

Original Paper

Assessing Associations Between COVID-19 Symptomology and Adverse Outcomes After Piloting Crowdsourced Data Collection: Cross-sectional Survey Study

Natalie Flaks-Manov¹, PhD; Jiawei Bai², PhD; Cindy Zhang³; Anand Malpani³, PhD; Stuart C Ray¹, MD; Casey Overby Taylor¹, PhD

¹Johns Hopkins University School of Medicine, Baltimore, MD, United States

²Johns Hopkins Bloomberg School of Public Health, Baltimore, MD, United States

³Johns Hopkins Whiting School of Engineering, Baltimore, MD, United States

Corresponding Author:

Casey Overby Taylor, PhD

Johns Hopkins University School of Medicine

3101 Wyman Park Dr.

Baltimore, MD, 21218

United States

Phone: 1 443 287 6657

Email: cot@jhu.edu

Abstract

Background: Crowdsourcing is a useful way to rapidly collect information on COVID-19 symptoms. However, there are potential biases and data quality issues given the population that chooses to participate in crowdsourcing activities and the common strategies used to screen participants based on their previous experience.

Objective: The study aimed to (1) build a pipeline to enable data quality and population representation checks in a pilot setting prior to deploying a final survey to a crowdsourcing platform, (2) assess COVID-19 symptomology among survey respondents who report a previous positive COVID-19 result, and (3) assess associations of symptomology groups and underlying chronic conditions with adverse outcomes due to COVID-19.

Methods: We developed a web-based survey and hosted it on the Amazon Mechanical Turk (MTurk) crowdsourcing platform. We conducted a pilot study from August 5, 2020, to August 14, 2020, to refine the filtering criteria according to our needs before finalizing the pipeline. The final survey was posted from late August to December 31, 2020. Hierarchical cluster analyses were performed to identify COVID-19 symptomology groups, and logistic regression analyses were performed for hospitalization and mechanical ventilation outcomes. Finally, we performed a validation of study outcomes by comparing our findings to those reported in previous systematic reviews.

Results: The crowdsourcing pipeline facilitated piloting our survey study and revising the filtering criteria to target specific MTurk experience levels and to include a second attention check. We collected data from 1254 COVID-19-positive survey participants and identified the following 6 symptomology groups: abdominal and bladder pain (Group 1); flu-like symptoms (loss of smell/taste/appetite; Group 2); hoarseness and sputum production (Group 3); joint aches and stomach cramps (Group 4); eye or skin dryness and vomiting (Group 5); and no symptoms (Group 6). The risk factors for adverse COVID-19 outcomes differed for different symptomology groups. The only risk factor that remained significant across 4 symptomology groups was influenza vaccine in the previous year (Group 1: odds ratio [OR] 6.22, 95% CI 2.32-17.92; Group 2: OR 2.35, 95% CI 1.74-3.18; Group 3: OR 3.7, 95% CI 1.32-10.98; Group 4: OR 4.44, 95% CI 1.53-14.49). Our findings regarding the symptoms of abdominal pain, cough, fever, fatigue, shortness of breath, and vomiting as risk factors for COVID-19 adverse outcomes were concordant with the findings of other researchers. Some high-risk symptoms found in our study, including bladder pain, dry eyes or skin, and loss of appetite, were reported less frequently by other researchers and were not considered previously in relation to COVID-19 adverse outcomes.

Conclusions: We demonstrated that a crowdsourced approach was effective for collecting data to assess symptomology associated with COVID-19. Such a strategy may facilitate efficient assessments in a dynamic intersection between emerging infectious diseases, and societal and environmental changes.

KEYWORDS

COVID-19; coronavirus; symptoms; symptomology; crowdsourcing; adverse outcomes; data quality

Introduction

COVID-19 represents a global public health concern [1-3]. While extensive measures are being implemented to control the outbreak, the high speed of transmission makes collecting data needed to inform clinical management and public health planning a challenge. Efficiently collecting high-quality data to characterize disease severity enables accurate information to be disseminated in a timely manner for such planning.

To understand and predict the adverse health outcomes in patients affected by COVID-19, many scientific efforts studying sociodemographic, clinical, and symptomatic risk factors are underway. Findings from those efforts, however, are not all consistent, with conflicting evidence on the risk factors associated with adverse COVID-19 outcomes [4-6]. Furthermore, infected people have reported a wide range of symptoms, from asymptomatic to severe illness [2-12]. Common symptoms include fever, cough, fatigue, shortness of breath, and loss of the sense of smell or taste, and less frequent symptoms are gastrointestinal and neurological symptoms [4-6,13-16]. Although there has been a concerted effort to describe patients' symptoms [7,17,18], there is no evidence yet as to whether symptoms differ between people with different characteristics, such as chronic diseases and demographic backgrounds [19,20]. As individual symptoms cannot predict COVID-19 adverse outcomes [21], knowledge of a patient's profile of symptoms (ie, symptomology) holds promise to improve estimations of the risk of adverse outcomes [22].

A crowdsourcing model is a useful way to rapidly collect information in the context of the COVID-19 pandemic [23,24]. Recent work to classify different types of crowdsourcing used to tackle the COVID-19 crisis [23] found that the most common configuration to deal with information and knowledge management problems was open crowdsourcing (described as a one-to-many configuration with potentially unlimited contributors, and without any form of preselection). Most initiatives falling under this category, however, demonstrated a desire to locate and assemble information. The *COVID Near You* website [25], for example, uses crowdsourced data to visualize maps to identify current and potential pandemic hotspots. An important emphasis for crowdsourced data, however, is to collect high-quality data. Indeed, the risk of bias can be great when building COVID-19 diagnosis and prognosis prediction models trained on small or low-quality data sets. The majority of COVID-19 prediction models to date, for example, show a high risk of bias (n=226, 97%) [26].

To eliminate substandard crowd data submissions, we used a "crowdsourcing via a broker" strategy with broker services that allowed for filtering participants and their responses, and testing data quality before finalizing the crowdsourcing data collection strategy. We chose to use the Amazon Mechanical Turk (MTurk) crowdsourcing platform that provides filtering

mechanisms via setting qualifications [27-30]. Through the MTurk platform, entities known as "requesters" can hire independent contractors, known as "workers," to perform a wide variety of remote jobs, known as "human intelligence tasks" (HITs). A worker's reputation is indicated by their HIT acceptance rate [30]. The emphasis on obtaining good-quality data through setting qualifications, however, has created some concern about "superworkers." These are experienced and very active MTurk workers to whom researchers often target survey distribution. This oversampling from experienced workers can lead to an issue of worker nonnaivete as workers are frequently exposed to common methods in research studies. Recent research shows that nonsuperworkers can also produce high-quality data [31], and our strategy thus incorporated a pilot phase with broad inclusion criteria according to experience qualifications. Rather than defaulting to experienced workers, the pilot data collection allowed us to determine what filtering criteria were best suited to our needs.

In this paper, we describe (1) a pipeline to enable data quality and population representation checks in a pilot setting prior to deploying the final survey to MTurk workers, (2) an assessment of COVID-19 symptomology among MTurk worker survey respondents who reported a previous positive COVID-19 result, and (3) an assessment of the associations of symptomology groups and underlying chronic conditions with adverse outcomes due to COVID-19.

Methods

Study Design and Instrument

This was a cross-sectional study. We developed 2 web-based surveys using Qualtrics. One survey was for individuals (ie, individual survey) who indicated a self-reported positive test for COVID-19, and another survey was for individuals whose relatives (ie, family survey), living in the same house, tested positive for COVID-19. We hosted both surveys on MTurk between August and December 2020. To improve the quality of data collection through MTurk and to make our study sample more representative of the target population, we followed the best practices suggested by Young et al [30].

A few restrictions were implemented to exclude certain survey responses from the final data analysis. First, only those participants who provided an existing COVID-19 test type (nasal/throat/blood/sputum) answer in response to our screening question could continue with the survey. Second, participants could fill the survey only once for themselves (individual survey) and for 1 family member (family survey). Third, a quality control question was included during the questionnaire, which stated, "Do not answer this question (Please click NEXT to go to the next question)." If the question was answered, the survey responses were excluded from the data analyses. Fourth, in the family survey, we asked the participants about their confidence level in their responses regarding their family

member. Responses with low reported confidence were excluded from the final analyses.

Ethical Considerations

This study was judged as imposing only minimal risks on participants and was determined to be exempt research by the Johns Hopkins University (IRB00248053) Institutional Review Board.

Variables and Definitions

The web-based surveys had 5 blocks as follows ([Multimedia Appendix 1](#) and [Multimedia Appendix 2](#)):

1. *Introduction and screening*: Introduction to the aim of the study, including the estimated amount of time needed to complete the survey, the compensation amount, information about the voluntary nature of the survey, and instructions on not filling out the questionnaire more than once.
2. *Symptoms of COVID-19*: We asked participants to select all the symptoms that they experienced following a COVID-19 infection.
3. *Adverse outcomes of COVID-19*: We asked questions about hospitalizations related to COVID-19 and connection to a mechanical ventilator.
4. *Medical history*: We asked questions about background medical conditions, smoking status, and influenza vaccine status in a previous season.
5. *Demographic characteristics*: Participants reported on their age, sex they were assigned at birth, race, ethnicity, last year's income (monthly and yearly), and the highest level of education.

Survey measures were from the Johns Hopkins University COVID-19 community response survey guidance toolkit that draws from multiple sources [32]. An additional data source we used beyond the toolkit to compile COVID-19 symptoms was Twitter [18,32,33].

Recruitment

Population

The inclusion criteria for this study were individuals living in the United States, adults (aged 18 years or older), and MTurk workers with a self-reported positive COVID-19 result. For the family survey, the participants could complete the survey for 1 family member, even if the family member, who lived in the same household, was below 18 years old. Thus, the target population of this study was COVID-19 patients living in the United States and having sufficient skills to use the MTurk platform. The participants were compensated according to a standard minimum wage and our estimate of completion time (about 5-10 min).

Crowdsourcing Pipeline

Before posting the final survey to MTurk, we conducted a pilot from August 5, 2020, to August 14, 2020, to assess the quality of responses among workers with different levels of experience. The pilot analysis stratified the worker sample into the following 3 experience groups: those who previously completed 100-499 HITs, 500-999 HITs, and 1000+ HITs. First, a worker would complete a qualification test asking them to verify that they or

a family member tested positive for COVID-19. If qualified, the worker could then start the MTurk HIT that included a link to the 26-question web-based Qualtrics survey. The first part of the survey was a screening question (age ≥ 18 years) and comprehension check. In response to the comprehension check, if a worker selected an invalid COVID-19 test type (eg, urine test), they could not continue the survey.

For those passing the screening test, responses were labeled as "high quality" according to the following criteria: sufficient time taken (threshold of more than 60 s); matching codes and IDs between Qualtrics and MTurk; each code being associated with only 1 worker; and worker had not taken the survey previously (ie, nonduplicate response). A worker's response was included in the "high-quality" group if they passed all of these criteria.

Separately, we assessed "nonduplicate responses." A nonduplicate response indicates that the respondent completed the survey only once. This criterion was considered under the assumption that workers who attempted to complete the survey multiple times to receive more compensation did not read through survey instructions carefully, and thus, they may provide lower quality responses than those who attempted to complete the survey once.

The general characteristics of age, sex, race, education, and income were extracted and compared among experience groups. Chi-square analysis was conducted to evaluate if there was a significant difference between experience groups in the number of high-quality and nonduplicate responses. Findings from this analysis were used to refine our filtering criteria in the final crowdsourcing pipeline.

Statistical Analysis

Outcomes

We assessed the following 2 primary adverse outcomes related to COVID-19: hospital admission due to COVID-19 and use of mechanical ventilation during admission.

Statistical Analyses

We used descriptive statistics to characterize the total cohort of participants. Bivariate analyses, using Pearson χ^2 tests, were performed to assess differences in participant characteristics between those hospitalized and those not hospitalized, and between those who needed mechanical ventilation during admission and those who did not need mechanical ventilation. We then fitted multivariate logistic regression models to identify the association of COVID-19 symptoms with hospitalization and mechanical ventilation due to COVID-19, adjusted for sociodemographic characteristics and comorbid conditions. Thereafter, hierarchical cluster analysis was conducted to search for patterns based on COVID-19 symptoms. The similarity measure was cosine similarity, and the linkage method was Ward minimum variance. To describe clusters, we calculated frequencies of the risk factors for each cluster of symptoms. We then developed logistic regression models for hospitalization and mechanical ventilation as outcomes, using symptomology groups as risk factors. Finally, we developed a logistic regression model for each symptomology group to identify the significant

risk factors for hospitalization among individuals with different symptomology. All analyses were performed using R version 3.6.2 (R Foundation for Statistical Computing).

Validation Assessment

To validate our findings, we performed a comparison with existing systematic review or meta-analysis papers that assessed symptoms as risk factors for COVID-19 adverse outcomes. Articles for which the analyses occurred prior to our data collection were selected for comparison.

For each article and this study, individual symptoms were checked for being reported as (1) a significant risk factor for an adverse outcome (“yes”) and (2) a nonsignificant risk factor for an adverse outcome (“no”). We also noted if a symptom was not assessed (“NA”). When synthesizing findings across studies, if we found a statistically significant association between an adverse outcome and a symptom that was not studied by others, we labeled it “New.” If there was agreement between this study and at least one other study in identifying a symptom as a risk factor (significant or nonsignificant), we labeled it “1.” Symptoms we did not assess were labeled “NA.”

Results

Pilot Findings

Pilot survey data were collected from 259 respondents who passed both the qualification test and the screening questions, and of these, 147 (56.8%) were considered to have “high quality” responses. For the experience groups 100-499, 500-999, and 1000+ HITs, the proportions of high-quality responses were 58% (48/83), 43% (41/95), and 72% (58/81), respectively (Table 1). There was no significant difference between the experience groups for obtaining high-quality responses ($P=.14$). There was, however, a significant difference between the groups for nonduplicate responses ($P<.001$). Comparisons of demographic characteristics across all experience groups among MTurk workers are shown in Multimedia Appendix 3.

Two modifications were made to our crowdsourcing pipeline following the pilot. First, we included only workers with 500+ prior HITs in our final filtering criteria. Given the differences in nonduplicate responses between groups, we reasoned that for tasks requiring a higher cognitive ability, workers with 500+ HITs may provide more high-quality responses than those with 100-499 HITs. Second, we added an attention check question to the Qualtrics survey (ie, “don’t answer this question”).

Table 1. Comparison of the approval rates and percentage of quality responses out of approved responses between the different experience groups.

Variable	Experience group			P value
	100-499 HITs ^a , n/N (%)	500-999 HITs, n/N (%)	1000+ HITs, n/N (%)	
High-quality responses	48/83 (58)	41/95 (43)	58/81 (72)	0.135
High-quality and nonduplicate responses	10/48 (21)	41/41 (100)	49/58 (85)	<0.001

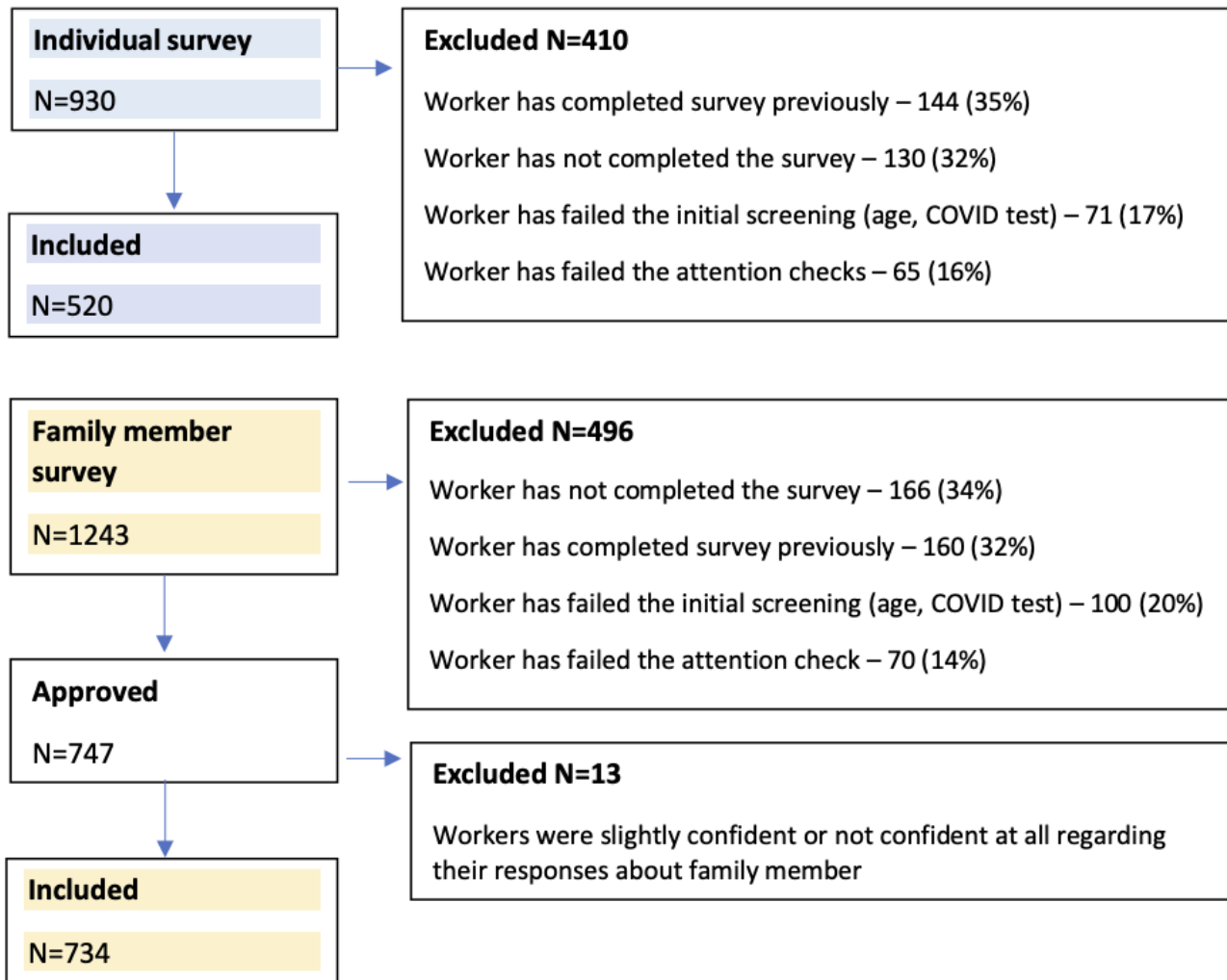
^aHIT: human intelligence task.

Survey Responses

After implementing our final crowdsourcing pipeline, we collected data from 930 individual surveys and 1243 family surveys; however, data from 410 individual surveys and 496 family surveys were excluded (late August to December 31, 2020). The reasons for exclusion were completion of the survey previously, noncompletion of the survey, initial screening failure for age or comprehension check, and attention check failure (Figure 1). Thus, we finally collected data from 1267 eligible COVID-19–positive participants, and of these, 520 were from individual surveys and 747 were from family surveys. Thirteen participants were further excluded as they were either only

slightly confident (n=12) or not confident at all (n=1) regarding their responses in the family survey. Thus, data from 1254 surveys were analyzed. The average time required to complete the general survey was 5.5 minutes.

