

Patient Self-Report Superior to Electronic Medical Record Abstraction for Identifying Positive COVID-19 Symptoms at Illness Onset



Gretchen L. Mockler, MD,¹ Samantha P. Novotny, BS,² Wei Hou, PhD,¹ Yuhang Liu, MS,³ Elinor R. Schoenfeld, PhD¹

Introduction: Most initial COVID-19 research focused on hospitalized patients. Presenting symptomatology in the outpatient setting was poorly characterized, making it difficult for primary care physicians to predict which patients would require hospitalization. The purpose of this study was to characterize the presenting symptoms of COVID-19 infection and baseline patient characteristics and evaluate for correlation with disease severity, duration, and chronicity in the outpatient setting.

Methods: A total of 107 adult, English-speaking patients with suspected and confirmed COVID-19 cases at the 3 primary care practices of Stony Brook University Hospital were studied between March and December 2020. Survey data were collected from patient telephone interviews and electronic medical record abstraction. The potential risk factors assessed included participant demographics, medical comorbidities, and the number and type of symptoms at illness onset. Outcome measures included symptom duration, hospitalizations, and persistence of symptoms at 12 weeks from study enrollment.

Results: Patient self-report survey elicited nearly twice as many symptoms described at illness onset as those recorded in the electronic medical record ($p < 0.0001$). A higher number of symptoms at illness onset was positively associated with symptom duration and chronicity. The presence of fever and hypoxia at the onset of illness were each positively associated with eventual hospitalization for COVID-19 disease.

Conclusions: Early in the setting of newly emerging infectious diseases, particularly those such as COVID-19 that involve multiple organ systems, patient self-report of symptoms using a complete review of systems rather than electronic medical record abstraction alone may be key for accurate disease identification and characterization as well as prediction of eventual disease severity, duration, and chronicity.

From the ¹Department of Family, Population & Preventive Medicine, Renaissance School of Medicine, Stony Brook University Hospital, Stony Brook, New York; ²Renaissance School of Medicine, Stony Brook University Hospital, Stony Brook, New York; and ³Department of Applied Mathematics & Statistics, College of Engineering and Applied Sciences, Stony Brook University, Stony Brook, New York

Address correspondence to: Gretchen L. Mockler, MD, Department of

Family, Population & Preventive Medicine, Renaissance School of Medicine, Stony Brook University Hospital, HSC L3, Room 086, Stony Brook NY 11794-8036. E-mail: Gretchen.mockler@stonybrookmedicine.edu. 2773-0654/\$36.00

<https://doi.org/10.1016/j.focus.2022.100005>

INTRODUCTION

As coronavirus disease 2019 (COVID-19) ravaged nations around the world in early 2020, the U.S. reported the highest number of infections worldwide.¹ New York State, New York City, and the New York City suburbs rapidly became COVID-19 epicenters, with an extraordinary burden of disease and mortality. As of February 5, 2022, there have been over 76 million COVID-19 cases and 900,000 deaths in the U.S. Suffolk County, New York, where our study was based, has reported over 417,000 cases and 4,200 deaths.¹

Most initial COVID-19 research focused on hospitalized patients.^{2,3} Presenting symptomatology in the outpatient setting and nonrespiratory symptoms were poorly characterized. This made it difficult for primary care physicians to predict which patients would require hospitalization or decompensate while being managed at home, and hospital beds were limited. To address these knowledge gaps, we initiated a study using patient interviews and electronic medical record (EMR) review to characterize the presenting symptoms of COVID-19 infection in the outpatient setting and evaluate for correlations with disease severity, duration, and chronicity.

