



RESEARCH ARTICLE

Predictors of incident viral symptoms ascertained in the era of COVID-19

Gregory M. Marcus ^{*}, Jeffrey E. Olgin, Noah D. Peyser, Eric Vittinghoff , Vivian Yang, Sean Joyce, Robert Avram , Geoffrey H. Tison, David Wen , Xochitl Butcher, Helena Eitel, Mark J. Pletcher

Division of Cardiology, Department of Medicine, University of California, San Francisco, California, United States of America

* greg.marcus@ucsf.edu



Abstract

Background

In the absence of universal testing, effective therapies, or vaccines, identifying risk factors for viral infection, particularly readily modifiable exposures and behaviors, is required to identify effective strategies against viral infection and transmission.

Methods

We conducted a world-wide mobile application-based prospective cohort study available to English speaking adults with a smartphone. We collected self-reported characteristics, exposures, and behaviors, as well as smartphone-based geolocation data. Our main outcome was incident symptoms of viral infection, defined as fevers and chills plus one other symptom previously shown to occur with SARS-CoV-2 infection, determined by daily surveys.

Findings

Among 14, 335 participants residing in all 50 US states and 93 different countries followed for a median 21 days (IQR 10–26 days), 424 (3%) developed incident viral symptoms. In pooled multivariable logistic regression models, female biological sex (odds ratio [OR] 1.75, 95% CI 1.39–2.20, $p < 0.001$), anemia (OR 1.45, 95% CI 1.16–1.81, $p = 0.001$), hypertension (OR 1.35, 95% CI 1.08–1.68, $p = 0.007$), cigarette smoking in the last 30 days (OR 1.86, 95% CI 1.35–2.55, $p < 0.001$), any viral symptoms among household members 6–12 days prior (OR 2.06, 95% CI 1.67–2.55, $p < 0.001$), and the maximum number of individuals the participant interacted with within 6 feet in the past 6–12 days (OR 1.15, 95% CI 1.06–1.25, $p < 0.001$) were each associated with a higher risk of developing viral symptoms. Conversely, a higher subjective social status (OR 0.87, 95% CI 0.83–0.93, $p < 0.001$), at least weekly exercise (OR 0.57, 95% CI 0.47–0.70, $p < 0.001$), and sanitizing one's phone (OR 0.79, 95% CI 0.63–0.99, $p = 0.037$) were each associated with a lower risk of developing viral symptoms.

OPEN ACCESS

Citation: Marcus GM, Olgin JE, Peyser ND, Vittinghoff E, Yang V, Joyce S, et al. (2021) Predictors of incident viral symptoms ascertained in the era of COVID-19. PLoS ONE 16(6): e0253120. <https://doi.org/10.1371/journal.pone.0253120>

Editor: Antonio Palazón-Bru, Universidad Miguel Hernandez de Elche, SPAIN

Received: September 4, 2020

Accepted: May 31, 2021

Published: June 17, 2021

Copyright: © 2021 Marcus et al. 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 author and source are credited.

Data Availability Statement: All relevant data are within the manuscript and its [Supporting information](#) files.

Funding: Drs. GM, JO, and MP received funding from awards IU2CEB021881-01 and 3U2CEB021881-05S1 from the NIH/ NIBIB (<https://www.nibib.nih.gov/>). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing interests: The authors have declared that no competing interests exist.

Interpretation

While several immutable characteristics were associated with the risk of developing viral symptoms, multiple immediately modifiable exposures and habits that influence risk were also observed, potentially identifying readily accessible strategies to mitigate risk in the COVID-19 era.

Introduction

The global SARS-CoV-2 pandemic has affected communities in every habitable continent and every state in the US. Given what is generally known about respiratory viruses, strategies to mitigate transmission have included government orders to practice regular hand hygiene, physical distancing, including the closures of public locations commonly associated with community gatherings, and more recently to wear masks [1–3]. Studies thus far have largely focused on comparisons among those seeking medical care for the disease [4–7] or evaluations of large administrative datasets [8]. It can be difficult to track individual-level characteristics and behaviors, particularly as they are dynamic and changing over time, as they relate to incident disease. Members of the public may benefit by understanding strategies under their direct control that may influence their own risk of infection and viral transmission.

Tracking viral infection is hindered by the absence of universal and repeated testing. In the absence of such testing, recent evidence suggests that symptoms themselves may be useful markers of SARS-CoV-2 infection [9]. While the virus may be asymptomatic, a variety of symptom clusters associated with the disease have been identified, often including fever, but ranging from typical respiratory symptoms to gastrointestinal afflictions to somewhat idiosyncratic findings such as anosmia/ageusia and conjunctivitis [10–18]. In the past, ascertainment of viral symptoms has relied on assessments of those seeking medical care or retrospective surveys that may be prone to recall bias. Given the current near-ubiquity of smartphones and use of related mobile apps, technology is now available to regularly and repeatedly query large numbers of individuals over time, providing access to symptom development as it arises. Although monitoring for viral symptoms may be neither sufficiently sensitive nor specific for SARS-CoV-2 infection, these outcomes are, by their nature, inherently experienced by the individual, potentially providing valuable information that may be best leveraged by modern mobile technology.

We sought to use prospectively collected information about exposures and modifiable behaviors, along with daily symptom reporting, to identify risk factors for incident viral symptoms using a globally-available, smartphone mobile application-based study, the COVID-19 Citizen Science Study.

Methods

We launched the COVID-19 Citizen Science Study, a mobile application-based study compatible with Android or iOS operating systems, on March 26, 2020. The mobile application was built by investigators and developers at the University of California, San Francisco, using the NIH-supported Eureka digital and mobile research platform. Enrollment is open to any adult with a smartphone, and study information has been broadcast via press release, social media, and to participants in the Eureka-based Health eHeart Study. Participants were encouraged to recruit additional individuals. Updated study information, including number of participants,

maps of symptom clusters, and location of participants around the world can be found at <https://covid19.eurekaplatform.org/>. All participation is remote, without geographic restriction. Verification of cell phone numbers via text was required before proceeding from study registration to remote-based study consent and subsequent study participation. Retention strategies include daily notifications, data visualizations, and intermittent study update blog posts. Our Citizen Scientist participants contribute study question ideas that are then included into the study and reported to participants as participant-generated.

Participants complete surveys written in lay language meeting Flesch-Kincaid criteria for an 8th grade reading level (<https://readabilityformulas.com>). At baseline, surveys collected information about demographics, education, occupation, SARS-CoV-2 status (referred to in surveys as “the novel coronavirus, the virus that causes COVID-19”), behaviors, living conditions, attitudes regarding the SARS-CoV-2 pandemic, local government restrictions related to the disease, medical conditions, and medications (the surveys are included in the [S1 File](#)). Questionnaires included pre-tested survey instruments employed in multiple Health eHeart Study and Eureka Digital Research Platform studies for demographics and past medical conditions; however, as the study was launched during the beginning of the SARS-CoV-2 pandemic, questionnaires specific to related risk factors and viral symptoms were new. Perceived socioeconomic status was assessed using the MacArthur subjective social status ladder [[19,20](#)]. All participants received an optional invitation to share their smartphone-based geolocation data.

Participants receive a daily survey, timed to occur synchronously to their local same time of day when they engaged with the first baseline survey, via mobile application-based push notification. The daily survey includes queries about current viral symptoms, updated according to new information, using “check all that apply” including: “A scratchy throat”; “A cough (worse than usual if you have a baseline cough)”; “A painful sore throat”; “A temperature greater than 100.4°F or 38.0°C”; “A runny nose”; “Symptoms of fever or chills”; “Muscle aches (worse than usual if you have baseline muscle aches)”; “Shortness of breath”; “Nausea, vomiting or diarrhea” (added March 30, 2020); “Unable to taste or smell” (added March 31, 2020); “Red or painful eyes” (added April 13, 2020); or “none of the above.” The daily survey then includes questions regarding current symptoms among household contacts and the number of individuals outside the household the participant interacted with within six feet (about 1.83 meters) in the previous 24 hours.

Participants received weekly surveys to update information regarding sleep, exercise, hand hygiene, social and physical distancing behaviors, habits such alcohol consumption, and SARS-CoV-2 infection status. All surveys remained open for 24 hours.

For those that consented to geolocation tracking, smartphone-based geolocation using a combination of the Global Positioning System and cell phone tower triangulation was collected every 5 minutes for Android phones and whenever the phone accelerometer exhibited movement (in order to minimize battery drain) for iOS smartphones. Geolocation latitude and longitude coordinates were clustered within an individual for every day of the study using the HDBScan clustering algorithm [[21,22](#)]. The most prevalent cluster of geolocation coordinates for a user was defined as “home.” Time spent at a cluster was calculated as the time difference between the current location cluster and any future cluster change. Daily time spent at home was calculated as the time spent at the cluster identified as “home” divided by the time between the first and last coordinate collected daily. In addition, distance travelled was calculated as the sum of the successive distances between all consecutive coordinates collected within a user on a daily basis. Long-distance travel was defined as movement of at least 1,000 kilometers within 24 hours.

Occupation was dichotomized into healthcare workers versus not; sleep was determined as the average number of hours per day over each week; exercise, defined as physical activity for

at least 20 minutes that resulted in breathing heavily or to “break a sweat,” was dichotomized into more or less than once weekly; alcohol was assessed as average daily standard drinks; cigarette, e-cigarette, and marijuana use were dichotomized into any use in the last 30 days versus not; household symptoms were dichotomized into any versus none in the previous 6–12 days; and the maximum number of contacts within six feet (about 1.83 meters) reported in the previous 6–12 days were derived from the daily survey responses. The lag time of 6–12 days was employed to allow for the incubation period of SARS-CoV-2 [23] and other common respiratory viruses [24] and in order to accommodate days without responses and the expectation that viral symptoms would last several days.

For the current analyses, all participants reporting a previous positive test for SARS-CoV-2 and those with any symptoms upon entry to the study were excluded. Those with baseline medical conditions that might themselves contribute to the symptoms of interest, including atrial fibrillation, coronary artery disease, congestive heart failure, chronic obstructive pulmonary disease, and asthma, were excluded. Two sensitivity analyses were conducted: one also excluded all participants reporting anemia; a second included participants with atrial fibrillation, coronary artery disease, congestive heart failure, chronic obstructive pulmonary disease, and asthma. Based on the daily surveys, incident viral symptoms were defined as the first report of a combination of fever or chills plus at least one other symptom on the same day. The absence of completion of a daily survey was assumed to represent an absence of symptoms in statistical analyses. Given the protean manifestations of COVID-19 [5,10–18,25], we allowed for any of these viral symptoms, requiring fever given that each of those symptoms could possibly occur in the absence of a viral infection. Follow-up for the current study ended May 3, 2020. The study was approved by the University of California, San Francisco Institutional Review Board. All participants provided informed electronic consent.

Statistical analyses

Normally distributed continuous variables are presented as means \pm SD and compared using t-tests, where continuous variables with skewed distributions are presented as medians with interquartile ranges (IQR) and compared using Wilcoxon rank sum tests. Categorical variables were compared using chi-squared tests. Pooled logistic regression models were used to identify factors associated with incident symptoms, potentially including baseline characteristics (demographics, medical conditions, habits, and behaviors related to viral infection risk such as hand hygiene), and time-updated information from daily and weekly surveys. Exposures that expected to influence the risk of viral infection that would then manifest as future symptoms several days later were evaluated using survey data for 6–12 days earlier. Consequently, only participants with at least one daily survey at least 6 days after the first were included in the pooled logistic regression models. Beginning with the subset of variables associated with incident symptoms at $p < 0.1$ in pooled logistic regression models adjusting only for age, sex, race, and calendar date (with linear and non-linear components), backward deletion was used to select multivariable models retaining covariates with $p < 0.05$. Statistical analyses were performed using Stata, version 16 (College Station, TX). Two-tailed p -values < 0.05 were considered statistically significant.

Results

After exclusions were applied, 14,335 participants were available and contributed to the incident analyses. Differences between these participants and those that entered the study reporting at least one viral symptom are shown in [Table 1](#). Participants resided in all 50 states and in 93 countries outside the US. While a mean $42\% \pm 12\%$ of all participants completed the daily

Table 1. Baseline characteristics of participants with and without prevalent viral symptoms.

	Prevalent Symptoms N = 374	Included in Incident Analyses N = 14,335	p-value
Age Category			<0.001
18–29	63 (16.8%)	1,961 (13.7%)	
30–39	112 (29.9%)	3,225 (22.5%)	
40–49	86 (23.0%)	2,873 (20.0%)	
50–59	55 (14.7%)	2,839 (19.8%)	
60+	58 (15.5%)	3,437 (24.0%)	
Female Biological Sex	253 (67.6%)	9,322 (65.0%)	<0.001
Race/Ethnicity			0.005
White	290 (77.5%)	11,342 (79.1%)	
Black	7 (1.9%)	164 (1.1%)	
Hispanic (any race)	43 (11.5%)	1,307 (9.1%)	
Asian or Pacific Islander	17 (4.5%)	1,150 (8.0%)	
Other (including multiracial)	17 (4.5%)	372 (2.6%)	
Highest Level of Education			<0.001
Less than high school	4 (1.1%)	57 (0.4%)	
High school graduate	14 (3.7%)	471 (3.3%)	
Some college	80 (21.4%)	2,034 (14.2%)	
College graduate	119 (31.8%)	5,039 (35.2%)	
Post-graduate	149 (39.8%)	6,566 (45.8%)	
Other	8 (2.1%)	167 (1.2%)	
MacArthur Subjective Social Status Ladder	7.0 (5.0–8.0)	7.0 (6.0–8.0)	<0.001
Health Care Worker	73 (19.5%)	3,044 (21.2%)	0.42
Anemia	78 (20.9%)	1,440 (10.0%)	<0.001
Cancer	24 (6.4%)	558 (3.9%)	0.044
Diabetes	26 (7.0%)	701 (4.9%)	0.18
High Blood Pressure	90 (24.1%)	3,128 (21.8%)	0.37
HIV	2 (0.5%)	77 (0.5%)	0.44
Immunodeficiency	23 (6.1%)	322 (2.2%)	<0.001
Children home from college living with you	20 (5.3%)	1,098 (7.7%)	0.096
School-age children living with you	103 (27.5%)	3,575 (24.9%)	0.51
At least weekly exercise	240 (64.3%)	10,627 (74.2%)	<0.001
Median drinks per day (IQR)	0.2 (0.0–0.9)	0.4 (0.0–1.0)	0.002
Median hours of sleep (IQR)	5.0 (5.0–5.0)	7.0 (7.0–8.0)	0.076
Cigarettes: any use in last 30 days	30 (8.1%)	703 (4.9%)	0.003
E-cigarettes: any use in last 30 days	24 (6.4%)	342 (2.4%)	<0.001
Marijuana: any use in last 30 days	42 (11.4%)	1,448 (10.2%)	0.17
Any pets at home	223 (59.6%)	8,512 (59.4%)	0.040
Received flu shot within the year	251 (67.1%)	10,772 (75.1%)	0.002

<https://doi.org/10.1371/journal.pone.0253120.t001>

survey each day, 95%–100% of all participants completed at least one daily survey per week throughout the study period, and weekly surveys were completed $66 \pm 26\%$ of the time (S1 Table).

Over a median follow-up of 21 days (IQR 10–26 days), 424 (3%) participants developed incident viral symptoms. S2 Table shows the specific symptoms reported. Fig 1 illustrates the locations of participants with and without symptoms. Fig 2 provides a sample summary of enrollment, survey completion, symptom development, and follow-up over time.

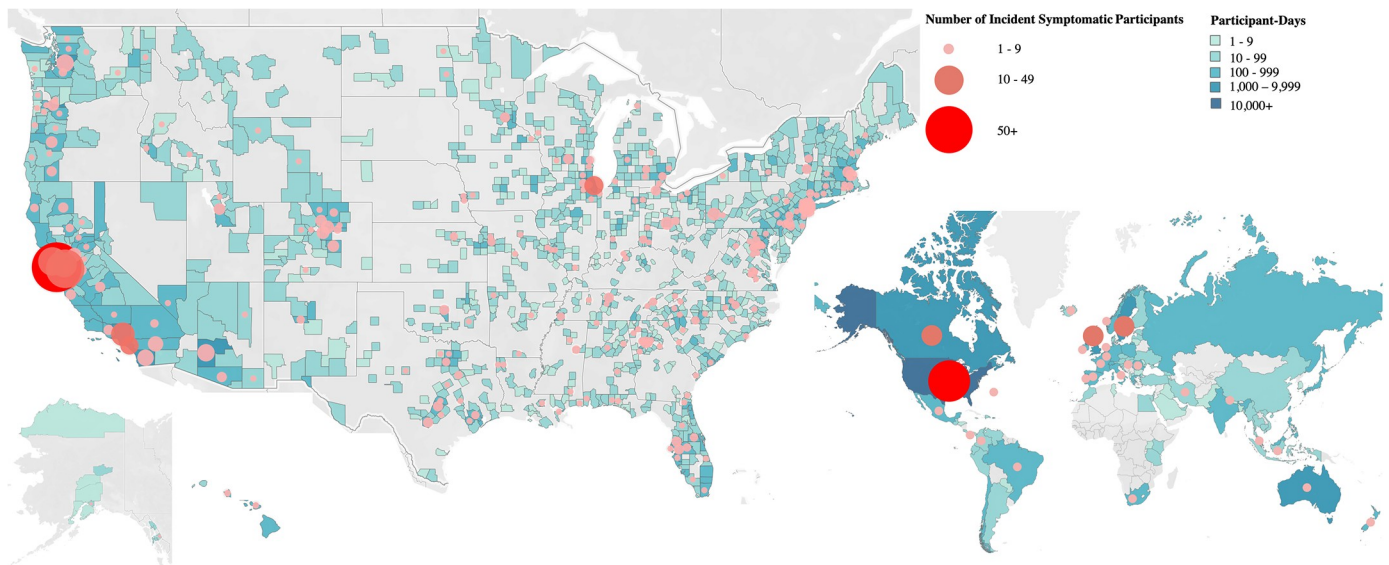


Fig 1. Location of study participants. Blue shading represents gradations of the number of participant-days within the US by county (left) and in the world by nation (right). Red shading depicts the number of symptomatic participants by location. Created with software provided by Tableau (www.tableau.com; San Francisco, CA) and published with their permission under the Creative Commons Attribution License (CC BY 4.0).

<https://doi.org/10.1371/journal.pone.0253120.g001>

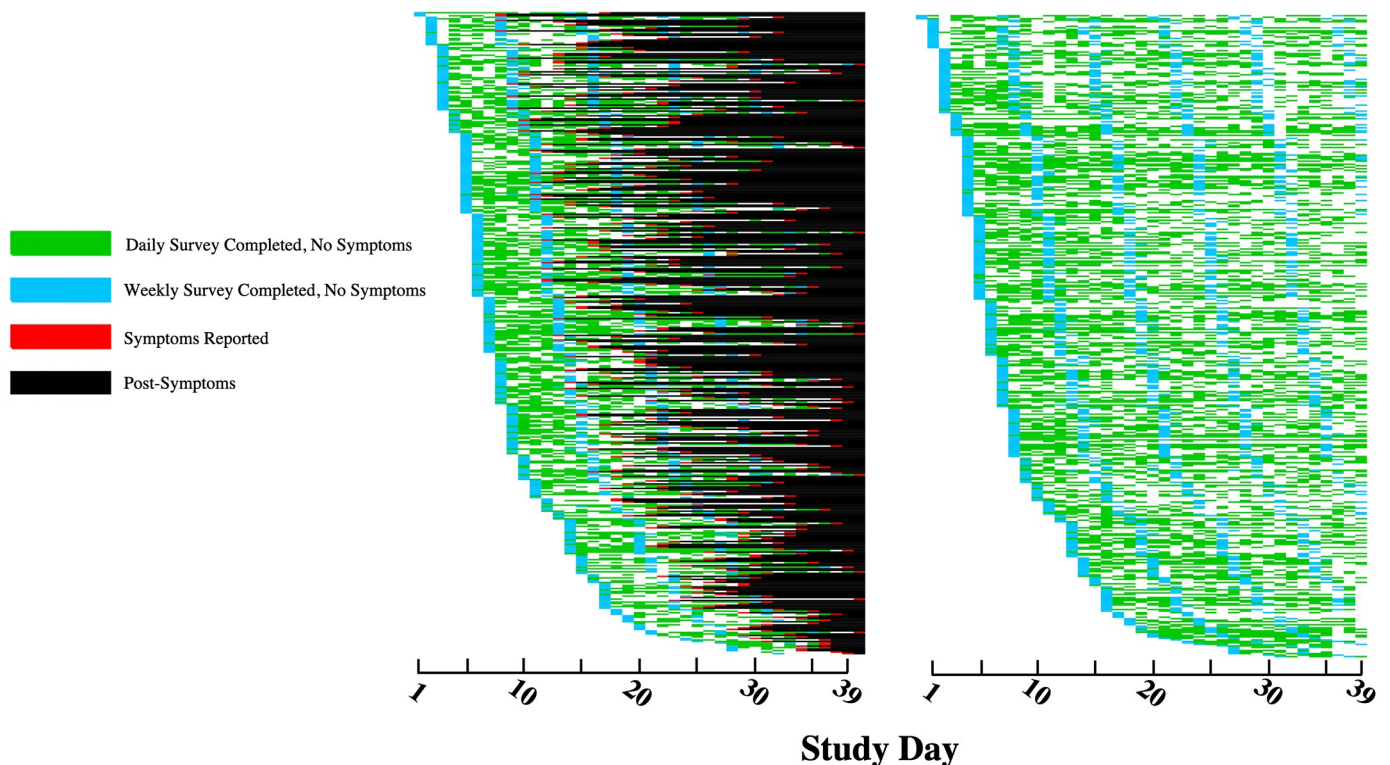


Fig 2. Heat map of symptomatic and sample of asymptomatic patients displaying time of enrollment, survey completion, time of symptom development, and follow-up. The left plot depicts participants that developed symptoms. The right plot depicts participants who did not develop symptoms matched in a one-to-one fashion with each symptomatic case by time of enrollment. Each row represents a unique study participant ($n = 424$ for each plot). The X-axis represents days of the current study. Blue = weekly survey completed (the first blue represents the enrollment visit). Green = daily survey completed (the daily survey contents are included in the weekly survey). Red = symptoms developed. Black = after development of symptoms. White = no data entry prior to or in the absence of symptoms.

<https://doi.org/10.1371/journal.pone.0253120.g002>

In minimally adjusted logistic models adjusting only for age, sex, race, ethnicity, and date, a higher level of education and subjective social scale, exercising at least once weekly, a longer average sleep duration, and sanitizing one's phone were each associated with a lower risk while a history of anemia, hypertension or some immunodeficiency, cigarette smoking, e-cigarette use, marijuana use, having pets at home, having household members with viral symptoms, and the number of individuals with which the participant interacted with within six feet (about 1.83 meters) each predicted a higher risk of incident viral symptoms (Table 2). Pertinent characteristics that failed to exhibit statistically significant relationships included HIV status, hand washing practices, reported government restrictions, and, per geolocation measurements, amount of time at home and daily distance traveled. In the backwards stepwise logistic model, the following were retained: a higher level of subjective social status, exercise, and sanitizing one's phone were each associated with a significantly lower risk of developing viral symptoms, whereas female sex, a history of anemia, hypertension, recent cigarette smoking, recent household contacts with viral symptoms, and the maximum number of individuals recently in contact with the participant within six feet (about 1.83 meters) were each associated with a significantly heightened risk of developing viral symptoms (Table 3).

In sensitivity analyses, excluding all participants with anemia did not meaningfully change the results (S3 Table). After including participants with atrial fibrillation, coronary artery disease, congestive heart failure, chronic obstructive pulmonary disease, and asthma: those with congestive heart failure and anemia exhibited a higher risk for incident symptoms, and all of the previously observed statistically significant relationships remained (S4 Table).

Discussion

Among an international cohort involving collection of prospective and time-updated data, female sex, anemia, hypertension, recent cigarette smoking, living with someone with viral symptoms, and the maximum number of recent contacts within six feet (about 1.83 meters) outside the home each predicted a higher risk of developing viral symptoms. Conversely, a higher self-perceived social status, regular exercise, and sanitizing one's phone were each associated with a lower risk of subsequently reporting viral symptoms.

As of July 25, 2020, there were more than 15 million confirmed cases of SARS-CoV-2 infections and more than 640,000 COVID-19-related deaths around the world [26]. In response, tremendous investment and efforts are being dedicated to enhance the availability of testing [27], identify effective therapies [28], and ultimately to develop a vaccine [29]. Although these remedies are being pursued at an unprecedented pace, the number of infections and deaths continues to grow, and, even after these new technologies, drugs, and vaccines are developed, additional time will be required to disseminate them. While studies of hospitalized patients are valuable, ultimately the characteristics, behaviors, and exposures of individuals in the general community associated with the development of viral symptoms can be helpful in several ways: to identify those at highest risk of developing these symptoms, which may help prioritize protecting the most vulnerable; to provide novel insights regarding the biology of current viral disease; and ideally to identify low-risk and modifiable behaviors that individuals might practice or avoid to reduce their own individual-level risk.

Clearly, self-reported symptoms of a viral infection do not equal SARS-CoV-2 infection. However, the specific viral symptoms repeatedly queried in our population were developed based on available evidence regarding the nature of SARS-CoV-2 infection [10–18], and similar viral diseases, including the common cold and influenza for example, very likely share common properties related to an individual's susceptibility to infection [30]. In addition, independent of commonalities across easily transmissible viral diseases, there are shared

Table 2. Minimally adjusted odds of incident symptoms.

Characteristic	Odds Ratio	95% CI	p-value	Group p-value
Highest level of education				
Less than high school	reference			<0.001*
High school graduate	0.44	0.16, 1.16	0.10	<0.001†
Some college	0.51	0.21, 1.24	0.14	0.015#
College grad	0.29	0.12, 0.69	0.005	
Post-grad	0.26	0.11, 0.63	0.003	
Other	0.32	0.09, 1.08	0.07	
MacArthur Subjective Social Status Ladder	0.82	0.77, 0.86	<0.001	
Health care worker	1.01	0.81, 1.28	0.90	
Anemia	1.63	1.31, 2.03	<0.001	
Cancer	1.02	0.61, 1.69	0.94	
Diabetes	1.60	1.14, 2.26	0.007	
High blood pressure	1.61	1.28, 2.02	<0.001	
HIV	1.03	0.47, 2.25	0.94	
Immunodeficiency	1.58	1.26, 1.97	<0.001	
Children home from college living with you	1.11	0.78, 1.57	0.57	
School-age children living with you	1.20	0.98, 1.47	0.08	
At least weekly exercise	0.48	0.39, 0.58	<0.001	
Average drinks per day	0.90	0.76, 1.07	0.25	
Average sleep duration (hours)	0.84	0.76, 0.93	<0.001	
Cigarettes: any use in last 30 days	2.43	1.79, 3.30	<0.001	
E-cigarettes: any use in last 30 days	2.00	1.41, 2.84	<0.001	
Marijuana: any use in last 30 days	1.32	1.01, 1.74	0.045	
Any pets at home	1.31	1.07, 1.62	0.01	
Hand washing practices				
Much more	reference			0.54*
Somewhat more	1.17	0.94, 1.46	0.16	0.87†
A little more	1.05	0.71, 1.57	0.80	0.52#
No change	1.21	0.62, 2.37	0.57	
Somewhat less	n/a	n/a, n/a	n/a	
Much less	n/a	n/a, n/a	n/a	
Flu shot	0.90	0.72, 1.13	0.35	
Sanitized phone	0.73	0.58, 0.92	0.007	
Local government ordinances				
School closures	0.76	0.57, 1.03	0.07	
Restricted gatherings by venue	0.84	0.57, 1.23	0.37	
Restricted gatherings by number	1.06	0.71, 1.56	0.79	
Recommended working from home	0.83	0.58, 1.18	0.29	
Shelter in place	0.89	0.66, 1.20	0.44	
Other restrictions	0.95	0.77, 1.16	0.59	
Any household symptoms, 6–12 days ago	2.07	1.68, 2.56	<0.001	
Maximum contacts (per 10), 6–12 days ago	1.21	1.12, 1.31	<0.001	
Distance traveled, 6–12 days ago				
0	reference			0.29*
<= 1 km	0.86	0.60, 1.24	0.43	0.22†
>1–10 km	0.87	0.58, 1.29	0.49	0.15#
>10–50 km	1.07	0.70, 1.63	0.75	

(Continued)

Table 2. (Continued)

Characteristic	Odds Ratio	95% CI	p-value	Group p-value
>50 km	0.52	0.25, 1.08	0.08	
Any travel >1000km, 6–12 days ago	0.36	0.09, 1.47	0.15	
Time at home, 6–12 days ago				
<25%	reference			0.64*
25–50%	1.07	0.77, 1.50	0.69	0.55†
50–75%	0.91	0.60, 1.39	0.66	0.51#
75–100%	1.18	0.88, 1.58	0.27	

Models were adjusted for age, sex, race, ethnicity, and date.

* overall heterogeneity.

† heterogeneity of non-reference levels.

linear trend.

<https://doi.org/10.1371/journal.pone.0253120.t002>

Table 3. Independent predictors of incident symptoms.

Characteristic	Odds ratio	95% CI	p-value	Group p-value
Age category				
18–29	reference			0.22*
30–39	0.83	0.60, 1.16	0.28	0.21†
40–49	0.87	0.63, 1.22	0.42	0.15#
50–59	0.96	0.69, 1.34	0.81	
60+	0.69	0.48, 1.00	0.049	
Race/ethnicity				
White	reference			0.41*
Black	1.04	0.46, 2.37	0.92	0.26†
Hispanic (any race)	1.10	0.80, 1.50	0.55	0.80#
Asian or Pacific Islander	0.73	0.48, 1.10	0.14	
Other (including multiracial)	1.31	0.79, 2.16	0.30	
Female Biological Sex	1.75	1.39, 2.20	<0.001	
MacArthur Subjective Social Status Ladder	0.87	0.83, 0.93	<0.001	
Anemia	1.45	1.16, 1.81	0.001	
High blood pressure	1.35	1.08, 1.68	0.007	
At least weekly exercise	0.57	0.47, 0.70	<0.001	
Cigarettes: any use in last 30 days	1.86	1.35, 2.55	<0.001	
Sanitized phone	0.79	0.63, 0.99	0.037	
Any household symptoms, 6–12 days ago	2.06	1.67, 2.55	<0.001	
Maximum contacts (per 10), 6–12 days ago	1.15	1.06, 1.25	<0.001	
Calendar date (linear)	0.93	0.90, 0.97	<0.001	
Calendar date (non-linear)	1.05	1.01, 1.09	0.019	

Derived backwards stepwise elimination of covariates (see [Methods](#)).

* overall heterogeneity.

† heterogeneity of non-reference levels.

linear trend.

<https://doi.org/10.1371/journal.pone.0253120.t003>

phenomena related to the human immune system's general vulnerability to viral infection [31]. Finally, due to the lack of universal testing, there is some evidence that surveillance for viral symptoms may itself have several advantages, often reflecting, either directly or indirectly, underlying COVID-19 disease [9,32,33].

The higher incidence of viral symptoms among women in our cohort runs counter to the prevailing evidence that men are at higher risk of both SARS-CoV-2 infection and related morbidity and mortality [34,35]. Indeed, population-based studies examining seropositivity for antibodies against SARS-CoV-2 have generally failed to demonstrate any differences by biological sex [36,37]. While this may suggest the viral symptoms detected by the current study fail to capture patterns most relevant to SARS-CoV-2, these differences may have occurred because our models adjusted for related mediators (such as smoking) [38] or because women are either more likely to experience or report more mild symptoms. Previous community-based studies often adjust or weight for sex distributions based on the population, which can often hinder a direct assessment of biological sex as a predictor itself [33,39]. Anemia and hypertension as risk factors are consistent with the general notion that other systemic comorbidities enhance the susceptibility to viral infection. While anemia may be a marker of general, non-specific disease, hypertension, not generally considered a risk factor for infectious diseases, has emerged as a consistent predictor of SARS-CoV-2 infection and associated complications [4–7]. The reasons for this are unclear, although angiotensin-converting-enzyme-2 (ACE2)-dependent cellular entry of the virus has been posited as a biologically plausible mechanism, assuming some connection with ACE2-dysregulation that is also associated with hypertension [40,41]. While optimizing blood pressure control reduces overall morbidity and mortality [42], the consequences on incident viral infection have not yet been fully elucidated.

Three at least theoretically readily modifiable exposures, each bolstered by biological plausibility and previous evidence, arose as risk factors for incident viral infection: smoking, household contacts, and the maximum number of recent interactions with other individuals within six feet (about 1.83 meters). The impact of smoking on the risk of SARS-CoV-2 has been difficult to study, as reliance on hospitalization data fails to provide a foundational study base to make comparisons. Conflicting data on the subject exists, with some evidence that smokers may have a lower risk of SARS-CoV-2 infection and other evidence they may experience a higher risk [43]—the great majority of these studies rely on data among those already infected rather than providing information from cohort studies that include a geographically heterogeneous cohort such as ours. Smoking may reduce the effectiveness of the immune response and may also upregulate ACE2, rendering individuals more prone to infection [44]. Or this observation may be related to an overall greater propensity to the symptoms of interest in the current study, rather than SARS-CoV-2 infection per se. The observation that sick household contacts predicted incident symptoms may provide evidence that these symptoms were in fact often due to a transmissible disease. For example, although the current analyses excluded those with prevalent symptoms (which reduces the chance symptoms arose from some chronic, ever-present, problem), shared symptoms within a household may have represented some common exposure or predisposition, such as an allergy—however, household symptoms *preceding* participant symptoms arose as a statistically significant predictor of incident symptoms, supporting viral infections as a culprit. Although physical distancing as a method to mitigate spread of infection is supported by the general understanding of the nature of infectious diseases, particularly respiratory viruses, our observation from prospective, repeatedly updated, individual-level data that the number of human to human physical interactions predicted viral symptoms may provide useful evidence in support of physical distancing.

Protective factors included a higher subjective social status, at least weekly exercise, and sanitizing one's phone. We utilized the MacArthur subjective social status ladder as a validated

single-item question to capture socioeconomic status [19,20]. A higher self-perceived social status may influence viral infection risk in several ways: more education may translate into a better understanding of disease risks and healthy behaviors, and employment among those of a higher socioeconomic status may be more flexible and less often involve high-risk environments. Conversely, stress and the allostatic load related to social determinants of health among those with a lower subjective social status may adversely affect the immune response to infection [45]. Regular exercise is an established means to improve immune function and the response to viral infection [46], now with evidence for beneficial effects specifically in the COVID-19 era. We recognize that sanitizing one's phone as a protective factor may simply serve as a marker of more fastidious behaviors to minimize risk in general, but the ubiquity and frequent use of the smartphone, likely while shopping, while at work, and while interacting with others, would seem to make it a potentially potent fomite that could result in repeated exposure throughout the day and into the home.

Although we examined multiple predictors of incident viral symptoms, we do not believe that adjustment for multiple hypothesis testing would be appropriate for several reasons. First, all of our predictors had biological plausibility. Second, all covariates were adjusted for one-another in our multivariable models. Third, beyond such mutual adjustment, we employed a backward stepwise elimination of covariates to only retain those achieving our prespecified statistical significance. Of note, the majority of statistically significant findings did exhibit particularly small *p* values that would have withstood even the most conservative multiple hypothesis testing adjustment (such as Bonferroni), but, for the reasons described above, we do not believe this would be appropriate as it would risk reclassifying true positives as false positives. These are also standard and well-accepted approaches to this sort of analysis [24].

Our study has several important limitations. The outcome of interest was viral symptoms, which relied on self-report. These findings therefore do not directly reflect any particular disease, including infection with SARS-CoV-2. Indeed, while we selected fever as a required symptom in hopes of capturing infection and at least one other symptom known to be associated with SARS-CoV-2 to help with specificity, the symptoms described are more applicable to respiratory viruses in general and could also occur outside the realm of infection. Importantly, we were able to leverage the prospective nature of our study with repeated assessments and exclude those with prevalent symptoms at baseline, which should help mitigate against contamination by those with chronic and non-infectious conditions. Some investigators have proposed and validated the creation of symptom-based scores for the accurate inference of COVID-19 [47], and future similar efforts may be able to harness the relative accessibility of patient or research participant self-report without reliance on biological assays. Multiple studies have sought to identify particular symptoms or combinations of symptoms most indicative of SARS-CoV-2 infection [14,25,32,33,47]—while the symptoms we employed generally match those associated with the disease in these studies, variability in populations and study design currently preclude the identification of an optimal approach to strike the intended balance regarding sensitivity versus specificity in such longitudinal cohort studies. Indeed, there is evidence the evolving nature of COVID-19 may produce variable manifestations over time, potentially hindering attempts to accurately characterize a true diagnosis by specific symptoms alone [48]. It is important to emphasize that we incorporated information into the study as it arose from the medical literature, such as (as described in the methods) gastrointestinal symptoms, loss of taste and smell, and eye symptoms, soon after they were reported in reputable sources—therefore, not every symptoms was assessed at baseline, which could have led to under-ascertainment of SARS-CoV-2-related symptoms in some. However, those updates were incorporated into both baseline and daily surveys for all participants as they arose. As the study required smartphone use, it is possible our population represents a more technically

savvy and perhaps more highly educated and affluent group than the general population. However, this would primarily limit generalizability and should not serve as a threat to internal validity. We assumed that the absence of a completed daily survey represented an absence of symptoms, which may have resulted in a loss of power to detect some relationships but addressed missing data in a fashion that did not risk creating spurious false positive associations. In addition, the participants were fairly geographically diverse, representing every state in the US and multiple countries. Although less than 80% of the study participants were non-Hispanic white, African American representation was relatively poor. Finally, although the data were collected prospectively and in a time-updated fashion, the study was observational, prone to residual and unmeasured confounding that should temper assumptions of causal effects.

In conclusion, female sex, anemia, hypertension, recent cigarette smoking, living with someone with recent viral symptoms, and the maximum number of recent contacts within six feet (about 1.83 meters) outside the home each predicted a higher risk of developing viral symptoms during the current COVID-19 pandemic. At the same time, a higher subjective social status, regular exercise, and sanitizing one's phone each predicted a lower risk of developing viral symptoms.

Supporting information

S1 Table. Proportion of participants completing at least one daily survey per week and the proportion completing weekly surveys.

(DOCX)

S2 Table. Specific symptoms reported in addition to fevers and chills.

(DOCX)

S3 Table. Sensitivity analysis of independent predictors of incident symptoms excluding anemia. Derived backwards stepwise elimination of covariates (see [Methods](#)). * overall heterogeneity. † heterogeneity of non-reference levels. # linear trend.

(DOCX)

S4 Table. Sensitivity analysis of independent predictors of incident symptoms including participants with atrial fibrillation, coronary artery disease, congestive heart failure, chronic obstructive pulmonary disease, and asthma. Derived backwards stepwise elimination of covariates (see [Methods](#)). * overall heterogeneity. † heterogeneity of non-reference levels. # linear trend.

(DOCX)

S1 File. Baseline survey.

(PDF)

S1 Dataset.

(DTA)

Author Contributions

Conceptualization: Gregory M. Marcus.

Data curation: Gregory M. Marcus, Jeffrey E. Olgin, Noah D. Peyser, Vivian Yang, Sean Joyce, Robert Avram, Geoffrey H. Tison, Xochitl Butcher, Mark J. Pletcher.

Formal analysis: Gregory M. Marcus, Eric Vittinghoff, Robert Avram.

Funding acquisition: Gregory M. Marcus, Jeffrey E. Olgin, Mark J. Pletcher.

Methodology: Gregory M. Marcus, Jeffrey E. