>
<td>Male/Female, n</td>
<td>97/179</td>
</tr>
<tr>
<td>Race, n (% total)</td>
<td></td>
</tr>
<tr>
<td>White</td>
<td>163 (59%)</td>
</tr>
<tr>
<td>Black or African American</td>
<td>18 (6%)</td>
</tr>
<tr>
<td>American Indian or Alaska Native</td>
<td>2 (0.7%)</td>
</tr>
<tr>
<td>Native Hawaiian or Pacific Islander</td>
<td>1 (0.4%)</td>
</tr>
<tr>
<td>Unknown</td>
<td>66 (24%)</td>
</tr>
<tr>
<td>Declined</td>
<td>26 (9%)</td>
</tr>
<tr>
<td>Ethnicity, n (% total)</td>
<td></td>
</tr>
<tr>
<td>Hispanic or Latino</td>
<td>38 (14%)</td>
</tr>
<tr>
<td>Non-Hispanic</td>
<td>164 (59%)</td>
</tr>
<tr>
<td>Unknown</td>
<td>53 (20%)</td>
</tr>
<tr>
<td>Declined</td>
<td>22 (8%)</td>
</tr>
<tr>
<td>Smoking status, n (% total)</td>
<td></td>
</tr>
<tr>
<td>Current</td>
<td>6 (2%)</td>
</tr>
<tr>
<td>Past</td>
<td>69 (25%)</td>
</tr>
<tr>
<td>Never</td>
<td>138 (50%)</td>
</tr>
<tr>
<td>Unavailable data</td>
<td>63 (23%)</td>
</tr>
<tr>
<td>Chronic respiratory disease, n (% total)</td>
<td></td>
</tr>
<tr>
<td>COPD</td>
<td>14 (9%)</td>
</tr>
<tr>
<td>Asthma</td>
<td>88 (54%)</td>
</tr>
<tr>
<td>ILD</td>
<td>21 (13%)</td>
</tr>
</tbody>
</table>

SD = standard deviation, COVID = Coronavirus disease 2019, n = number, COPD = chronic obstructive pulmonary disease, ILD = interstitial lung disease.
PFT did not correlate with pulmonary symptoms. Interestingly, patients reporting cough had not only had no significant difference in FEV\textsubscript{1} (90.6\% vs. 84.4\%, respectively; p = 0.17), but had a higher FEV\textsubscript{1}/FVC (0.792 vs. 0.737, p = 0.01) when compared to those without cough. Those with underlying lung disease were as likely to report pulmonary symptoms: (92\% vs. 88\%, p = 0.5) as those without lung disease.
Follow-up visits of 105 subjects (38%) allowed determination of total symptom duration, defined as number of days from initial SARS-CoV-2 PCR positive to follow-up visits. Median symptom duration was 16 days (interquartile range, IQR = 11–31 days; range = 2–304 days), with approximately one-third of subjects reporting symptoms persisting ≥4 weeks. When grouped by symptom duration of <4 weeks (short-lived) vs. ≥4 weeks (prolonged), several distinguishing characteristics emerged, with the prolonged symptoms group having a higher prevalence of each symptom (Table 2), and a higher burden of multiple symptoms at the time of evaluation (8.2 vs. 3.3 symptoms, p < 0.0001, Figure 3).

There was no association between symptoms lasting ≥4 weeks and other factors such as underlying disease (e.g., chronic lung disease, systemic hypertension, diabetes), laboratory, or PFT abnormalities. Although we had only medication prescription data on 73 subjects, patients on oral corticosteroids (OCS; n = 19) were much less likely to have symptoms lasting >4 weeks compared with those not on OCS, (47% vs. 89%, p = 0.0003).

We included subjects with complete symptoms data (143) in elastic net analysis to predict prolonged symptoms (≥4 weeks), and all, except vomiting, were selected as predictors for prolonged symptoms with AUC of 0.99. By regression model analysis, all 13 variables increased the likelihood of an individual likely to experience prolonged symptoms. Anosmia has the largest odds ratio of 23.01, followed by myalgia (OR = 12.78), and dyspnea (OR = 10.78) (Figure 4).

### Seroconversion

Of 231 patients, 78% tested positive for antibodies following natural infection. We found higher seroconversion in males vs. females (88% vs. 75%, p = 0.03). Those who seroconverted were of older age (56.3 years-old...
vs. 44 years-old; \( p < 0.0001 \), more likely to report fever (82% vs. 69%, \( p = 0.04 \)), and less likely to report GI symptoms (e.g., abdominal pain; 63% vs. 84%; \( p = 0.0023 \)) than those without seroconversion. No difference in other symptoms or comorbidities was noted between these groups, although seroconversion rates were lower in patients on corticosteroids (66% vs. 86%, \( p = 0.02 \)).

Patients with complete data sets (98) were included in the elastic net analysis to determine predictors of seroconversion after COVID-19. Eight variables (except lymphocyte counts) were selected as predictors for seroconversion (AUC 0.73). Oral corticosteroid use (odds ratio, OR = 0.19) and elevated creatinine (OR = 0.24) decreased the likelihood of seroconversion after natural infection, whereas male sex (OR = 4.09) and leukocytosis (OR = 1.15) increased the likelihood of seroconversion (Figure 5).

**Discussion**

Despite the higher morbidity and mortality of acute COVID-19 in older adults, most symptomatic infections affect those in the 18–50 age group\(^9\text{–}^{11}\); their prolonged COVID-19 symptoms contribute to loss of societal productivity and to a high healthcare burden\(^12\text{,}^{13}\). Reports of cohorts surveyed after ICU/hospital stays or self-reported by subjects at large identified multiple\(^7\text{–}^{15}\) post-COVID symptoms that persisted in 20–80% of patients.\(^4\text{,}^{5\text{,}^{14\text{–}20}\text{ We report for the first time the varied symptomatology, symptom burden, and natural seroconversion rates of patients who present to an outpatient clinic specifically designed to manage ambulatory patients experiencing post-acute COVID-19 symptoms.}

By establishing a CPCCR and anticipating the need to manage patients with persistent multi-system symptoms and provide comprehensive long-term COVID-19 care, we captured the clinical presentation of individuals with persistent functional and/or radiographic abnormalities and those with recurrent or prolonged infections. The CPCCR was designed with guided safety protocols to address uncertainty of illness duration and infectivity.\(^7\text{,}^{21}\) Our study of a PASC cohort spanned 6 months during the first year of the COVID-19 pandemic, prior to vaccine availability, at a specialized quaternary medical center caring for individuals with chronic respiratory and immunologic disorders. As the long-term impact of COVID-19 affects healthcare resources and societal productivity across the globe, despite effective vaccination, our study
provides novel insight into ambulatory presentation of adults following acute infection. The clinical factors predictive of prolonged symptoms are useful to identify individuals at risk who may require enhanced prophylaxis, earlier evaluation, and targeted interventions with emerging therapeutics.

Interestingly, prolonged symptoms were not associated with lung disease or dysfunction, consistent with other reports, which is reassuring for those with underlying respiratory conditions. In addition, our finding that corticosteroid use correlated with a reduced symptom burden and duration may encourage their use during acute COVID-19.

The seroconversion rates, detected by IgG and/or IgM antibodies against SARS-CoV-2 spike protein following infection were high (78%) and favorably influenced by male sex, older age, leukocytosis, and self-reported fever. Because we did not account for the severity of COVID-19 infection, these findings may be related to a higher severity of infection described in this demographic. Our findings are consistent with reports that the strongest predictors of seroconversion include cough, anosmia/dysgeusia, and fever. Elevated creatinines decreased the likelihood of seroconversion, as did OCS use, a finding that was not previously identified in hospitalized severe COVID-19 cases. The overall clinical impact of the dichotomous effect of OCS in ambulatory patients, reducing duration of PASC symptoms while impairing seroconversion, remains to be tested in larger prospective studies.

In addition to caring for discharged patients surviving severe COVID-19 pneumonia and/or acute respiratory distress syndrome (ARDS) following prolonged hospital and ICU stay, ambulatory centers are evaluating individuals with a surprisingly heterogeneous clinical presentation even in those with milder acute COVID-19. Our clinical experience suggests that despite this protean presentation, PASC patients may be phenotypically classified in four distinct groups: (1) Prolonged multisystem complaints, in previously relatively healthy, younger patients, with mild initial acute illness ("long-COVID" or "long-haulers"); (2) Symptoms previously identified following severe illness (post-ICU/post-hospitalization for acute disease); (3) Sequelae of discrete organ dysfunction resulting from severe acute infection (e.g., post-ARDS pulmonary fibrosis); (4) Exacerbation of comorbid lung disease (e.g., asthma). All these proposed distinct phenotypes were represented in our study, and we found that Group 1 presents a unique challenge, due to its novelty and unclear pathogenesis. Since current case definitions and symptom surveys for PASC do not differentiate yet among these proposed groups, it may be useful to evaluate such sub-phenotyping in future studies of therapeutic interventions and outcomes in these patients. Indeed, while vaccination decreased COVID-19 mortality and hospitalization, there is increasing need for interventions to mitigate the long-term sequelae of COVID-19, given their negative impact on societal function and healthcare and economic burden.

**Study Limitations**

Limitations of this study include its retrospective design, imposed by data collection in a real-world clinical environment early during the COVID-19 pandemic. As the onset of symptoms was variable (i.e., occurring later during PASC), and since data were collected during clinical follow-up rather than at standardized intervals, the longitudinal analysis of symptom duration may have been biased toward shorter duration. Any missing data points were accounted for with statistical analysis. The generalizability of our results may also be limited by the composition of a cohort that included many patients with chronic respiratory disease, predominantly identifying as white, indicating that our novel findings should be confirmed in future studies that include diverse populations.

**Conclusions**

Our study of unvaccinated ambulatory patients with PASC identified clinical features that predict prolonged course and seroconversion after SARS-CoV-2 infection. Most patients presented with multiple symptoms. A greater number of initial symptoms, especially if those included anosmia, were indicative of prolonged symptoms. Persistence of pulmonary symptoms did not correlate with abnormal pulmonary
function testing or with pre-existing pulmonary disease. Natural post-infection seroconversion rates were high, especially in older, male individuals who experienced fever during the acute COVID-19. Oral corticosteroid was associated with reduced burden of prolonged symptoms, but with lower natural seroconversion.

Our results may identify ambulatory patients at greatest risk for prolonged morbidity, and therefore targeted for COVID-19 prevention and treatment and proposes several distinct post-COVID phenotypes that may aid in future PASC study design.

**ACKNOWLEDGMENTS**

Authors acknowledge the significant contributions of Joy Zimmer for data acquisition and organization from EMR.

**HUMAN SUBJECTS**

Approval from BRANY institutional review board was obtained prior to initiation of the study (IRB #20-12-582, Approval 12/15/2020).

**FUNDING SOURCES**

Wollowick Chair of COPD Research, Department of Medicine, National Jewish Health (IP)

**DISCLOSURES**

IP has received consulting honoraria from CeraMedix and Allinaire Inc. VPG has received consulting honoraria from AstraZeneca for Advisory Board and from Regeneron for Speaker Bureau. None of the above disclosures represent a conflict with this manuscript.

**AUTHOR CONTRIBUTIONS**

Acquisition of data: NG, RD, JJE, RK, PZ. Conception and design: NG, BM, IP. Analysis and interpretation: NG, IP, BM, SYL, VPG. Writing of manuscript: NG, VPG, CO, IP, BM, JJE. Editing and approval of manuscript: NG, JJE, VPG, RD, CO, RD, BM, IP.


**From:** 1Division of Pulmonary, Critical Care, and Sleep Medicine (VPG, CO, RK, RD, IP), Department of Medicine, National Jewish Health, Denver, CO; 2Division of Pulmonary Sciences and Critical Care Medicine (SYL, IP), Department of Medicine, University of Colorado, Anschutz Medical Campus, Denver, CO; 3Division of Occupational Medicine (SYL), Department of Medicine, NJH, Denver, CO; 4Colorado School of Public Health (SYL), University of Colorado, Anschutz Medical Campus, Denver, CO; 5The NJH Cohen Family Asthma Institute (VPG), NJH, Denver, CO; 6Scripps Institute(BM), La Jolla, CA; 7Division of Respiratory and Mycobacterial Infections (PZ, JJE), Department of Medicine NJH, Denver, CO

**Submitted:** 5/10/2022

**Accepted:** 7/1/2022

**Conflicts of interest:** none

This work is licensed under a Creative Commons Attribution-ShareAlike 4.0 International License.

**REFERENCES**