


# BMJ Open Elucidating symptoms of COVID-19 illness in the Arizona CoVHORT: a longitudinal cohort study

Sana M Khan ,<sup>1</sup> Leslie V Farland,<sup>1,2</sup> Collin J Catalfamo,<sup>1</sup> Erika Austhof,<sup>1</sup> Melanie L Bell,<sup>1,2</sup> Zhao Chen,<sup>1</sup> Felina Cordova-Marks,<sup>2,3</sup> Kacey C Ernst,<sup>1</sup> Pamela Garcia-Filion,<sup>4</sup> Kelly M Heslin,<sup>1</sup> Joshua Hoskinson,<sup>2</sup> Megan L Jehn,<sup>5</sup> Emily C S Joseph,<sup>2</sup> Connor P Kelley,<sup>1</sup> Yann Klimentidis,<sup>1</sup> Stephanie Russo Carroll,<sup>6,7</sup> Lindsay N Kohler,<sup>1,2,3</sup> Kristen Pogreba-Brown,<sup>1</sup> Elizabeth T Jacobs<sup>1,2</sup>

**To cite:** Khan SM, Farland LV, Catalfamo CJ, *et al.* Elucidating symptoms of COVID-19 illness in the Arizona CoVHORT: a longitudinal cohort study. *BMJ Open* 2022;**12**:e053403. doi:10.1136/bmjopen-2021-053403

► Prepublication history and additional supplemental material for this paper are available online. To view these files, please visit the journal online (<http://dx.doi.org/10.1136/bmjopen-2021-053403>).

Received 17 May 2021

Accepted 20 December 2021



© Author(s) (or their employer(s)) 2022. Re-use permitted under CC BY-NC. No commercial re-use. See rights and permissions. Published by BMJ.

For numbered affiliations see end of article.

## Correspondence to

Sana M Khan;  
sanakhan@email.arizona.edu

## ABSTRACT

**Objective** To elucidate the symptoms of laboratory-confirmed COVID-19 cases as compared with laboratory-confirmed negative individuals and to the untested general population among all participants who reported symptoms within a large prospective cohort study.

**Setting and design** This work was conducted within the framework of the Arizona CoVHORT, a longitudinal prospective cohort study conducted among Arizona residents.

**Participants** Eligible participants were any individual living in Arizona and were recruited from across Arizona via COVID-19 case investigations, participation in testing studies and a postcard mailing effort.

**Primary and secondary outcome measures** The primary outcome measure was a comparison of the type and frequency of symptoms between COVID-19-positive cases, tested but negative individuals and the general untested population who reported experiencing symptoms consistent with COVID-19.

**Results** Of the 1335 laboratory-confirmed COVID-19 cases, 180 (13.5%) reported having no symptoms. Of those that did report symptoms, the most commonly reported were fatigue (82.2%), headache (74.6%), aches, pains or sore muscles (66.3%), loss of taste or smell (62.8) and cough (61.9%). In adjusted logistic regression models, COVID-19-positive participants were more likely than negative participants to experience loss of taste and smell (OR 12.1; 95% CI 9.6 to 15.2), bone or nerve pain (OR 3.0; 95% CI 2.2 to 4.1), headache (OR 2.6; 95% CI 2.2 to 3.2), nausea (OR 2.4; 95% CI 1.9 to 3.1) or diarrhoea (OR 2.1; 95% CI 1.7 to 2.6). Fatigue (82.9) and headache (74.9) had the highest sensitivities among symptoms, while loss of taste or smell (87.2) and bone or nerve pain (92.9) had the high specificities among significant symptoms associated with COVID-19.

**Conclusion** When comparing confirmed COVID-19 cases with either confirmed negative or untested participants, the pattern of symptoms that discriminates SARS-CoV-2 infection from those arising from other potential circulating pathogens may differ from general reports of symptoms among cases alone.

## Strengths and limitations of this study

- To our knowledge, no prior research has compared the prevalence of non-specific symptoms such as headache, fever and runny nose between confirmed COVID-19-positive cases, confirmed COVID-19-negative cases and a general, untested comparison group.
- While we have a variety of recruitment methods, the majority of our population was recruited from COVID-19 case investigations, testing and vaccination centres; this may lead to a participant population with greater access to health services than the general population.
- We cannot know the COVID-19 status of the untested participants; it is possible that some had already been infected but were asymptomatic or exhibited few symptoms and were not captured using our study design.

## INTRODUCTION

In late 2019, the novel coronavirus SARS-CoV-2 was first recognised in China among patients who presented with pneumonia and the first scientific report appeared shortly thereafter.<sup>1</sup> On 11 March 2020, the WHO declared COVID-19 a pandemic. The pathogen has had multiple impacts on individual and societal well-being arising from both biological effects of the virus and policy-based mitigation. The majority of those infected with acute COVID-19 will go on to recover, though approximately 10%–20% of patients with COVID-19 overall will develop a severe case of disease, and may suffer from stroke, pneumonia or acute respiratory distress syndrome and require intensive care and ventilation.<sup>2,3</sup>

Individuals are likely to be most infectious during the early phases of the disease, when symptoms may be comparatively mild;



therefore, it is important to elucidate the reported symptom patterns of patients with COVID-19 compared with both laboratory-confirmed negative individuals and population-based controls. Several risk factors have been associated with disease susceptibility and severity including increasing age,<sup>4</sup> male sex<sup>2 5 6</sup> and current or former smoking,<sup>3</sup> which may also affect symptomology. Further, important differences in disease incidence and severity by race and ethnicity have emerged, with Native Americans, African Americans and Latinos having higher COVID-19 prevalence, hospitalisation and mortality rates compared with non-Hispanic whites.<sup>7</sup> It is presently not well known if reports of symptoms or symptom patterns vary by these factors as well.