Regarding family survey respondents, 68.3% (501/734) provided answers about a first-degree family member, 25.7% (189/734) provided answers about a second-degree family member, and only 6.0% (44/734) provided answers about a third-degree relative. There were no statistically significant differences in characteristics or outcomes between the individual respondents and the persons the respondents completed the family survey for, except for age (Multimedia Appendix 4). Therefore, the analysis presented here combined data from both surveys.

Figure 1. Study Inclusion and Exclusion of Amazon Mechanical Turk Worker Responses.

Demographic Characteristics

Over 90% (1159/1254, 92.4%) of the participants were up to 65 years old, and only 1.2% (15/1254) were less than 18 years old. Moreover, 52.0% (652/1254) were male, 81.2% (1018/1254) were white, 79.5% (997/1254) were not Hispanic or Latino, 68.4% (858/1254) had a bachelor's degree or any postgraduate degree, 14.4% (180/1254) had yearly income of US \$75,000 or more, 39.6% (496/1254) were smokers, and 46.8% (587/1254) had an influenza vaccine in the last season (Multimedia Appendix 5). Eight responders mentioned "passed away" in the family survey. As reflected in our sample, the MTurk worker population tended to be younger than the overall US population, with household incomes below that of the average US population [27].

Findings From Assessing Individual Symptoms Associated With Adverse COVID-19 Outcomes

Hospitalization

Overall, 47.6% (597/1254) of participants were hospitalized due to COVID-19. Bivariate analysis showed statistically significant differences between hospitalized and nonhospitalized COVID-19 participants for most demographic factors, except gender (Multimedia Appendix 5). Chronic conditions, including depression, hypertension, asthma, alcohol disorder, anemia,

weight loss, ulcer, lung/respiratory disease, bladder problems, bowel disease, and angina, were associated with more COVID-19 hospitalizations (Multimedia Appendix 6). COVID-19 symptoms associated with higher risk for hospitalization were cough with sputum, sneezing, abdominal pain, vomiting, confusion, bladder pain, dry eyes, dry skin, skin rash, and seizure (Multimedia Appendix 7).

From the logistic regression analysis of the total study population (Multimedia Appendix 8), we found statistically significant associations between the following participant characteristics and COVID-19 hospitalization compared with baseline: being in any age group over 24 years; having a bachelor's degree (odds ratio [OR] 2.57, 95% CI 1.43-4.66); smoking every day, smoking some days, or past smoking with quitting less than a year ago (OR 2.06, 95% CI 1.18-3.61; OR 3.41, 95% CI 2.2-5.33; and OR 3.39, 95% CI 1.88-6.24, respectively); and influenza vaccine in the last season (OR 3.09, 95% CI 2.18-4.41). Chronic conditions associated with higher risk for hospitalization were depression (OR 1.77, 95% CI 1.18-2.67), asthma (OR 3.83, 95% CI 2.22-6.78), diabetes (OR 2.66, 95% CI 1.46-4.96), and bladder problems (OR 5.51, 95% CI 1.28-27.13). COVID-19 symptoms associated with higher risk for hospitalization were abdominal pain (OR 2.02, 95% CI 1.15-3.59), bladder pain (OR 3.20, 95% CI 1.25-9.3), cough

with sputum (OR 2.60, 95% CI 1.77-3.86), fever with a temperature over 100.4°F (OR 1.50, 95% CI 1.04-2.17), and shortness of breath (OR 2.75, 95% CI 1.8-4.23).

Mechanical Ventilation

Overall, 66.8% (399/597) of hospitalized participants were connected to a mechanical ventilator (31.8% of all participants). There were 11 hospitalized participants from the family survey whose mechanical ventilation use was unknown to the survey respondents, and these participants were not included in the subsequent mechanical ventilation analysis. Smoking every day (OR 3.51, 95% CI 1.45-9.1), influenza vaccine in the last season (OR 3.65, 95% CI 2.29-5.89), loss of appetite (OR 2.07, 95% CI 1.09-4.02), tiredness and fatigue (OR 2.36, 95% CI 1.04-5.44), and vomiting (OR 2.68, 95% CI 1.3-5.71) were significantly associated with higher risk for mechanical ventilation (Multimedia Appendix 8).

Findings From Assessing COVID-19 Symptomology

We identified the following 6 symptomology groups using hierarchical cluster analysis (Figure 2): Group 1, abdominal and bladder pain; Group 2, flu-like symptoms (loss of smell/taste/appetite); Group 3, hoarseness and sputum production; Group 4, joint aches and stomach cramps; Group 5, skin or eye dryness and vomiting; and Group 6, no symptoms. We found sociodemographic and clinical differences between the symptomology groups (Table 2).

The flu-like symptoms group (Group 2) mostly represented the general study population. The abdominal and bladder pain group (Group 1) and the skin or eye dryness group (Group 5) had the highest hospitalization frequencies (153/227, 67.4% and 134/196, 68.4%, respectively). Both groups were characterized by a lower chance of high income (19/227, 8.4% and 21/196, 10.7%, respectively), more smoking (121/227, 53.3% and 102/196, 52.0%, respectively), and more influenza vaccinations (144/227, 63.4% and 102/196, 52.0%, respectively). The group with abdominal and bladder pain symptoms (Group 1) had higher proportions of Hispanic participants (82/227, 36.1%), asthma patients (64/227, 28.2%), alcohol disorder patients (64/227, 28.2%), and anemia patients (54/227, 23.8%). The group with skin or eye dryness (Group 5) had higher proportions of patients with depression (64/196, 32.7%), diabetes (28/196, 14.3%), weight loss (27/196, 13.8%), and ulcers (25/196, 12.8%). The group with joint aches and stomach cramps (Group 4) had lower proportions of hospitalization (65/158, 41.1%) and mechanical ventilation (31/158, 47.7%). Compared with the general study population, the asymptomatic group (Group 6) was younger (age 18-44 years; 70/85, 82.4%), had more males (54/85, 63.5%), had less white participants (60/85, 70.6%), had less Hispanic participants (7/85, 8.2%), had more participants with a high income (18/85, 21.2%), had less smokers (26/85, 30.6%), had less influenza vaccinations reported (30/85, 35.3%), had a higher proportion of participants with no chronic conditions (45/85, 52.9%), and had a very low risk for hospitalization (12/85, 14.1%).

Figure 2. Dendrogram for COVID-19 symptom clusters.

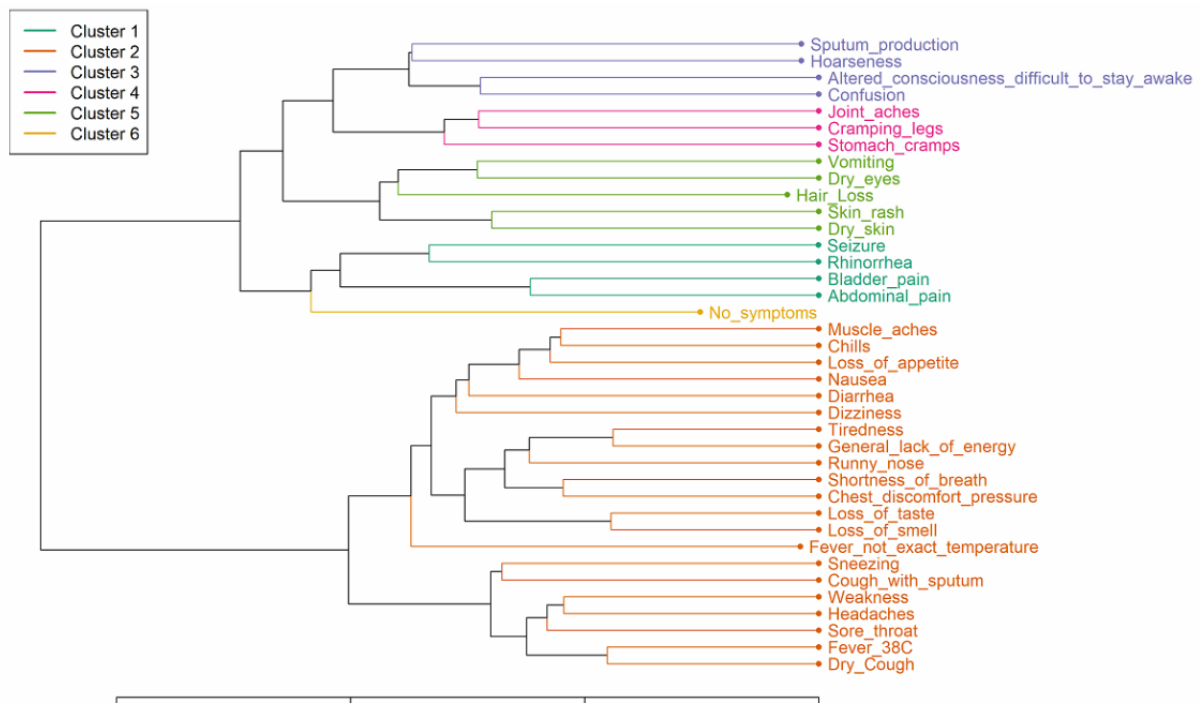


Table 2. Descriptive characteristics of COVID-19 symptomology groups.

Characteristic	Group 1 (abdominal and bladder pain) (N=227), n (%)	Group 2 (flu-like symptoms) (N=1139), n (%)	Group 3 (hoarseness, sputum production) (N=144), n (%)	Group 4 (joint aches, stomach cramps) (N=158), n (%)	Group 5 (skin or eye dryness) (N=196), n (%)	Group 6 (no symptoms) (N=85), n (%)
Hospitalization	153 (67.4)	621 (54.5)	79 (54.9)	65 (41.1)	134 (68.4)	12 (14.1)
Mechanical ventilation (among hospitalized patients)	112 (73.2)	366 (58.9)	46 (58.2)	31 (47.7)	86 (64.2)	10 (83.3)
Demographic characteristics						
Male gender	107 (47.1)	588 (51.6)	64 (44.4)	87 (55.1)	91 (46.4)	54 (63.5)
Age 18-44 years	148 (65.2)	727 (63.8)	82 (56.9)	94 (59.5)	133 (67.9)	70 (82.4)
Age ≥45 years	76 (33.5)	379 (33.3)	54 (37.5)	58 (36.7)	57 (29.1)	15 (17.6)
White race	186 (81.9)	929 (81.6)	118 (81.9)	134 (84.8)	168 (85.7)	60 (70.6)
Hispanic or Latino ethnicity	82 (36.1)	230 (20.2)	23 (16.0)	27 (17.1)	43 (21.9)	7 (8.2)
US \$75,000 or more yearly income	19 (8.4)	162 (14.2)	20 (13.9)	28 (17.7)	21 (10.7)	18 (21.2)
Smoking	121 (53.3)	450 (39.5)	53 (36.8)	45 (28.5)	102 (52.0)	26 (30.6)
Flu vaccination	144 (63.4)	533 (46.8)	62 (43.1)	69 (43.7)	102 (52.0)	30 (35.3)
Chronic conditions						
Depression	38 (16.7)	285 (25.0)	35 (24.3)	37 (23.4)	64 (32.7)	12 (14.1)
Obesity	32 (14.1)	165 (14.5)	35 (24.3)	37 (23.4)	29 (14.8)	6 (7.1)
Asthma	64 (28.2)	156 (13.7)	29 (20.1)	19 (12.0)	29 (14.8)	6 (7.1)
Alcohol or substance use disorder	64 (28.2)	128 (11.2)	19 (13.2)	17 (10.8)	25 (12.8)	7 (8.2)
Diabetes, uncomplicated	12 (5.3)	116 (10.2)	15 (10.4)	19 (12.0)	28 (14.3)	1 (1.2)
Mental illness	36 (15.9)	98 (8.6)	23 (16.0)	22 (13.9)	17 (8.7)	4 (4.7)
Migraines	23 (10.1)	81 (7.1)	15 (10.4)	23 (14.6)	15 (7.7)	4 (4.7)
Weight loss	20 (8.8)	76 (6.7)	14 (9.7)	15 (9.5)	27 (13.8)	3 (3.5)
Anemia	54 (23.8)	70 (6.1)	12 (8.3)	13 (8.2)	22 (11.2)	3 (3.5)
High cholesterol	7 (3.1)	69 (6.1)	14 (9.7)	20 (12.7)	9 (4.6)	3 (3.5)
Ulcer	10 (4.4)	68 (6.0)	5 (3.5)	11 (7.0)	25 (12.8)	3 (3.5)
No chronic condition	34 (15.0)	264 (23.2)	23 (16.0)	31 (19.6)	29 (14.8)	45 (52.9)

Symptomology Groups Associated With Adverse COVID-19 Outcomes

Our findings from the logistic regression models, using symptomology groups as risk factors for adverse COVID-19 outcomes and adjusted for all sociodemographic characteristics and comorbid conditions, showed the following 3 groups

associated with hospitalization: abdominal and bladder pain group (Group 1; OR 1.5, 95% CI 1.01-2.34); flu-like symptoms group (Group 2; OR 3.33, 95% CI 1.97-5.79); and skin or eye dryness group (Group 5; OR 1.63, 95% CI 1.07-2.52). No symptomology group was associated with a high risk for mechanical ventilation (Table 3).

Table 3. Associations between COVID-19 symptomology groups and adverse COVID-19 outcomes.

Symptomology group ^a	Hospitalization ^b			Mechanical ventilation ^b		
	OR ^c	95% CI	P value	OR	95% CI	P value
Abdominal and bladder pain group (Group 1)	1.54	1.01-2.34	.04	1.12	0.66-1.92	.68
Flu-like symptoms group (Group 2)	3.33	1.97-5.79	<.001	0.17	0.04-0.54	.01
Hoarseness and sputum production group (Group 3)	1.51	0.92-2.48	.10	1.09	0.57-2.11	.80
Joint aches and stomach cramps group (Group 4)	0.56	0.35-0.88	.01	0.54	0.27-1.07	.08
Skin or eye dryness group (Group 5)	1.63	1.07-2.52	.02	1.25	0.76-2.05	.39

^aGroup 6 (no symptoms) is excluded.

^bMultivariate logistic models adjusted for sociodemographic characteristics and comorbid conditions.

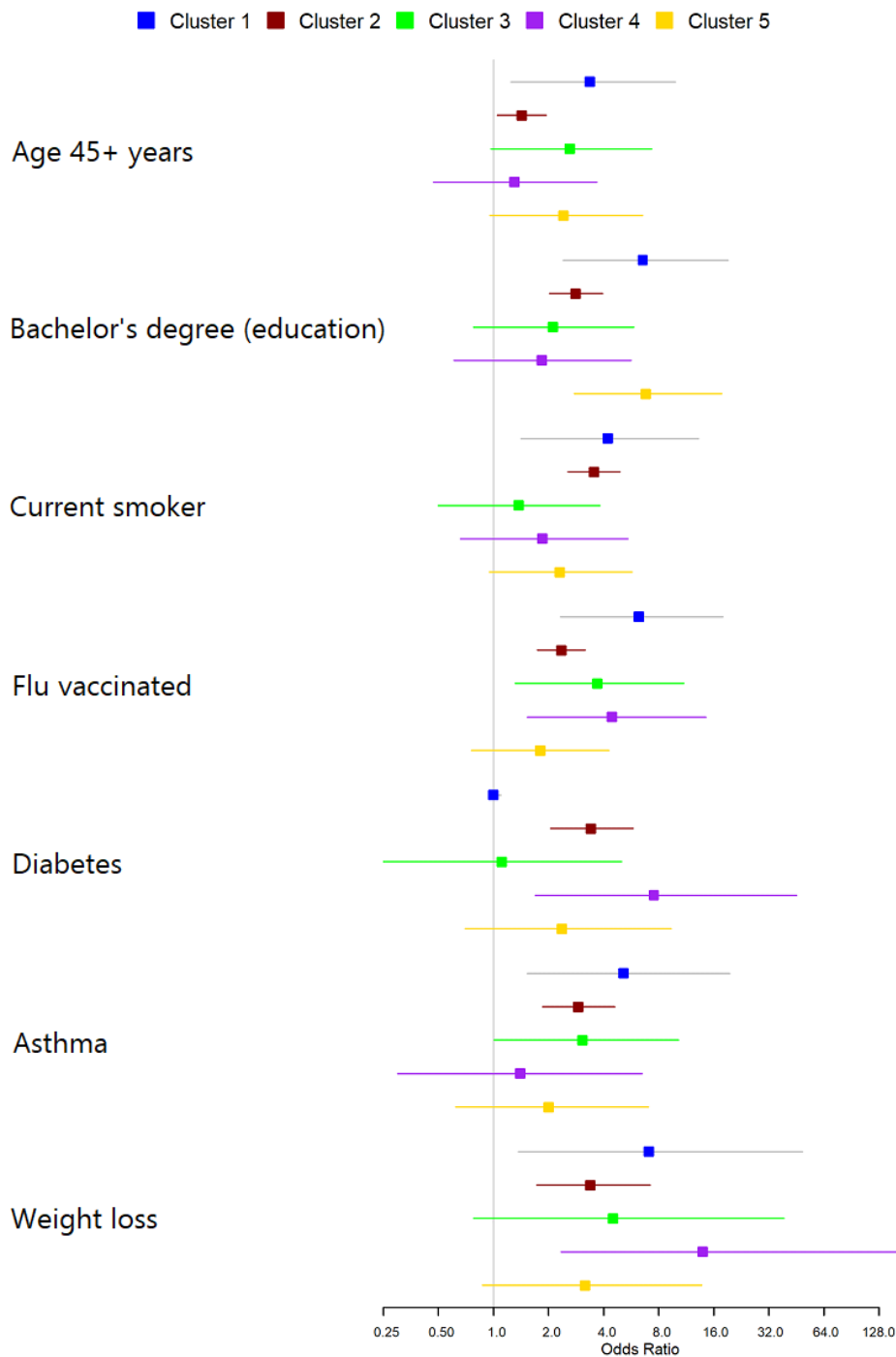
^cOR: odds ratio.

Risk Factors for COVID-19 Hospitalization Among Symptomology Groups

Finally, we developed 5 logistic regression models for symptomology groups to compare the risk factors for COVID-19 hospitalization among those groups (asymptomatic participants were excluded from this analysis). The results of those models are presented as a forest plot of significant variables in at least one symptomology group (Figure 3). The risk factors differed between participants from different symptomology groups. The only risk factor that was significant for 4 out of 5 groups was

influenza vaccine in the last season (Group 1: OR 6.22, 95% CI 2.32-17.92; Group 2: OR 2.35, 95% CI 1.74-3.18; Group 3: OR 3.7, 95% CI 1.32-10.98; Group 4: OR 4.44, 95% CI 1.53-14.49). Smoking (OR 4.22, 95% CI 1.42-13.26) and asthma (OR 5.14, 95% CI 1.53-19.56) were significant risk factors for hospitalization in the abdominal and bladder pain group (Group 1). Weight loss was a risk factor in the joint aches and stomach cramps group (Group 4; OR 13.9, 95% CI 2.34-161.64) and in the abdominal and bladder pain group (Group 1; OR 7.05, 95% CI 1.37-49.01). Diabetes was a risk factor in the joint aches and stomach cramps group (Group 4; OR 7.5, 95% CI 1.69-45.28).

Figure 3. Risk factors for hospitalization among individuals in different symptomology groups.



Findings From the Validation Assessment

A comparison of our findings with those of other studies can be found in [Multimedia Appendix 9](#). At the time of our analysis, we found 3 systematic review or meta-analysis studies mapping the association of symptoms with the risk of adverse outcomes of COVID-19 [19-21].

We found agreement between this study and previous studies for 18 symptoms, 6 of which were associated with adverse outcomes (abdominal pain, cough, dyspnea/shortness of breath, fever, fatigue, and vomiting). In addition, we assessed 14 symptoms that were not previously studied by others, 6 of which

were associated with adverse outcomes (bladder pain, dry eyes, dry skin, loss of appetite, seizure, and skin rash).

Discussion

Principal Findings

Our results identified individual symptoms and behaviors associated with COVID-19 adverse outcomes. Among these, some were well-known and some were new. We also identified 6 symptomology groups, with 3 groups showing statistically significant associations with COVID-19 outcomes. Furthermore, the findings of this work increase our understanding of the

MTurk population and show that with precautionary measures, high-quality data can be obtained.

Well-known single COVID-19 symptoms identified (ie, abdominal pain, cough, fever, and shortness of breath) were associated with hospitalization [5,6]. Less common symptoms identified, such as bladder pain, eye dryness, and skin dryness were also associated with adverse COVID-19 outcomes. We provided additional validation of our findings by comparing the results with the findings of systematic review and meta-analysis studies. The individual symptoms we identified as being associated with adverse COVID-19 outcomes were consistent with the symptoms in those studies.

Our analysis of chronic conditions and associations with COVID-19 adverse outcomes showed that patients with preexisting asthma, diabetes, depression, and bladder problems were at high risk for hospitalization, similar to the findings in previous studies. Although previous studies have shown an increased risk of severe COVID-19 among people with obesity [34], our study did not find a significant increase in the risk of hospitalization among obese people. This result may be due to the participants in our sample being younger than those in other studies, resulting in a weaker link between obesity and chronic diseases that are the actual drivers of COVID-19 severity.

When studying behaviors influencing adverse COVID-19 outcomes, like previous studies, we found that smoking increased the risk of severe COVID-19 outcomes [35-37]. Current smokers and past smokers who quit less than a year ago had a higher risk of hospitalization, and every day smokers also had a higher risk for mechanical ventilation. Our finding showing an effect of influenza vaccination on adverse outcomes contradicts the findings in some other studies. For example, it has been previously reported that influenza vaccination could be considered a protective factor against severe cases of COVID-19 infection [38,39]. Our data, however, suggested that COVID-19-positive respondents who were vaccinated against influenza in Autumn 2019 had higher odds of hospitalization and mechanical ventilation after adjusting for demographic factors, chronic conditions, and COVID-19 symptoms, as the influenza vaccination status might be associated with preexisting comorbidities and a person's demographics. This is not an isolated finding as others have reported that there is a positive association between influenza vaccination rates and COVID-19 death rates [40], that influenza vaccination coverage in a country is a risk factor associated with higher infection rates of COVID-19 [41], and that there is a need to investigate the potential impact of influenza vaccination on COVID-19 risk and severity [42].

In addition to studying individual symptoms and behaviors, this study identified 6 COVID-19 symptomology groups by cluster analysis and assessed their associations with adverse outcomes of the disease. Three symptomology groups (flu-like symptoms, abdominal and bladder pain symptoms, and eye and skin dryness symptoms) were highly associated with a high risk for hospitalization. While the characteristics of respondents in the flu-like symptoms group were similar to the characteristics of the general population, the abdominal and bladder pain group included survey respondents who had lower income, and were

more likely to have smoked and to be influenza vaccinated. They also tended to have chronic conditions, such as asthma and anemia, and alcohol disorder. The survey respondents in the eye and skin dryness group were generally older and had a greater possibility of being white. They were also more likely to have smoked and to be influenza vaccinated. This group also had a very high percentage of survey respondents with depression, diabetes, and ulcers.

Characterizing patients according to clusters using artificial intelligence devices and machine learning is a pioneering method in a variety of infectious and noninfectious diseases. The use of scientific methods to identify clusters of patients with similar characteristics and specific disease risks might improve awareness of heterogeneity in symptomology, and may enable targeted interventions to reduce disease severity. Other studies of COVID-19 disease trajectories have been able to identify vulnerable population clusters that could benefit from specific health resources, and have provided insights for public health targets for managing the pandemic [43,44]. One previous study identified 3 symptomatic groups and 1 asymptomatic group among COVID-19 patients [43]. However, that study did not analyze the associations between the symptomology groups and COVID-19 outcomes. Our analysis of 6 symptomatic groups found that the risk factors for COVID-19 adverse outcomes differed between participants from the different symptomology groups. For the asymptomatic COVID-19 group, other studies have shown that asymptomatic carriers account for 15% to 60% of the infected population and play a key role in disease transmission [45]. Adding to our understanding of asymptomatic carriers, our findings indicated that the asymptomatic symptomology group had a low percentage of hospitalization; a high percentage of young non-Hispanic men with high income; and a low percentage of people with chronic conditions, smoking, and influenza vaccination. These characteristics add to those described in a review study of asymptomatic COVID-19 carriers' characteristics that found young age alone to be a significant factor for having no symptoms [46-48]. Another study of Mexican outpatients found a lower frequency of smokers and influenza vaccination among asymptomatic responders [43].

The percentage of those connected to a mechanical ventilator among hospitalized patients may seem high in our study (61.8%); however, the management of patients hospitalized with COVID-19 has changed considerably over the course of the pandemic. More than half of the study population had been hospitalized, and two-thirds of them were on ventilators. Since the survey was conducted in the first months of the COVID-19 pandemic, many people who got sick with COVID-19 were hospitalized and then connected to a mechanical ventilator. Over time, fewer people with COVID-19 were hospitalized, and among those who were hospitalized, only patients with more severe disease were put on ventilators. Other studies have also shown a high percentage (68%) of ventilator use among hospitalized COVID patients [49].

This work also showed that with precautionary measures to ensure high-quality data collection, a crowdsourcing model can be used to collect data to characterize symptomology for COVID-19 diagnosis and prognosis. There are many studies

assessing health data on MTurk as a source of high-quality and rapidly collected data, and it has demonstrated good reliability [30,50,51]. However, to improve data quality on MTurk, there are recommendations to include workers with an “approval rate” above 95% and keep the “number of HITs approved” to at least 100 [30,31,51]. Prior studies have not investigated data quality from workers by comparing survey responses of 3 experience levels (according to the number of HITs approved) in a pilot study. By launching a pilot study, we found no difference in the approval rate of workers from different experience groups; thus, all could provide adequate data to satisfy the basic approval criterion. For specific tasks requiring higher cognitive ability, however, workers with more experience may provide higher quality data. In our case, we found that those with 500+ HITs submitted fewer duplicate responses than those with 100-499 HITs. While this may exacerbate the superworker issue, the tradeoff of quality data for the use of more experienced workers may be necessary depending on the task. To provide additional validation of our findings, we compared the findings of individual symptoms associated with COVID-19 to the findings of other researchers and identified many concordant findings.

Limitations

A major limitation of this study was the self-reported data, which can be less reliable than physiological assessments. Our crowdsourced approach, however, allowed for reaching many participants, which helped mitigate the noise, and the fast data collection process was helpful during this pandemic. In addition, during this pandemic, many risk factors of COVID-19 were discovered through social media and other self-reported surveys [52-56]. To use those data sources, crowdsourced practices are emerging in research fields such as *infodemiology* (defined as

collecting and analyzing data in real time through an electronic medium with the aim to inform public health decision makers) [57-59]. Another growing field is *digital epidemiology*, in which researchers are using internet data for epidemiological purposes [60-62]. The techniques of capturing relevant real-world data are promising but need to be further developed to meet the possible public health challenges in the future. Second, some of our findings warrant further validation. The risk factors first reported in our study, such as bladder pain symptoms and eye or skin dryness symptoms, need to be more extensively studied so that they can be used in clinical assessments. Furthermore, the influence of influenza vaccination on COVID-19 adverse outcomes should be further investigated as it appears now that humans will have to co-exist with both diseases for a long time even after this pandemic.

Conclusions

Our work demonstrated that a crowdsourced approach was effective for collecting data to assess the symptomology associated with COVID-19. Conducting a pilot study to assess data quality and population representation facilitated refining the filtering criteria for our final data collection strategy. We validated our approach by comparing the findings from assessing individual symptoms associated with COVID-19 to those identified by others and found highly concordant results. In our assessment of symptomology groups, we discovered that the bladder pain and skin or eye dryness groups had a high risk of COVID-19 hospitalization. Given these findings, we believe that a crowdsourcing strategy, such as the one proposed here, should be considered by others for quick and cost-effective assessments in a rapidly changing spectrum of infectious diseases, and societal and environmental factors.

Acknowledgments

We would like to thank Dr Dina Demner-Fushman (National Library of Medicine, National Institutes of Health) for her helpful feedback. We would also like to thank the Johns Hopkins COVID-19 Research Response Program Community Research Working Group for providing access to the toolkit used by this research team to create a survey that includes harmonized data elements. This group is led by Dr Shruti Mehta (Public Health), Dr Jason Farley (Nursing), and Dr Jacky Jennings (Medicine). This work was supported in part by a Microsoft Investigator Fellowship awarded to COT.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Individual survey of COVID-19 symptoms.

[\[DOCX File , 27 KB-Multimedia Appendix 1\]](#)

Multimedia Appendix 2

Family survey of COVID-19 symptoms.

[\[DOCX File , 31 KB-Multimedia Appendix 2\]](#)

Multimedia Appendix 3

Comparison of demographic characteristics across all experience groups among approved Amazon Mechanical Turk workers.

[\[DOCX File , 27 KB-Multimedia Appendix 3\]](#)

Multimedia Appendix 4

Descriptive characteristics of participants in the individual and family surveys.

[\[DOCX File , 32 KB-Multimedia Appendix 4\]](#)

Multimedia Appendix 5

Demographic characteristics of the study participants.

[\[DOCX File , 27 KB-Multimedia Appendix 5\]](#)

Multimedia Appendix 6

Chronic condition characteristics of the study participants.

[\[DOCX File , 31 KB-Multimedia Appendix 6\]](#)

Multimedia Appendix 7

COVID-19 symptom characteristics of the study participants.

[\[DOCX File , 30 KB-Multimedia Appendix 7\]](#)

Multimedia Appendix 8

Associations between symptoms and adverse COVID-19 outcomes adjusted for sociodemographic factors and chronic conditions (multivariate logistic regression).

[\[DOCX File , 41 KB-Multimedia Appendix 8\]](#)

Multimedia Appendix 9

Comparison of our findings with those of systematic review and meta-analysis studies regarding the association between COVID-19 symptoms and adverse outcomes.

[\[DOCX File , 18 KB-Multimedia Appendix 9\]](#)

References

1. WHO Coronavirus (COVID-19) Dashboard. World Health Organization. URL: <https://covid19.who.int/> [accessed 2022-11-10]
2. Symptoms of COVID-19. Centers for Disease Control and Prevention. URL: <https://www.cdc.gov/coronavirus/2019-ncov/symptoms-testing/symptoms.html> [accessed 2022-11-10]
3. CDC COVID-19 Response Team. Severe Outcomes Among Patients with Coronavirus Disease 2019 (COVID-19) - United States, February 12-March 16, 2020. *MMWR Morb Mortal Wkly Rep* 2020 Mar 27;69(12):343-346 [FREE Full text] [doi: [10.15585/mmwr.mm6912e2](https://doi.org/10.15585/mmwr.mm6912e2)] [Medline: [32214079](https://pubmed.ncbi.nlm.nih.gov/32214079/)]
4. Booth A, Reed AB, Ponzo S, Yassae A, Aral M, Plans D, et al. Population risk factors for severe disease and mortality in COVID-19: A global systematic review and meta-analysis. *PLoS One* 2021 Mar 4;16(3):e0247461 [FREE Full text] [doi: [10.1371/journal.pone.0247461](https://doi.org/10.1371/journal.pone.0247461)] [Medline: [33661992](https://pubmed.ncbi.nlm.nih.gov/33661992/)]
5. Berek MA, Aziz MA, Islam MS. Impact of age, sex, comorbidities and clinical symptoms on the severity of COVID-19 cases: A meta-analysis with 55 studies and 10014 cases. *Heliyon* 2020 Dec;6(12):e05684 [FREE Full text] [doi: [10.1016/j.heliyon.2020.e05684](https://doi.org/10.1016/j.heliyon.2020.e05684)] [Medline: [33344791](https://pubmed.ncbi.nlm.nih.gov/33344791/)]
6. Li J, Huang DQ, Zou B, Yang H, Hui WZ, Rui F, et al. Epidemiology of COVID-19: A systematic review and meta-analysis of clinical characteristics, risk factors, and outcomes. *J Med Virol* 2021 Mar 25;93(3):1449-1458 [FREE Full text] [doi: [10.1002/jmv.26424](https://doi.org/10.1002/jmv.26424)] [Medline: [32790106](https://pubmed.ncbi.nlm.nih.gov/32790106/)]
7. Jain V, Yuan J. Predictive symptoms and comorbidities for severe COVID-19 and intensive care unit admission: a systematic review and meta-analysis. *Int J Public Health* 2020 Jun 25;65(5):533-546 [FREE Full text] [doi: [10.1007/s00038-020-01390-7](https://doi.org/10.1007/s00038-020-01390-7)] [Medline: [32451563](https://pubmed.ncbi.nlm.nih.gov/32451563/)]
8. Rodriguez-Morales AJ, Cardona-Ospina JA, Gutiérrez-Ocampo E, Villamizar-Peña R, Holguin-Rivera Y, Escalera-Antezana JP, Latin American Network of Coronavirus Disease 2019-COVID-19 Research (LANCOVID-19). Electronic address: <https://www.lancovid.org>. Clinical, laboratory and imaging features of COVID-19: A systematic review and meta-analysis. *Travel Med Infect Dis* 2020 Mar;34:101623 [FREE Full text] [doi: [10.1016/j.tmaid.2020.101623](https://doi.org/10.1016/j.tmaid.2020.101623)] [Medline: [32179124](https://pubmed.ncbi.nlm.nih.gov/32179124/)]
9. Zheng Z, Peng F, Xu B, Zhao J, Liu H, Peng J, et al. *J Infect* 2020 Aug;81(2):e16-e25 [FREE Full text] [doi: [10.1016/j.jinf.2020.04.021](https://doi.org/10.1016/j.jinf.2020.04.021)] [Medline: [32335169](https://pubmed.ncbi.nlm.nih.gov/32335169/)]
10. Ahmed A, Ali A, Hasan S. Comparison of Epidemiological Variations in COVID-19 Patients Inside and Outside of China-A Meta-Analysis. *Front Public Health* 2020;8:193 [FREE Full text] [doi: [10.3389/fpubh.2020.00193](https://doi.org/10.3389/fpubh.2020.00193)] [Medline: [32574293](https://pubmed.ncbi.nlm.nih.gov/32574293/)]

11. Li L, Huang T, Wang Y, Wang Z, Liang Y, Huang T, et al. COVID-19 patients' clinical characteristics, discharge rate, and fatality rate of meta-analysis. *J Med Virol* 2020 Jun 23;92(6):577-583 [FREE Full text] [doi: [10.1002/jmv.25757](https://doi.org/10.1002/jmv.25757)] [Medline: [32162702](https://pubmed.ncbi.nlm.nih.gov/32162702/)]
12. Sun P, Qie S, Liu Z, Ren J, Li K, Xi J. Clinical characteristics of hospitalized patients with SARS-CoV-2 infection: A single arm meta-analysis. *J Med Virol* 2020 Jun;92(6):612-617 [FREE Full text] [doi: [10.1002/jmv.25735](https://doi.org/10.1002/jmv.25735)] [Medline: [32108351](https://pubmed.ncbi.nlm.nih.gov/32108351/)]
13. Ioannou GN, Locke E, Green P, Berry K, O'Hare AM, Shah JA, et al. Risk Factors for Hospitalization, Mechanical Ventilation, or Death Among 10 131 US Veterans With SARS-CoV-2 Infection. *JAMA Netw Open* 2020 Sep 01;3(9):e2022310 [FREE Full text] [doi: [10.1001/jamanetworkopen.2020.22310](https://doi.org/10.1001/jamanetworkopen.2020.22310)] [Medline: [32965502](https://pubmed.ncbi.nlm.nih.gov/32965502/)]
14. Nobel YR, Phipps M, Zucker J, Lebowitz B, Wang TC, Sobieszczyk ME, et al. Gastrointestinal Symptoms and Coronavirus Disease 2019: A Case-Control Study From the United States. *Gastroenterology* 2020 Jul;159(1):373-375.e2 [FREE Full text] [doi: [10.1053/j.gastro.2020.04.017](https://doi.org/10.1053/j.gastro.2020.04.017)] [Medline: [32294477](https://pubmed.ncbi.nlm.nih.gov/32294477/)]
15. Gupta A, Madhavan MV, Sehgal K, Nair N, Mahajan S, Sehrawat TS, et al. Extrapulmonary manifestations of COVID-19. *Nat Med* 2020 Jul 10;26(7):1017-1032. [doi: [10.1038/s41591-020-0968-3](https://doi.org/10.1038/s41591-020-0968-3)]
16. Wang H, Li X, Yan Z, Sun X, Han J, Zhang B. Potential neurological symptoms of COVID-19. *Ther Adv Neurol Disord* 2020 Mar 28;13:1756286420917830 [FREE Full text] [doi: [10.1177/1756286420917830](https://doi.org/10.1177/1756286420917830)] [Medline: [32284735](https://pubmed.ncbi.nlm.nih.gov/32284735/)]
17. Grant MC, Geoghegan L, Arbyn M, Mohammed Z, McGuinness L, Clarke EL, et al. The Prevalence of Symptoms in 24,410 Adults Infected by the Novel Coronavirus (SARS-CoV-2; COVID-19): A Systematic Review and Meta-Analysis of 148 Studies from 9 Countries. *SSRN Journal* 2020:3582819. [doi: [10.2139/ssrn.3582819](https://doi.org/10.2139/ssrn.3582819)]
18. Santosh R, Schwartz HA, Eichstaedt J, Ungar L, Guntuku SC. Detecting Emerging Symptoms of COVID-19 using Context-based Twitter Embeddings. In: Proceedings of the 1st Workshop on NLP for COVID-19 (Part 2) at EMNLP 2020. 2020 Presented at: 1st Workshop on NLP for COVID-19 (Part 2) at EMNLP 2020; December 2020; Online. [doi: [10.18653/v1/2020.nlpCOVID19-2.35](https://doi.org/10.18653/v1/2020.nlpCOVID19-2.35)]
19. Li J, Chen Z, Nie Y, Ma Y, Guo Q, Dai X. Identification of Symptoms Prognostic of COVID-19 Severity: Multivariate Data Analysis of a Case Series in Henan Province. *J Med Internet Res* 2020 Jun 30;22(6):e19636 [FREE Full text] [doi: [10.2196/19636](https://doi.org/10.2196/19636)] [Medline: [32544071](https://pubmed.ncbi.nlm.nih.gov/32544071/)]
20. Talukder A, Razu SR, Alif SM, Rahman MA, Islam SMS. Association Between Symptoms and Severity of Disease in Hospitalised Novel Coronavirus (COVID-19) Patients: A Systematic Review and Meta-Analysis. *J Multidiscip Healthc* 2022;15:1101-1110 [FREE Full text] [doi: [10.2147/JMDH.S357867](https://doi.org/10.2147/JMDH.S357867)] [Medline: [35592814](https://pubmed.ncbi.nlm.nih.gov/35592814/)]
21. Cui L, Wang C, Wu Z, Peng D, Huang J, Zhang C, et al. Symptomatology differences of major depression in psychiatric versus general hospitals: A machine learning approach. *J Affect Disord* 2020 Jan 01;260:349-360. [doi: [10.1016/j.jad.2019.09.030](https://doi.org/10.1016/j.jad.2019.09.030)] [Medline: [31521873](https://pubmed.ncbi.nlm.nih.gov/31521873/)]
22. Sudre CH, Lee KA, Lochlainn MN, Varsavsky T, Murray B, Graham MS, et al. Symptom clusters in COVID-19: A potential clinical prediction tool from the COVID Symptom Study app. *Sci Adv* 2021 Mar 19;7(12):eabd4177 [FREE Full text] [doi: [10.1126/sciadv.abd4177](https://doi.org/10.1126/sciadv.abd4177)] [Medline: [33741586](https://pubmed.ncbi.nlm.nih.gov/33741586/)]
23. Vermicelli S, Cricelli L, Grimaldi M. How can crowdsourcing help tackle the COVID - 19 pandemic? An explorative overview of innovative collaborative practices. *R&D Management* 2020 Nov 15;51(2):183-194. [doi: [10.1111/radm.12443](https://doi.org/10.1111/radm.12443)]
24. Desai A, Warner J, Kuderer N, Thompson M, Painter C, Lyman G, et al. Crowdsourcing a crisis response for COVID-19 in oncology. *Nat Cancer* 2020 May 21;1(5):473-476 [FREE Full text] [doi: [10.1038/s43018-020-0065-z](https://doi.org/10.1038/s43018-020-0065-z)] [Medline: [35121981](https://pubmed.ncbi.nlm.nih.gov/35121981/)]
25. Crowdsourcing COVID-19. Harvard College. URL: <https://hms.harvard.edu/news/crowdsourcing-covid-19> [accessed 2022-11-10]
26. Wynants L, Van Calster B, Collins GS, Riley RD, Heinze G, Schuit E, et al. Prediction models for diagnosis and prognosis of covid-19: systematic review and critical appraisal. *BMJ* 2020 Apr 07;369:m1328 [FREE Full text] [doi: [10.1136/bmj.m1328](https://doi.org/10.1136/bmj.m1328)] [Medline: [32265220](https://pubmed.ncbi.nlm.nih.gov/32265220/)]
27. Difallah D, Filatova E, Ipeirotis P. Demographics and Dynamics of Mechanical Turk Workers. In: WSDM '18: Proceedings of the Eleventh ACM International Conference on Web Search and Data Mining. 2018 Presented at: Eleventh ACM International Conference on Web Search and Data Mining; February 5-9, 2018; Marina Del Rey, CA, USA p. 135-143. [doi: [10.1145/3159652.3159661](https://doi.org/10.1145/3159652.3159661)]
28. Moss AJ, Rosenzweig C, Robinson J, Litman L. Demographic Stability on Mechanical Turk Despite COVID-19. *Trends Cogn Sci* 2020 Sep;24(9):678-680 [FREE Full text] [doi: [10.1016/j.tics.2020.05.014](https://doi.org/10.1016/j.tics.2020.05.014)] [Medline: [32553445](https://pubmed.ncbi.nlm.nih.gov/32553445/)]
29. Mortensen K, Hughes TL. Comparing Amazon's Mechanical Turk Platform to Conventional Data Collection Methods in the Health and Medical Research Literature. *J Gen Intern Med* 2018 Apr 4;33(4):533-538 [FREE Full text] [doi: [10.1007/s11606-017-4246-0](https://doi.org/10.1007/s11606-017-4246-0)] [Medline: [29302882](https://pubmed.ncbi.nlm.nih.gov/29302882/)]
30. Young J, Young KM. Don't Get Lost in the Crowd: Best Practices for Using Amazon's Mechanical Turk in Behavioral Research. *Journal of the Midwest Association for Information Systems* 2019;2019(2):Article 2 [FREE Full text] [doi: [10.17705/3jmw.000050](https://doi.org/10.17705/3jmw.000050)]
31. Robinson J, Rosenzweig C, Moss AJ, Litman L. Tapped out or barely tapped? Recommendations for how to harness the vast and largely unused potential of the Mechanical Turk participant pool. *PLoS One* 2019 Dec 16;14(12):e0226394 [FREE Full text] [doi: [10.1371/journal.pone.0226394](https://doi.org/10.1371/journal.pone.0226394)] [Medline: [31841534](https://pubmed.ncbi.nlm.nih.gov/31841534/)]

32. COVID-19 Community Response Survey. PhenX Toolkit. URL: <https://www.phenxtoolkit.org/protocols/view/940301> [accessed 2022-11-14]
33. Sarker A, Lakamana S, Hogg-Bremer W, Xie A, Al-Garadi MA, Yang YC. Self-reported COVID-19 symptoms on Twitter: an analysis and a research resource. *J Am Med Inform Assoc* 2020 Aug 01;27(8):1310-1315 [FREE Full text] [doi: [10.1093/jamia/ocaa116](https://doi.org/10.1093/jamia/ocaa116)] [Medline: [32620975](https://pubmed.ncbi.nlm.nih.gov/32620975/)]
34. Steenblock C, Hassanein M, Khan EG, Yaman M, Kamel M, Barbir M, et al. Obesity and COVID-19: What are the Consequences? *Horm Metab Res* 2022 Aug 09;54(8):496-502 [FREE Full text] [doi: [10.1055/a-1878-9757](https://doi.org/10.1055/a-1878-9757)] [Medline: [35724688](https://pubmed.ncbi.nlm.nih.gov/35724688/)]
35. Reddy RK, Charles WN, Sklavounos A, Dutt A, Seed PT, Khajuria A. The effect of smoking on COVID-19 severity: A systematic review and meta-analysis. *J Med Virol* 2021 Feb 13;93(2):1045-1056 [FREE Full text] [doi: [10.1002/jmv.26389](https://doi.org/10.1002/jmv.26389)] [Medline: [32749705](https://pubmed.ncbi.nlm.nih.gov/32749705/)]
36. Patanavanich R, Glantz S. Smoking Is Associated With COVID-19 Progression: A Meta-analysis. *Nicotine Tob Res* 2020 Aug 24;22(9):1653-1656 [FREE Full text] [doi: [10.1093/ntr/ntaa082](https://doi.org/10.1093/ntr/ntaa082)] [Medline: [32399563](https://pubmed.ncbi.nlm.nih.gov/32399563/)]
37. van Zyl-Smit RN, Richards G, Leone FT. Tobacco smoking and COVID-19 infection. *The Lancet Respiratory Medicine* 2020 Jul;8(7):664-665. [doi: [10.1016/s2213-2600\(20\)30239-3](https://doi.org/10.1016/s2213-2600(20)30239-3)]
38. Yang M, Rooks BJ, Le TT, Santiago IO, Diamond J, Dorsey NL, et al. Influenza Vaccination and Hospitalizations Among COVID-19 Infected Adults. *J Am Board Fam Med* 2021 Feb 23;34(Supplement):S179-S182. [doi: [10.3122/jabfm.2021.s1.200528](https://doi.org/10.3122/jabfm.2021.s1.200528)]
39. Zein JG, Whelan G, Erzurum SC. Safety of influenza vaccine during COVID-19. *J. Clin. Trans. Sci* 2020 Sep 17;5(1):e49. [doi: [10.1017/cts.2020.543](https://doi.org/10.1017/cts.2020.543)]
40. Walach H, Klement R. Influenza Vaccination Rates Predict 30% of the Variance in Covid-19 Related Deaths in Europe – A Modeling Approach. Zenodo. URL: <https://zenodo.org/record/4067724#.Y2y0auRBzDc> [accessed 2022-11-10]
41. Lisewski AM. Association between Influenza Vaccination Rates and SARS-CoV-2 Outbreak Infection Rates in OECD Countries. *SSRN Journal* 2020:3558270. [doi: [10.2139/ssrn.3558270](https://doi.org/10.2139/ssrn.3558270)]
42. De Wals P, Divangahi M. Could seasonal influenza vaccination influence COVID-19 risk? medRxiv. URL: <https://www.medrxiv.org/content/10.1101/2020.09.02.20186734v1.full> [accessed 2022-11-10]
43. Fernández-Rojas MA, Luna-Ruiz Esparza MA, Campos-Romero A, Calva-Espinosa DY, Moreno-Camacho JL, Langle-Martínez AP, et al. Epidemiology of COVID-19 in Mexico: Symptomatic profiles and presymptomatic people. *Int J Infect Dis* 2021 Mar;104:572-579 [FREE Full text] [doi: [10.1016/j.ijid.2020.12.086](https://doi.org/10.1016/j.ijid.2020.12.086)] [Medline: [33434668](https://pubmed.ncbi.nlm.nih.gov/33434668/)]
44. San-Cristobal R, Martín-Hernández R, Ramos-Lopez O, Martinez-Urbistondo D, Micó V, Colmenarejo G, et al. Longwise Cluster Analysis for the Prediction of COVID-19 Severity within 72 h of Admission: COVID-DATA-SAVE-LIFES Cohort. *J Clin Med* 2022 Jun 10;11(12):3327 [FREE Full text] [doi: [10.3390/jcm11123327](https://doi.org/10.3390/jcm11123327)] [Medline: [35743398](https://pubmed.ncbi.nlm.nih.gov/35743398/)]
45. Nikolai LA, Meyer CG, Kreamsner PG, Velavan TP. Asymptomatic SARS Coronavirus 2 infection: Invisible yet invincible. *Int J Infect Dis* 2020 Nov;100:112-116 [FREE Full text] [doi: [10.1016/j.ijid.2020.08.076](https://doi.org/10.1016/j.ijid.2020.08.076)] [Medline: [32891737](https://pubmed.ncbi.nlm.nih.gov/32891737/)]
46. Zhang W, Wu X, Zhou H, Xu F. Clinical characteristics and infectivity of asymptomatic carriers of SARS-CoV-2 (Review). *Exp Ther Med* 2021 Feb;21(2):115 [FREE Full text] [doi: [10.3892/etm.2020.9547](https://doi.org/10.3892/etm.2020.9547)] [Medline: [33335578](https://pubmed.ncbi.nlm.nih.gov/33335578/)]
47. Hu Z, Song C, Xu C, Jin G, Chen Y, Xu X, et al. Clinical characteristics of 24 asymptomatic infections with COVID-19 screened among close contacts in Nanjing, China. *Sci China Life Sci* 2020 May 04;63(5):706-711 [FREE Full text] [doi: [10.1007/s11427-020-1661-4](https://doi.org/10.1007/s11427-020-1661-4)] [Medline: [32146694](https://pubmed.ncbi.nlm.nih.gov/32146694/)]
48. Wang Y, Liu Y, Liu L, Wang X, Luo N, Li L. Clinical Outcomes in 55 Patients With Severe Acute Respiratory Syndrome Coronavirus 2 Who Were Asymptomatic at Hospital Admission in Shenzhen, China. *J Infect Dis* 2020 May 11;221(11):1770-1774 [FREE Full text] [doi: [10.1093/infdis/jiaa119](https://doi.org/10.1093/infdis/jiaa119)] [Medline: [32179910](https://pubmed.ncbi.nlm.nih.gov/32179910/)]
49. Anesi G, Jablonski J, Harhay MO, Atkins JH, Bajaj J, Baston C, et al. Characteristics, Outcomes, and Trends of Patients With COVID-19-Related Critical Illness at a Learning Health System in the United States. *Ann Intern Med* 2021 May;174(5):613-621 [FREE Full text] [doi: [10.7326/M20-5327](https://doi.org/10.7326/M20-5327)] [Medline: [33460330](https://pubmed.ncbi.nlm.nih.gov/33460330/)]
50. Hauser DJ, Schwarz N. Attentive Turkers: MTurk participants perform better on online attention checks than do subject pool participants. *Behav Res Methods* 2016 Mar 12;48(1):400-407. [doi: [10.3758/s13428-015-0578-z](https://doi.org/10.3758/s13428-015-0578-z)] [Medline: [25761395](https://pubmed.ncbi.nlm.nih.gov/25761395/)]
51. Moss A, Litman L. After the Bot Scare: Understanding What's Been Happening With Data Collection on MTurk and How to Stop It. CloudResearch. URL: <https://www.cloudresearch.com/resources/blog/after-the-bot-scare-understanding-whats-been-happening-with-data-collection-on-mturk-and-how-to-stop-it/> [accessed 2022-11-10]
52. Menni C, Valdes AM, Freidin MB, Sudre CH, Nguyen LH, Drew DA, et al. Real-time tracking of self-reported symptoms to predict potential COVID-19. *Nat Med* 2020 Jul 11;26(7):1037-1040 [FREE Full text] [doi: [10.1038/s41591-020-0916-2](https://doi.org/10.1038/s41591-020-0916-2)] [Medline: [32393804](https://pubmed.ncbi.nlm.nih.gov/32393804/)]
53. Quer G, Radin JM, Gadaleta M, Baca-Motes K, Ariniello L, Ramos E, et al. Wearable sensor data and self-reported symptoms for COVID-19 detection. *Nat Med* 2021 Jan 29;27(1):73-77. [doi: [10.1038/s41591-020-1123-x](https://doi.org/10.1038/s41591-020-1123-x)] [Medline: [33122860](https://pubmed.ncbi.nlm.nih.gov/33122860/)]
54. Zens M, Brammertz A, Herpich J, Südkamp N, Hinterseer M. App-Based Tracking of Self-Reported COVID-19 Symptoms: Analysis of Questionnaire Data. *J Med Internet Res* 2020 Sep 09;22(9):e21956 [FREE Full text] [doi: [10.2196/21956](https://doi.org/10.2196/21956)] [Medline: [32791493](https://pubmed.ncbi.nlm.nih.gov/32791493/)]

55. Peters A, Rospleszcz S, Greiser KH, Dallavalle M, Berger K, Complete list of authors available under: Collaborators. The Impact of the COVID-19 Pandemic on Self-Reported Health. *Dtsch Arztebl Int* 2020 Dec 11;117(50):861-867 [[FREE Full text](#)] [doi: [10.3238/arztebl.2020.0861](https://doi.org/10.3238/arztebl.2020.0861)] [Medline: [33295275](https://pubmed.ncbi.nlm.nih.gov/33295275/)]
56. Dreyer NA, Reynolds M, DeFilippo Mack C, Brinkley E, Petruski-Ivleva N, Hawaldar K, et al. Self-reported symptoms from exposure to Covid-19 provide support to clinical diagnosis, triage and prognosis: An exploratory analysis. *Travel Med Infect Dis* 2020 Nov;38:101909 [[FREE Full text](#)] [doi: [10.1016/j.tmaid.2020.101909](https://doi.org/10.1016/j.tmaid.2020.101909)] [Medline: [33152512](https://pubmed.ncbi.nlm.nih.gov/33152512/)]
57. Eysenbach G. Infodemiology: the epidemiology of (mis)information. *The American Journal of Medicine* 2002 Dec;113(9):763-765. [doi: [10.1016/s0002-9343\(02\)01473-0](https://doi.org/10.1016/s0002-9343(02)01473-0)]
58. Eysenbach G. Infodemiology and infoveillance tracking online health information and cyberbehavior for public health. *Am J Prev Med* 2011 May;40(5 Suppl 2):S154-S158. [doi: [10.1016/j.amepre.2011.02.006](https://doi.org/10.1016/j.amepre.2011.02.006)] [Medline: [21521589](https://pubmed.ncbi.nlm.nih.gov/21521589/)]
59. Mavragani A. Tracking COVID-19 in Europe: Infodemiology Approach. *JMIR Public Health Surveill* 2020 Apr 20;6(2):e18941 [[FREE Full text](#)] [doi: [10.2196/18941](https://doi.org/10.2196/18941)] [Medline: [32250957](https://pubmed.ncbi.nlm.nih.gov/32250957/)]
60. Salathé M. Digital epidemiology: what is it, and where is it going? *Life Sci Soc Policy* 2018 Jan 04;14(1):1 [[FREE Full text](#)] [doi: [10.1186/s40504-017-0065-7](https://doi.org/10.1186/s40504-017-0065-7)] [Medline: [29302758](https://pubmed.ncbi.nlm.nih.gov/29302758/)]
61. He Z, Zhang CJP, Huang J, Zhai J, Zhou S, Chiu JW, et al. A New Era of Epidemiology: Digital Epidemiology for Investigating the COVID-19 Outbreak in China. *J Med Internet Res* 2020 Sep 17;22(9):e21685 [[FREE Full text](#)] [doi: [10.2196/21685](https://doi.org/10.2196/21685)] [Medline: [32805703](https://pubmed.ncbi.nlm.nih.gov/32805703/)]
62. Tarkoma S, Alghnam S, Howell MD. Fighting pandemics with digital epidemiology. *EClinicalMedicine* 2020 Sep;26:100512 [[FREE Full text](#)] [doi: [10.1016/j.eclinm.2020.100512](https://doi.org/10.1016/j.eclinm.2020.100512)] [Medline: [32864592](https://pubmed.ncbi.nlm.nih.gov/32864592/)]

Abbreviations

HIT: human intelligence task

MTurk: Amazon Mechanical Turk

OR: odds ratio

Edited by A Mavragani; submitted 23.02.22; peer-reviewed by H Matsui, T Karen; comments to author 04.07.22; revised version received 21.09.22; accepted 02.11.22; published 06.12.22

Please cite as:

Flaks-Manov N, Bai J, Zhang C, Malpani A, Ray SC, Taylor CO

Assessing Associations Between COVID-19 Symptomology and Adverse Outcomes After Piloting Crowdsourced Data Collection: Cross-sectional Survey Study

JMIR Form Res 2022;6(12):e37507

URL: <https://formative.jmir.org/2022/12/e37507>

doi: [10.2196/37507](https://doi.org/10.2196/37507)

PMID: [36343205](https://pubmed.ncbi.nlm.nih.gov/36343205/)

©Natalie Flaks-Manov, Jiawei Bai, Cindy Zhang, Anand Malpani, Stuart C Ray, Casey Overby Taylor. Originally published in JMIR Formative Research (<https://formative.jmir.org>), 06.12.2022. This is an open-access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Formative Research, is properly cited. The complete bibliographic information, a link to the original publication on <https://formative.jmir.org>, as well as this copyright and license information must be included.