METHODS

Study Population

IRB approval for this longitudinal, observational study was obtained. Patients aged ≥ 18 years who presented to any of the 3 suburban primary care practices of Stony Brook University Hospital with physician-suspected or test-confirmed COVID-19 between February and August 2020 and who agreed to telephone interview and EMR review by study staff were eligible. Of the primary care practices involved in this study, all 3 were faculty practices, 1 was a patient-centered medical home, and 1 included residents being precepted by faculty. The size of the 3 practices included in this study totaled 24,463 patients: 15,942 in internal-medicine primary care, 7,821 in family medicine, and 700 in preventive medicine. Participants were followed from April through December 2020. Study staff were trained in interviewing and EMR data abstraction. Data were collected through telephone and EMR review and recorded into a REDCap database. Patient interview questionnaires were designed in regular consultation with emerging Centers for Disease Control and Prevention data regarding COVID-19 and queried demographics, medical comorbidities, medications, prevention of transmission practices, travel, COVID-19 symptom course and treatment, and test results. Parallel EMR data were abstracted from clinician notes and from laboratory and radiology reports. Participants were surveyed weekly

until 2 weeks after COVID-19 symptoms resolved or a maximum of 3 months.

Measures

Questionnaire responses were used to characterize potential risk factors including participant demographics, medical comorbidities, travel, and number and type of symptoms at illness onset including *a priori* 5 key initial COVID-19 signs and symptoms commonly used in the existing literature as surrogates for disease severity: fever, cough, shortness of breath, hypoxia and confusion.²⁻⁴ Fever, including low-grade fever, was defined as temperature $\geq 99^{\circ}\text{F}$.^{5,6} Outcomes included increased disease severity, for which hospitalization was used as a marker; increased symptom duration; and symptom chronicity, defined as persistent symptoms at 12 weeks after study enrollment.

Statistical Analysis

Patient characteristics were described using medians (range) and frequencies (percentages) when appropriate. Percentages of pre-existing medical comorbidities, number of symptoms at the onset of illness, COVID-19 testing, and escalating medical care were calculated. Percentages of each individual symptom were calculated at the onset of illness and at a 12-week follow-up. To evaluate the association of risk factors with hospitalization (ever versus never) and symptom chronicity (persistent versus resolved), univariable and multivariable analyses were performed. To evaluate for an association of risk factors with duration of symptoms, Spearman correlation was used. For univariable analyses, chi-square tests were used for categorical risk factors (e.g., medical comorbidities, travel history). Wilcoxon ranks sum tests were used for numeric risk factors (e.g., number of symptoms at illness onset). For multivariable analyses, hospitalization (ever versus never) and symptom-resolved status (persistent versus resolved) were analyzed as the outcomes in logistic regression. Explanatory variables included the number of symptoms, presence of 5 key signs, symptoms at the onset of illness, medical comorbidities, history of travel, and demographic characteristics (i.e., age, sex, and race). All analyses were hypothesis driven and were performed using SAS, version 9.4 (SAS Institute, Cary, NC).

RESULTS

A total of 178 patients were deemed eligible for the study; 146 were successfully contacted, and 107 were enrolled. A total of 14 participants were lost to interview follow-up, but EMR follow-up was completed. Among participants, the largest group by age was those aged 40–64 years, and the majority were female and Caucasian. Half had traveled either domestically or internationally during the pandemic period (after January 1, 2020). Nearly one fifth of participants had no pre-existing medical conditions. Close to 50% of participants had a

Table 1. Participant Demographics, Characteristics, and COVID-19 Test Status (N=107)

Participant	n	%
Age (years)		
<40	13	12.1
40–64	51	47.7
65–74	26	24.3
≥75	17	15.9
Median, range	61, 21–91	
Sex (female)	70	65.4
Race (Caucasian)	87	82.1
Ethnicity (Hispanic/Latino)	16	15.0
Travel		
International	9	8.6
Domestic	44	41.9
Works in a healthcare facility	23	22.1
Number of pre-existing comorbidities	18	16.8
0		
1–2	61	57.0
3–4	24	22.4
5–6	4	3.7
Median, range	1, 0–6	
High-risk medical comorbidities		
Obesity	50	47.2
Diabetes	11	10.3
Hypertension	42	39.2
Other cardiovascular disease (e.g., CHF, CAD)	52	48.6
Chronic pulmonary disease (any)	32	29.9
Number of symptoms at the onset of illness		
Self-reported (median, range)	7, 0–18	
EMR-documented (median, range)	4, 0–13	
O ₂ sat documented	61	57.0
Hypoxia documented (O ₂ sat ≤93%)	15	14.0
Emergency room visit	38	29.9
Hospitalization	28	23.9
COVID-19 test results		
	(n/N tests)	%
PCR positive	37 (37/75)	49.3
Rapid antigen test positive	11 (11/19)	57.9
Antibody test positive	27 (27/45)	60.0
Any positive test result	58 (58/93)	62.4

CAD, coronary artery disease; CHF, congestive heart failure; EMR, electronic medical record; O₂sat, oxygen saturation; PCR, polymerase chain reaction.

diagnosis of obesity and/or cardiovascular disease, approximately one third had chronic pulmonary disease, and one tenth had diabetes (Table 1).

Participant self-report at telephone interview yielded nearly 2 times as many COVID-19-related symptoms at

illness onset as EMR abstraction (median=7 vs 4, $p<0.0001$). The most frequent symptoms at the onset of illness were fatigue, cough, myalgias, and headache. A total of 18% (19 of 107) of study participants had persistent symptoms at 12 weeks after study enrollment. In order of prevalence, these symptoms were fatigue, headache, myalgia, and shortness of breath (Table 2). Presence of fever or hypoxia at illness onset was positively associated with hospitalization in the multivariable logistic regression model (Table 3). A higher number of symptoms at illness onset was positively associated with increased symptom duration ($R=0.2556$, $p=0.0279$) and with chronic symptoms at 3 months from study enrollment ($p=0.04$) (Table 4).

DISCUSSION

This longitudinal study collected patient-reported and EMR data from 3 primary care practices to describe COVID-19 symptomatology, disease progression, and hospitalization in an epicenter of the pandemic's first wave. Patients self-reported nearly twice as many symptoms at illness onset as were recorded in the EMR. This gap in the documentation of patients' symptoms has been previously noted in other disease states^{7–9} but has a particular urgency for our current circumstances.

Rapid characterization of the full range of symptoms and presentations of newly emerging diseases such as COVID-19 is essential for accurate diagnosis and management. In the case of COVID-19, a higher number of distinct symptoms in the first 5 days of illness has been associated with the eventual need for supplemental oxygen.⁴ At the beginning of the COVID-19 pandemic, defining symptoms included cough, dyspnea, fever, and fatigue.^{2,3} Only later were gastroenterological, neurologic, and dermatological manifestations recognized.^{10–12}

Time constraints owing to increased patient volume during periods of emergency can prohibit extensive review of systems (ROS) documentation. Therefore, for research purposes when investigating new diseases, EMR data abstraction alone may be insufficient. Supplemental patient interviews with full ROS may be necessary to obtain comprehensive data and define pathognomonic symptoms of new diseases in real time. This method might also help to characterize newer viral variants such as the Delta or Omicron variants of COVID-19 as they emerge as well as atypical presentations particular to specific patient groups (e.g., elderly, pediatric, pregnant, immunosuppressed). To make this feasible, a full ROS questionnaire might be sent electronically to patients before their clinic visits or obtained by clinic support staff by telephone, both for physician review at the time of the patient's clinical visit and for

Table 2. Symptoms at Onset of Illness and 12 Weeks From Study Enrollment

Symptoms	Present at the onset of illness (n=107 participants)		Present at 12-week interview (n=19 participants with persistent symptoms)	
	n	%	n	%
Fatigue	83	77.6	15	79.0
Cough	76	71.0	5	26.3
Muscle aches	66	61.7	7	36.8
Headache	66	61.7	7	36.8
Subjective fever (feels feverish)	60	56.1	1	5.3
Fever (temperature $\geq 99^\circ\text{F}$)	60	56.1	0	
Chills	56	52.3	2	10.5
Shortness of breath	54	50.5	10	52.6
Other symptom	48	44.9	8	42.1
Chest discomfort, tightness or pain	48	44.9	4	21.1
Sore throat	47	43.9	5	26.3
Sinus congestion	41	38.3	6	31.6
Diarrhea	37	34.6	3	15.8
Nausea	33	30.8	2	10.5
Runny nose	29	27.1	4	21.1
Loss of taste	28	26.2	5	26.3
Loss of smell	27	25.2	5	26.3
Abdominal pain	26	24.3	3	15.8
Increased sputum production	25	23.4	2	10.5
Eye redness or pain/conjunctivitis	24	22.4	3	15.8
Confusion	22	20.6	5	26.3
Vomiting	16	15.0	0	

Table 3. Potential Risk Factors Versus Hospitalization for COVID-19 Disease (n=107)

Risk factor	Ever hospitalized (n=28)		Never hospitalized (n=79)		p-value ^a
	Median	IQR	Median	IQR	
Continuous					
Number of symptoms at the onset of illness	11.0	6.3	9.0	6.5	0.06
Age	64.5	17.3	60.0	18.0	0.06
Categorical	n	%	n	%	
Sex (female)	14	50.0	56	70.9	0.08
Race	3	10.7	16	20.5	0.38
Any of the 5 key signs/symptoms at illness onset	26	92.9	56	70.9	0.04
Fever at illness onset	21	75.0	38	48.1	0.03
Shortness of breath at illness onset	16	57.1	33	41.8	0.24
Hypoxia at illness onset ($\text{O}_2\text{sat} \leq 93\%$)	6	21.4	1	1.3	<0.01
Chest discomfort at illness onset	13	46.4	35	44.3	1.00
Confusion at illness onset	8	28.6	14	17.7	0.34
Obesity	13	46.4	37	47.4	1.00
Cardiovascular disease	18	64.3	34	43.0	0.09
Diabetes	2	7.1	9	11.4	0.78
Hypertension	16	57.1	26	32.9	0.03^b
Immunosuppression	3	10.7	3	3.8	0.37
Chronic pulmonary disease	8	28.6	24	30.4	1.00
History of domestic or international travel	10	35.7	39	50.7	0.26

Note: Boldface indicates statistical significance ($p < 0.05$).

IQR: Interquartile range

^aCategorical variable: chi-square test; Continuous variable: Wilcoxon rank sums test.

^bHypertension did not remain significant in the multivariable logistic regression model.

O_2sat , oxygen saturation.

Table 4. Potential Risk Factors Versus Chronic Symptoms at 12-Week Study Follow-Up Interview (n=96)

Risk factor	Persistent symptoms (n=19)		All symptoms resolved (n=77 ^a)		p-value ^b
	Median	IQR	Median	IQR	
Continuous					
Number of symptoms at onset of illness	13.0	9.5	9.0	5.0	0.04
Age	62.5	15.0	63.0	6.0	0.57
Categorical	n	%	n	%	
Sex (female)	14	73.7	46	59.7	0.39
Race	2	11.1	12	15.6	0.91
Any of the 5 key symptoms at onset of illness	16	84.2	56	72.7	0.46
Obesity	11	61.1	35	45.5	0.35
Cardiovascular disease	10	52.6	40	52.0	1.00
Diabetes	2	10.5	7	9.1	1.00
Hypertension	6	31.6	34	44.2	0.32
Immunosuppression	2	10.5	3	3.9	0.56
Chronic pulmonary disease	8	42.1	20	26.0	0.27
History of domestic or international travel	8	42.1	36	46.8	0.91

Note: Boldface indicates statistical significance ($p < 0.05$).

A total of 11 symptomatic participants were lost to follow-up before 12-week interview.

IQR: Interquartile range

^aOne participant never developed symptoms.

^bCategorical variables were analyzed with chi-square test, continuous variables were analyzed with Wilcoxon rank sums test.

research purposes. In the case of larger studies, automated natural language processing might be employed to maximize extraction of data from free text notes. Notably, real-time supplemental patient interviews are an important tool used in public health case investigation; thus, close collaboration between clinical and public health sectors may play a key role in the rapid and thorough characterization of new diseases.

The longitudinal aspect of this study allowed for analysis of associations between presenting symptoms and disease course, showing that the presence of hypoxia or fever at the time of illness onset were each associated with a higher risk of hospitalization and that a higher number of symptoms at illness onset was associated with increased symptom duration and with symptom chronicity 3 months from study enrollment. The symptoms that our study found to have persisted most commonly at the 3-month time point—fatigue, headache, myalgia, and shortness of breath—are consistent with the current Centers for Disease Control and Prevention description of symptoms of postacute sequelae of severe acute respiratory syndrome coronavirus 2 (SARS CoV-2).¹³

Longitudinal analyses are essential for characterizing atypical presentations, varying disease trajectories, and chronic symptoms of conditions such as COVID-19. Identifying the risk factors for adverse outcomes and chronicity may allow clinicians to more accurately risk stratify patients early in the disease course and enhance

management and follow-up of these high-risk patients. Future research is warranted to address the chronic symptoms and specific criteria for the diagnosis of postacute sequelae of SARS CoV-2.

Limitations

Our patient sample was limited to primary care patients at a single hospital system in Suffolk County, NY during the initial wave of COVID-19. It may not be generalizable to other populations. Small sample size limited our statistical power to analyze the correlations between potential risk factors and disease outcomes. Because this study occurred during a time when outpatient and inpatient resources were severely limited, many patients were managed from home through telehealth. This limited the ability of clinicians to obtain vital signs or a complete physical examination. Finally, COVID-19 testing was not widely available for the initial period of our study, and test types varied widely later during the study period, so we relied on clinician diagnosis alone when necessary. This method may be needed again in the case of other new infectious diseases for which testing is being developed and refined while the disease itself is still being clinically defined.

CONCLUSIONS

In our present times of rapid population growth, global warming, high potential for further pandemic diseases, and

increased virtual assessment of patients, it may be wise to return to a comprehensive ROS to evaluate for multiple organ system involvement in a newly emerging disease such as COVID-19. The outpatient primary care setting may offer the ideal opportunity for this approach, helping to rapidly characterize novel diseases and save patients' lives earlier in the trajectory of a future pandemic.

ACKNOWLEDGMENTS

The authors acknowledge Raja Jaber; Lisa Strano Paul, MD; Hrisanti Arcan, PhD; Christopher Aseervatham; Rachel Bright; Yuanzhen Cao; Yuwen Cheng; Christa Lynn Denmon; Sanket Desai; Samantha Fuller; Jenny Fung; Olivia Hu; Jessie Huang; Yujie Jiang; Emily Leong; Soumya Mandava; Christopher Medrano; Kelly Pellegrino; Chetan Potu; Olga Reykhardt; Priyal Sakhuja; Spencer Schell; Elisabeth Sorenson; Jiayu Tian; Danny Wu; and Jessica Ye and all the study participants, who volunteered their time and effort to contribute to the health and well-being of patients to come.

The sponsor had no role in the study design; collection, analysis, or interpretation of data; the writing of the report; or the decision to submit the report for publication. This study received approval with IRB Number 2020-00234.

This work was supported by the State University of New York Center-Scale Seed Grant, \$7,500, July 2020.

This study was presented in part at the North American Primary Care Research Group Annual Conference, November 2021.

No financial disclosures were reported by the authors of this paper.

CREDIT AUTHOR STATEMENT

Gretchen Mockler, MD: Conceptualization, Methodology, Funding acquisition, Resources, Investigation, Supervision, Project administration, Validation, Writing - original draft, review and editing, Visualization. Samantha Novotny, BS: Conceptualization, Investigation, Writing - original draft, review and editing. Wei Hou, PhD: Methodology, Software, Formal analysis, Resources, Supervision, Project administration. Yuhang Liu, MS: Software, Formal analysis, Validation. Elinor Schoenfeld, PhD: Conceptualization, Methodology, Resources, Software, Investigation, Data curation, Supervision, Project administration, Validation, Writing - review and editing.

REFERENCES

- Dong E, Du H, Gardner L. An interactive web-based dashboard to track COVID-19 in real time [published correction appears in *Lancet Infect Dis*. 2020;20(9):e215]. *Lancet Infect Dis*. 2020;20(5):533–534. [https://doi.org/10.1016/S1473-3099\(20\)30120-1](https://doi.org/10.1016/S1473-3099(20)30120-1).
- Huang C, Wang Y, Li X, et al. Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China. *Lancet*. 2020;395(10223):497–506. [https://doi.org/10.1016/S0140-6736\(20\)30183-5](https://doi.org/10.1016/S0140-6736(20)30183-5).
- Wang D, Hu B, Hu C, et al. Clinical characteristics of 138 hospitalized patients with 2019 novel coronavirus-infected pneumonia in Wuhan, China [published correction appears in *JAMA*. 2021;325(11):1113]. *JAMA*. 2020;323(11):1061–1069. <https://doi.org/10.1001/jama.2020.1585>.
- 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):eabd4177. <https://doi.org/10.1126/sciadv.abd4177>.
- Affronti M, Mansueto P, Soresi M, et al. Low-grade fever: how to distinguish organic from non-organic forms. *Int J Clin Pract*. 2010;64(3):316–321. <https://doi.org/10.1111/j.1742-1241.2009.02256.x>.
- Islam MA, Kundu S, Alam SS, Hossain T, Kamal MA, Hassan R. Prevalence and characteristics of fever in adult and paediatric patients with coronavirus disease 2019 (COVID-19): a systematic review and meta-analysis of 17515 patients. *PLoS One*. 2021;16(4):e0249788. <https://doi.org/10.1371/journal.pone.0249788>.
- Mizrahi B, Shilo S, Rossman H, et al. Longitudinal symptom dynamics of COVID-19 infection. *Nat Commun*. 2020;11(1):6208. <https://doi.org/10.1038/s41467-020-20053-y>.
- Valikodath NG, Newman-Casey PA, Lee PP, Musch DC, Niziol LM, Woodward MA. Agreement of ocular symptom reporting between patient-reported outcomes and medical records. *JAMA Ophthalmol*. 2017;135(3):225–231. <https://doi.org/10.1001/jamaophthalmol.2016.5551>.
- Fares CM, Williamson TJ, Theisen MK, et al. Low concordance of patient-reported outcomes with clinical and clinical trial documentation. *JCO Clin Cancer Inform*. 2018;2:1–12. <https://doi.org/10.1200/CCL18.00059>.
- Han C, Duan C, Zhang S, et al. Digestive symptoms in COVID-19 patients with mild disease severity: clinical presentation, stool viral RNA testing, and outcomes. *Am J Gastroenterol*. 2020;115(6):916–923. <https://doi.org/10.14309/ajg.000000000000664>.
- Tyson B, Erdodi L, Ray S, Agarwal P. Altered mental status in 71 deaths due to COVID-19. *Int J Neurosci*. 2020:1–4. <https://doi.org/10.1080/00207454.2020.1825422>.
- Recalcati S. Cutaneous manifestations in COVID-19: a first perspective. *J Eur Acad Dermatol Venerol*. 2020;34(5):e212–e213. <https://doi.org/10.1111/jdv.16387>.
- Nalbandian A, Sehgal K, Gupta A, et al. Post-acute COVID-19 syndrome. *Nat Med*. 2021;27(4):601–615. <https://doi.org/10.1038/s41591-021-01283-z>.