A recent meta-analysis of over 24 000 patients across nine countries reported on COVID-19 symptom presentation. In this work, the most commonly reported symptoms among people with COVID-19 were fever (78% of patients with COVID-19 reporting), cough (57%) and fatigue (31%).<sup>8</sup> Additionally, a systematic review published in February 2021 aimed to assess the diagnostic accuracy of symptoms associated with COVID-19; this review identified 44 studies which in total included over 26 000 participants. The review found that among 84 symptoms, cough and fever had high sensitivities and could be used as a prompt for further COVID-19 testing. However, a limitation of the review article is potential selection bias due to their sample being selected from primarily clinical settings.<sup>9</sup> Additional work examining symptoms in an unselected population is necessary to determine the syndromic presentation of COVID-19 in the general population. Another study conducted among European patients (n=1420) with mild or moderate COVID-19 found that the most frequently reported symptoms were headache (70%), loss of smell (70%) and obstruction of the nasal passages (68%).<sup>10</sup> The authors of a separate study, the objective of which was to develop a better symptom modelling algorithm to aid targeted testing, concluded that fever and cough should be used as the key symptoms for rapid COVID-19 screening given their high sensitivity.<sup>11</sup> However, a major limitation of studies conducted to date is the lack of comparison of patient-reported symptoms to those of uninfected individuals. To our knowledge, no prior research has compared the prevalence of non-specific symptoms such as headache, fever and runny nose between confirmed COVID-19-positive cases, confirmed COVID-19-negative cases and population-based comparison groups.

Since COVID-19 community transmission began, Arizona has experienced multiple, severe COVID-19 surges, with more than 1.1 million infections and 21 000 COVID-19-related deaths as of October 2021. To address this epidemiological challenge, in May 2020, we initiated a large prospective cohort in Arizona of racially and ethnically diverse residents in order to rigorously investigate the factors contributing to variability in natural COVID-19 history including incidence, progression, resolution and chronic outcomes of infection.<sup>12</sup>

This COVID-19 cohort, dubbed the Arizona CoVHORT, provides a rich data source for multiple areas of inquiry related to the pandemic. The objective of the present work was to determine which symptoms were reported with the greatest frequency among participants who tested positive for COVID-19 as compared with participants who tested negative for COVID-19 and untested participants, while controlling for potential confounders such as age, ethnicity, sex, body mass index (BMI) and smoking status. The findings of this paper will aid in the identification of symptoms that differentiate COVID-19 from other circulating infections or conditions, such as allergies.

## MATERIALS AND METHODS

### Study participants

The overall goal of the CoVHORT is to continuously enrol Arizonans into a cohort study to track both the acute and long-term phases of infection with SARS-CoV-2. The present analysis includes data through 1 October 2021. Several recruitment methods were employed, which have been described in detail previously.<sup>12</sup> Briefly, the primary sources of recruitment have been through case investigations in a partnership with the Arizona Department of Health Services and other research studies and testing sites at The University of Arizona and Arizona State University, both of which have allowed for inclusion of laboratory-confirmed COVID-19-positive and negative participants. By 1 October 2021, a total of 493 COVID-19-positive participants had been recruited through health department case investigations and 901 through our partnerships with studies and testing sites in Arizona.

A comprehensive mailing list was purchased that provides information on 2.2 million residents in Arizona. To recruit the population-based comparison group, a total of 17 500 postcards were mailed to a simple random sample of Pima County, Arizona residents in July 2020. Consistent with the Dillman method to maximise participation and minimise bias,<sup>13</sup> three phased mailings of recruitment postcards occurred every 2 weeks. Participant-provided information from returned surveys was used to exclude those who had already enrolled from subsequent phases of the mailing campaign. Each list was screened prior to each mailing to reduce the number of undeliverable postcards. We have completed all three phases of the mailing campaign in Pima County, with 17 294 postcards delivered in the first phase, 17 147 in the second phase and another 17 081 in the third phase. Method of recruitment is recorded for all participants allowing sensitivity analyses to be conducted within subgroups.

### Patient and public involvement

We encourage active participation from members of the Arizona CoVHORT. The public and members of the cohort are invited to webinars where they are able to provide input, ask questions and speak with the projects' principal investigators. We regularly revisit our survey instruments to ensure they are reflecting feedback from

participants and are centring their experiences and priorities. Study findings are disseminated at our study website (covhort.arizona.edu), along with a regularly updated participant dashboard containing descriptive data of the cohort population.

### Survey instruments

All participants included in the CoVHORT were sent identical structured electronic questionnaires at study entry, regardless of COVID-19 status. All participants were first asked, 'Since January 1, have you experienced a sudden illness that led you to believe you had COVID-19?' If they answered 'yes', all participants were asked to indicate which symptoms they had experienced since January 2020 from a list based on prior reports in the literature, as well as through an open-text field. Participants who responded 'no' were not asked about symptomology and were not included in this analysis. Regardless of symptom status, all participants were then asked if they had tested

for the virus that causes COVID-19 with a nasal swab, throat swab or saliva test since January 2020. Participants were classified as untested, positive or negative based on their results (table 1). Information regarding health and medical history was collected, along with other demographic data, including age, sex, race and ethnicity, as well as for weight, height and smoking status. From these data, we calculated the BMI ( $\text{kg}/\text{m}^2$ ) and categorised the participants as having a BMI of  $<25$ ,  $>25$ – $29.9$  and  $\geq 30$ , to aid in clinical interpretation, as well as reported BMI as a continuous variable (table 2).

### Statistical analysis

Data were analysed to describe the COVID-19 symptoms, estimate the prevalence of individual symptoms and identify differences among COVID-19-positive cases compared with COVID-19-negative individuals and untested participants. Individual variables were summarised and reported using appropriate statistical measures: mean (SD) for

**Table 1** Demographic characteristics of CoVHORT participants who reported symptoms and were laboratory confirmed positive for COVID-19, those who were tested and were negative for COVID-19 and those without COVID-19 test results in the CoVHORT population

Characteristics at study entry	Untested participants*† n=288	Laboratory-confirmed COVID-19 status	
		COVID-19 negative‡§ n=930	COVID-19 positive¶** n=1335
Age (years, mean (SD))	46.8 (14.2)	44.8 (14.0)	43.9 (16.1)
Gender (%)			
Male	89 (31.0)	234 (25.2)	342 (29.6)
Female	193 (67.3)	688 (74.0)	806 (69.8)
Non-binary	5 (1.7)	5 (0.5)	5 (0.4)
Ethnicity (n, %)			
Hispanic	44 (15.3)	130 (14.0)	245 (21.2)
Non-Hispanic	228 (79.4)	788 (85.1)	886 (76.7)
BMI ( $\text{kg}/\text{m}^2$ , mean (SD))	28.0 (6.6)	27.9 (6.9)	28.5 (6.9)
BMI ( $\text{kg}/\text{m}^2$ )			
$<18.5$	2 (0.7)	30 (3.2)	32 (2.8)
$18.5$ – $24.9$	105 (36.6)	348 (37.8)	375 (32.5)
$25.0$ – $29.9$	97 (33.8)	271 (29.5)	344 (29.8)
$30.0$ – $39.9$	59 (20.6)	216 (23.5)	313 (27.1)
$\geq 40$	19 (6.6)	55 (6.0)	83 (7.2)
Smoking status (n, %)			
Never	260 (90.6)	869 (93.9)	1079 (93.4)
Occasionally	11 (3.8)	28 (3.0)	41 (3.6)
Regularly	11 (3.8)	29 (3.1)	25 (2.2)

\*All participants in CoVHORT who did not have a COVID-19 test result.

†Ethnicity: prefer not to answer (n=2), missing (n=13). Smoking status: missing (n=5). BMI: missing (n=5).

‡PCR negative.

§Gender: non-binary gender includes any reported gender other than male or female, including transgender. Prefer not to answer (n=1), transgender male (n=2). Ethnicity: prefer not to answer (n=8), missing (n=4). Smoking status: missing (n=4). BMI: missing (n=10).

¶PCR positive.

\*\*Ethnicity: prefer not to answer (n=12), missing (n=12). Smoking status: missing (n=10). BMI: missing (n=8). BMI, body mass index.

**Table 2** Characteristics of COVID-19-positive study participants (n=1355) by reported number of COVID-19 symptoms

Characteristics at study entry	No symptoms* (n=180)	Any symptoms† (n=1155)	1–6 symptoms (n=486)	7–9 symptoms (n=364)	10–18 symptoms (n=305)	P value‡
Age (years, mean (SD))	41.9 (17.5)	43.9 (16.1)	44.5 (16.7)	45.1 (16.1)	41.6 (14.6)	0.03
Days since symptoms began (mean (SD))§	–	86.6 (95.7)	87.0 (100.7)	84.7 (87.8)	88.2 (97.1)	0.95
Days since positive test (mean (SD))¶	68.5 (73.4)	83.3 (91.4)	83.3 (94.9)	82.7 (86.0)	84.3 (92.2)	0.88
Sex (n, %)**						<0.001
Female	109 (60.6)	806 (69.8)	310 (63.8)	249 (68.4)	247 (81.0)	
Male	68 (37.8)	342 (29.6)	174 (35.8)	113 (31.0)	55 (18.0)	
Ethnicity (n, %)						0.05
Non-Hispanic	129 (72.1)	886 (77.5)	380 (79.0)	284 (79.3)	222 (73.0)	
Hispanic	48 (26.8)	245 (21.4)	94 (19.5)	72 (20.1)	79 (26.0)	
BMI (kg/m <sup>2</sup> , mean (SD))	27.6 (6.5)	28.5 (6.9)	27.4 (5.9)	29.0 (7.2)	29.6 (7.8)	<0.001
BMI (kg/m <sup>2</sup> )						
<18.5	7 (3.9)	32 (2.8)	14 (2.9)	11 (3.0)	7 (2.3)	0.95
18.5–24.9	71 (39.4)	375 (32.7)	175 (36.3)	107 (29.6)	93 (30.7)	Ref
25.0–29.9	45 (25.0)	344 (30.0)	159 (33.0)	108 (29.8)	77 (25.4)	0.81
30.0–39.9	46 (25.6)	313 (27.3)	116 (24.1)	103 (28.5)	94 (31.0)	0.01
≥40	8 (4.4)	83 (7.2)	18 (3.8)	33 (9.1)	32 (10.6)	<0.001
Smoking status (n, %)						0.38
Never	172 (95.6)	1079 (94.2)	457 (95.2)	338 (93.1)	284 (94.0)	
Occasionally or regularly	8 (4.5)	66 (5.8)	23 (4.8)	25 (6.9)	18 (6.0)	
Self-rated severity score††	–		3.6 (2.3)	5.6 (2.1)	6.8 (1.9)	<0.001

\*Sex: non-binary (n=1), transgender male (n=1). Missing values or prefer not to answer: days since positive test (n=2), ethnicity (n=3), BMI (n=3).

†Sex: non-binary (n=5), transgender male (n=1), transgender female (n=1). Missing values or prefer not to answer: days since symptom began (n=178), days since positive test (n=3), ethnicity (n=24), BMI (n=8), smoking status (n=10).

‡P values calculated using ordered logistic regression.

§Number of days between start of symptoms and survey completion.

¶Number of days between positive test date and survey completion.

\*\*Non-binary is a term for gender identities that fall outside of the traditional gender binary of male and female, and is how several participants self-identified.

BMI, body mass index.

continuous variables and per cent (%) for categorical variables. Among those who tested positive for COVID-19, we compared the participant characteristics at study entry and number of symptoms (0 symptom, 1–6 symptoms, 7–9 symptoms, 10–16 symptoms) using ordered logistic regression and reported p values to explore the factors associated with increasing severity. A logistic regression model was fit for each symptom to measure the association, as measured by ORs and 95% CIs, with COVID-19-positive status after adjusting for age, sex, ethnicity, BMI and smoking status. Confounders were selected based on background knowledge. Logistic models were performed using participants with complete data (n=1370) for all variables in the model. Additionally, we included sensitivity and specificity estimates for each individual symptom (online supplemental table 1). Statistical significance was

defined as p<0.05, with two-sided tests. Data analyses were conducted using Stata V.16.0 (College Station, Texas).

## RESULTS

As of 1 October 2021, the Arizona CoVHORT study had enrolled a total of 7012 participants, 2373 (33.8%) of whom reported symptoms associated with COVID-19 since January 2021. Of these participants, 1335 (56.3%) had laboratory-confirmed positive COVID-19 result, 930 (39.2%) had a laboratory-confirmed negative COVID-19 result and 288 (12.1%) were untested (table 1). The participants were of female majority (70.4%) and white (89.4%) and had a mean (SD) age of 44.5 (15.3) years. COVID-19-positive participants were younger (43.9 years) than COVID-19-negative participants (44.8 years), and

**Table 3** Symptom characteristics and ORs of CoVHORT participants using a logistic regression model adjusted for case status, age, sex, ethnicity, BMI and smoking status

Reported symptoms at study entry	COVID-19 positive* n=1155 n (%)	Untested participants† n=288 n (%)	COVID-19 negative‡ n=930 n (%)	Positive versus untested OR (95% CI)	Positive versus negative OR (95% CI)
Fatigue	957 (82.9)	236 (82.2)	680 (73.1)	1.1 (0.7 to 1.5)	1.81 (1.5 to 2.3)
Headache	861 (74.6)	167 (58.2)	495 (53.2)	2.1 (1.5 to 2.7)	2.6 (2.2 to 3.2)
Aches and pains or sore muscles	766 (66.3)	178 (62.0)	506 (54.4)	1.2 (0.9 to 1.6)	1.7 (1.4 to 2.0)
Loss of smell/taste	725 (62.8)	67 (23.4)	119 (12.8)	6.0 (4.3 to 8.3)	12.4 (9.8 to 15.7)
Cough	716 (61.9)	209 (72.8)	540 (58.1)	0.7 (0.5 to 0.9)	1.2 (1.0 to 1.4)
Fever	610 (52.8)	171 (59.6)	452 (48.6)	0.8 (0.6 to 1.0)	1.2 (1.0 to 1.5)
Runny nose/cold-like symptoms	684 (59.2)	139 (48.4)	451 (48.5)	1.5 (1.1 to 1.9)	1.6 (1.3 to 1.9)
Chills	563 (48.7)	132 (46.0)	339 (36.5)	1.2 (0.9 to 1.5)	1.7 (1.4 to 2.0)
Sore throat	543 (47.0)	161 (56.1)	507 (54.5)	0.6 (0.5 to 0.8)	0.7 (0.6 to 0.9)
Difficulty breathing or shortness of breath	475 (41.1)	132 (46.0)	319 (34.3)	0.8 (0.6 to 1.1)	1.4 (1.1 to 1.6)
Diarrhoea	348 (30.1)	71 (24.7)	162 (17.4)	1.3 (1.0 to 1.8)	2.0 (1.6 to 2.5)
Nausea	326 (28.2)	51 (17.8)	129 (13.9)	1.9 (1.3 to 2.7)	2.5 (1.9 to 3.1)
Chest pain or pressure	362 (31.3)	88 (30.7)	246 (26.5)	1.1 (0.8 to 1.4)	1.3 (1.1 to 1.6)
Bone pain/nerve pain	212 (18.4)	20 (7.0)	66 (7.1)	2.9 (1.8 to 4.8)	3.0 (2.2 to 4.0)
Vomiting	92 (8.0)	13 (4.5)	44 (4.7)	1.8 (1.0 to 3.5)	1.7 (1.1 to 2.4)
Other	101 (8.7)	17 (5.9)	32 (3.4)	1.8 (1.0 to 3.1)	2.9 (1.9 to 4.3)
Rash on skin	82 (7.1)	15 (5.2)	38 (4.1)	1.5 (0.8 to 2.8)	1.8 (1.2 to 2.8)
Discolouration of fingers/toes	29 (2.5)	4 (1.4)	18 (1.9)	2.2 (0.6 to 7.2)	1.3 (0.7 to 2.3)
Loss of speech or movement	12 (1.0)	1 (0.4)	7 (0.8)	2.8 (0.4 to 22.0)	1.3 (0.5 to 3.4)
Conjunctivitis	26 (2.3)	11 (3.8)	28 (3.0)	0.6 (0.3 to 1.3)	0.8 (0.5 to 1.4)

\*PCR-positive cases.

†Participants in CoVHORT who do not have a laboratory-confirmed result.

‡PCR or antibody negative.

BMI, body mass index.

participants who had not been tested for COVID-19 (46.8 years). COVID-19-positive participants were more likely to be Hispanic (22.2%), compared with COVID-19-negative participants (14.0%) and untested CoVHORT participants (16.0%). COVID-19-positive participants were more likely to have a BMI of greater than 30 kg/m<sup>2</sup> (34.3%) compared with COVID-19-negative participants (29.5%) and untested CoVHORT participants (27.2%). Of the 1335 laboratory-confirmed COVID-19-positive participants, the majority (86.5%) reported having experienced at least one symptom after diagnosis, while the remaining 180 participants (13.5%) were asymptomatic, having reported never experiencing any symptoms (table 2). When asked to self-rate the severity of their illness on a scale of 0–10, those who reported 10–18 symptoms reported a mean (SD) severity score of 6.8 (1.9), while participants with 7–9 symptoms reported a mean severity score of 5.6 (2.1), and participants with 1–6 symptoms reported a mean severity score of 3.6 (2.3) (table 2). We assessed days since symptom onset and days since test date with the survey completion date and found no significant difference between symptom groups (table 2).

As shown in table 3, other common symptoms that laboratory-confirmed COVID-19-positive participants reported at any time in their disease course included fatigue (82.9%), headache (74.6%), loss of taste or smell (62.8%), aches and pains or sore muscles (66.3%) and cough (61.9%). COVID-19-positive participants had greater odds of reporting loss of taste or smell, bone or nerve pain, headache, nausea and cold-like symptoms when compared with participants who tested negative for COVID-19 and participants who were never tested for COVID-19. While the magnitude of effect for these latter symptoms was smaller, all results were statistically significant. No differences between groups were observed for cough, fever, sore throat, loss of speech or movement, discolouration of fingers or toes and conjunctivitis. After adjusting for age, ethnicity, sex, BMI and smoking status, COVID-19-positive participants were more likely than negative participants to experience loss of taste and smell (OR 12.1; 95% CI 9.6 to 15.2), bone or nerve pain (OR 3.0; 95% CI 2.2 to 4.1), headache (OR 2.6; 95% CI 2.2 to 3.2), nausea (OR 2.4; 95% CI 1.9 to 3.1) and diarrhoea (OR 2.1; 95% CI 1.7 to 2.6) (table 3). Similarly, the

symptoms with the strongest association when comparing COVID-19-positive cases with the untested participants were loss of taste or smell (OR 5.8; 95% CI 4.2 to 7.9), bone/nerve pain (OR 2.9; 95% CI 1.8 to 4.6), headache (OR 2.1; 95% CI 1.6 to 2.7), nausea (OR 1.7; 95% CI 1.2 to 2.5) and cold-like symptoms (OR 1.5; 95% CI 1.1 to 2.0). Fatigue (82.9), headache (74.6) and aches and pains or sore muscles (66.3) were shown to have the highest sensitivities among symptoms, while loss of taste or smell (87.2) and bone or nerve pain (92.9) had high specificity among the significant symptoms (online supplemental table 1).

## DISCUSSION

We assessed the type and frequency of symptoms between COVID-19-positive cases, tested but negative individuals and the general untested population who reported experiencing symptoms consistent with COVID-19. We determined that laboratory-confirmed COVID-19 cases differed in age, ethnicity, BMI from COVID-19-negative participants and untested cohort members. These same factors were associated with reported symptom severity. The most commonly reported first symptoms among COVID-19-positive participants were sore throat, followed by headache, cough, runny nose/cold-like symptoms and fatigue. Discriminating symptoms for COVID-19 positivity included loss of taste and smell and bone or nerve pain as demonstrated by specificity analyses, while fatigue, headache, and aches and pains or sore muscles were shown to have the highest sensitivities among symptoms.

Individuals identifying as Hispanic in CoVHORT constituted 33.5% of the recruited COVID-19-positive participants, mirroring the broader state-wide case composition reported by the Arizona Department of Health Services.<sup>14</sup> By comparison, they constituted far fewer of the laboratory-negative and untested groups. As discussed by Macias Gil *et al*,<sup>15</sup> the burden of COVID-19 on communities of colour has been far more extreme due to extant healthcare disparities, with greater rates of hospitalisations and deaths among US Hispanics as compared with whites being reported in other studies.<sup>15</sup> Further, because publicly available COVID-19 data by race or ethnicity may have missing values, it is critical to continue to follow-up the health outcomes of this medically vulnerable group.

Differences in disease outcomes by body size have been well documented. In the first large study of patients with COVID-19 in the USA, obesity was determined to be a major risk factor for hospitalisation,<sup>3</sup> but it remains unclear whether this finding is attributable to comorbidities that are themselves associated with both larger body size and with severe COVID-19. In the present work, only those with a BMI greater than 30 kg/m<sup>2</sup> were at increased risk for being COVID-19 positive compared with those classified as normal weight or overweight. Disentangling the drivers of susceptibility and disease progression will require long-term follow-up in a large, diverse study population, particularly as several comorbidities, such as type

2 diabetes, are also strongly associated with larger body size. Future work from this cohort will include detailed investigations of the impact of body size on susceptibility to and recovery from COVID-19.

Another equivocal risk factor is smoking, which to date has not been clearly demonstrated to convey an increased risk for severe disease.<sup>3</sup> In the present work, there was no difference in COVID-19 test status by smoking status. A previous study in the USA indicated that current or former smokers were less likely to be hospitalised with COVID-19, but that former smokers were more likely to go on to develop severe disease after hospitalisation, and no differences in frequency of critical illness were observed for current smokers.<sup>3</sup> However, smoking is known to upregulate the production of the ACE2 receptor cells needed for SARS-CoV-2 to invade cells, though nicotine is known to block the ACE2 receptors.<sup>16</sup> This paradox complicates the relationship between smoking and COVID-19, and there is significant variability in the literature. Therefore, more work is needed to assess the role of smoking in COVID-19 progression, and future work from CoVHORT will include a detailed analysis of different smoking modalities such as vaping or e-cigarettes, cigar and cigarette smoking.

Several efforts have been made to identify and characterise the symptoms associated with COVID-19 to allow for more efficient and targeted screening practices, as well as to differentiate SARS-CoV-2 infection from other diseases, such as influenza.<sup>8-10 17</sup> However, these reports of COVID-19 symptoms have largely been confined to hospitalised or outpatient population and are lacking a symptomatic COVID-19-negative comparison group. Because many of the symptoms reported as being associated with COVID-19 are general symptoms that could be associated with conditions such as allergies or other infectious illnesses such as influenza, there is an urgent need to evaluate the prevalence of reported symptoms of confirmed COVID-19-positive cases as compared with confirmed COVID-19-negative individuals, as well as with the prevalence of symptoms in the general population.

The results of the present study demonstrate that in southern Arizona, the most common first symptom reported by COVID-19-positive participants was sore throat; other common first symptoms of COVID-19 included headache, cough, runny nose or cold-like symptoms and fatigue. While these are the same cluster of symptoms as reported by Larsen *et al* in a large meta-analysis of more than 50 000 subjects, with data captured by the WHO, the timing of appearance differed.<sup>11</sup> Specifically, the report by Larsen *et al* concluded that the order of symptom appearance was estimated to be fever, cough, nausea and vomiting, while in the current work, the first symptom reported by the majority of cases was sore throat, followed by headache, cough and runny nose; only 6% of participants had fever as their first symptom. Differences in the study population, including geographic location, sex, age, timing within the pandemic, severity of illness that prompted healthcare-seeking behaviour and testing, testing accessibility and race differences across the

spectrum of studies employed in the meta-analysis, may explain some of the inconsistent results for first reported symptoms.

An example of this variation in symptom reporting can be observed regarding the number of symptoms that women experienced as compared with men. Women were more likely to be classified in the category of the greatest number of symptoms than men, as were those with a BMI of greater than 30 kg/m<sup>2</sup>, compared with those with a BMI below that threshold, although these findings were not statistically significant. A greater proportion of smokers was observed in the asymptomatic category, as compared with any symptom category. These findings suggest that ascertaining the type and order of COVID-19-specific symptomology may be confounded by characteristics of the participants.

With regard to overall COVID-19 symptoms, the greatest differences between laboratory-confirmed positive and negative participants were observed for loss of smell and taste and bone or nerve pain, followed by vomiting, nausea and headache. A similar pattern was seen when comparing cases to the overall untested sample. To date, most work regarding symptoms has relied on the frequency of symptom occurrence among cases, with little ability to ascertain the degree to which these symptoms differentiate cases from non-cases. For instance, the largest meta-analysis of COVID-19 symptomology to date included data from 24410 cases from nine countries and reported that the most common symptoms were fever (78%), cough (57%) and fatigue (31%).<sup>8</sup> A smaller study within the USA found that the frequency of symptoms among cases was highest for cough (84%), fever (80%), aches and pains (63%), chills (63%) and fatigue (62%).<sup>17</sup> In comparison, herein we found that the most common symptoms reported by cases were fatigue, headache, loss of smell or taste, cough, aches or pains, or sore muscles.

A key finding of this work is that the discrimination of COVID-19-positive symptoms from others requires comparison groups. General symptoms reported differ from those which may be applied to differentiate COVID-19 from other infectious diseases or conditions that are present in the underlying population. The symptoms that demonstrated the greatest difference between COVID-19-positive participants and the prevalence of symptoms among laboratory-confirmed COVID-19-negative participants or in the general population were loss of smell and taste, bone or nerve pain, headache, nausea and fatigue.

The strengths of this study are its prospective nature, ability to capture data for laboratory-confirmed COVID-19-positive cases who have not been hospitalised and the presence of comparison groups among both those who tested negative for COVID-19 as well as a population base drawn from throughout Arizona. These aspects allowed us to compare symptoms between cases and laboratory-confirmed uninfected individuals. However, limitations of the work must also be considered. First and foremost, while we are able to recruit participants via follow-up to

COVID-19 testing, participants' test results and symptoms are self-reported. Furthermore, although we have self-reported, laboratory-confirmed negative participants, we cannot know the COVID-19 status of the untested participants. It is possible that some had already been infected but were asymptomatic or exhibited few symptoms. This would likely attenuate any associations between exposure and outcomes in this study. It is also important to acknowledge that participants who indicated not experiencing symptoms that led them to believe they had COVID-19, regardless of testing status, were not asked to indicate which symptoms they had experienced. These participants were not able to directly indicate that they experienced no symptoms from the provided list; however, because the majority of CoVHORT participants who undergo testing for COVID-19 enrol at a time point after receiving their test results, we believe that the likelihood that participants who indicate not experiencing symptoms actually experienced symptoms from our list offered to participants who indicate experiencing symptoms is low. Additionally, there may be differences in the source population for cases as compared with the laboratory-negative participants and untested participants due to the differences in recruitment strategies for these populations. For example, while postcards were mailed to a random selection of households, it is possible that Latinx participants were less likely to respond to this method than direct recruitment as cases during routine case follow-up. This could bias the association between being COVID-19 positive and Latinx away from the null. However, our race/ethnicity profile among cases is approximately similar to the overall distribution of cases throughout Arizona, suggesting a representative sample. Therefore, bias would potentially come from differential responses to other recruitment methods. This was an exploratory study, with a large number of statistical tests, and therefore care should be taken when considering *p* values.

In conclusion, the findings of this analysis from the Arizona CoVHORT study show variation in several individual characteristics between COVID-19-positive participants, negative participants and the untested population, which will be studied in future publications to assess the contributors to these observations. In addition, we found that in southern Arizona, COVID-19-positive participants most commonly reported a sore throat, headache, fatigue, cough or runny nose as the first symptom they noted. These results may aid in earlier identification of cases in the future and highlight the continued importance of addressing surveillance strategies as the pandemic continues.

#### Author affiliations

<sup>1</sup>Department of Epidemiology and Biostatistics, The University of Arizona, Tucson, Arizona, USA

<sup>2</sup>The University of Arizona Cancer Center, Tucson, Arizona, USA

<sup>3</sup>Department of Health Promotion Sciences, The University of Arizona, Tucson, Arizona, USA



<sup>4</sup>Department of Biomedical Informatics, The University of Arizona, Tucson, Arizona, USA

<sup>5</sup>School of Human Evolution and Social Change, Arizona State University, Tempe, Arizona, USA

<sup>6</sup>Department of Community, Environment, and Policy, The University of Arizona, Tucson, Arizona, USA

<sup>7</sup>Native Nations Institute at the Udall Center for Studies in Public Policy, The University of Arizona, Tucson, Arizona, USA

**Contributors** KP-B, LVF, MLJ, MLB, ZC, YK, KCE, ETJ, PG-F, SRC, LNK and FC-M conceptualised the study and developed the initial study protocol. SMK, KMH, CJC, EA, JH, EJ, CPK and KP-B participated in the design of the protocol and the drafting and reviewing of the manuscript. ETJ is responsible for the contents of this manuscript as the guarantor. All authors critically reviewed the draft of the manuscript and approved the final version.

**Funding** This work was supported by the BIO5 Institute at The University of Arizona.

**Competing interests** None declared.

**Patient consent for publication** Obtained.

**Ethics approval** This study involving human participants was reviewed and approved by the Institutional Review Board of The University of Arizona Human Subjects Protection Program (2003521636A00). Participants gave informed consent to participate in the study before taking part.

**Provenance and peer review** Not commissioned; externally peer reviewed.

**Data availability statement** No data are available. No additional data available.

**Supplemental material** This content has been supplied by the author(s). It has not been vetted by BMJ Publishing Group Limited (BMJ) and may not have been peer-reviewed. Any opinions or recommendations discussed are solely those of the author(s) and are not endorsed by BMJ. BMJ disclaims all liability and responsibility arising from any reliance placed on the content. Where the content includes any translated material, BMJ does not warrant the accuracy and reliability of the translations (including but not limited to local regulations, clinical guidelines, terminology, drug names and drug dosages), and is not responsible for any error and/or omissions arising from translation and adaptation or otherwise.

**Open access** This is an open access article distributed in accordance with the Creative Commons Attribution Non Commercial (CC BY-NC 4.0) license, which permits others to distribute, remix, adapt, build upon this work non-commercially, and license their derivative works on different terms, provided the original work is properly cited, appropriate credit is given, any changes made indicated, and the use is non-commercial. See: <http://creativecommons.org/licenses/by-nc/4.0/>.

#### ORCID iD

Sana M Khan <http://orcid.org/0000-0003-0699-2103>

## REFERENCES

- Zhu N, Zhang D, Wang W, *et al*. A novel coronavirus from patients with pneumonia in China, 2019. *N Engl J Med Overseas Ed* 2020;382:727–33.
- Guan W-J, Ni Z-Y, Hu Y, *et al*. Clinical characteristics of coronavirus disease 2019 in China. *N Engl J Med* 2020;382:1708–20.
- Petrilli CM, Jones SA, Yang J, *et al*. Factors associated with hospital admission and critical illness among 5279 people with coronavirus disease 2019 in New York City: prospective cohort study. *BMJ* 2020;369:m1966.
- Zhou F, Yu T, Du R, *et al*. Clinical course and risk factors for mortality of adult inpatients with COVID-19 in Wuhan, China: a retrospective cohort study. *Lancet* 2020;395:1054–62.
- Chen N, Zhou M, Dong X, *et al*. Epidemiological and clinical characteristics of 99 cases of 2019 novel coronavirus pneumonia in Wuhan, China: a descriptive study. *Lancet* 2020;395:507–13.
- Cai H. Sex difference and smoking predisposition in patients with COVID-19. *Lancet Respir Med* 2020;8:e20.
- Hygiene NDoHaM. *Age-Adjusted rates of lab confirmed COVID-19 non-hospitalized cases, estimated non-fatal hospitalized cases, and persons known to have died per 100,000 by race/ethnicity group*. Albany, NY, 2020.
- Grant MC, Geoghegan L, Arbyn M, *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. *PLoS One* 2020;15:e0234765.
- Struyf T, Deeks JJ, Dinnes J. Signs and symptoms to determine if a patient presenting in primary care or hospital outpatient settings has COVID-19. *Cochrane Database Syst Rev* 2021;2021.
- Lechien JR, Chiesa-Estomba CM, Place S, *et al*. Clinical and epidemiological characteristics of 1420 European patients with mild-to-moderate coronavirus disease 2019. *J Intern Med* 2020;288:335–44.
- Larsen JR, Martin MR, Martin JD, *et al*. Modeling the onset of symptoms of COVID-19. *Front Public Health* 2020;8:473.
- Catalfamo CJ, Heslin KM, Shilen A, *et al*. Design of the Arizona CoVHORT: a population-based COVID-19 cohort. *Front Public Health* 2021;9:620060.
- Dillman D, Smyth J, Christian L. *Internet, mail, and Mixed-Mode surveys: the tailored design method*. Hoboken, NJ: Wiley & Sons, 2009.
- Services ADoH. *COVID-19 data Dashboard*. Phoenix, Arizona: Arizona Department of Health Services, 2021. <https://www.azdhs.gov/preparedness/epidemiology-disease-control/infectious-disease-epidemiology/covid-19/dashboards/index.php>
- Macias Gil R, Marcelin JR, Zuniga-Blanco B, *et al*. COVID-19 pandemic: disparate health impact on the Hispanic/Latinx population in the United States. *J Infect Dis* 2020;222:1592–5.
- Smith JC, Sausville EL, Girish V, *et al*. Cigarette smoke exposure and inflammatory signaling increase the expression of the SARS-CoV-2 receptor ACE2 in the respiratory tract. *Dev Cell* 2020;53:514–29.
- Burke RM, Killerby ME, Newton S, *et al*. Symptom Profiles of a Convenience Sample of Patients with COVID-19 - United States, January–April 2020. *MMWR Morb Mortal Wkly Rep* 2020;69:904–8.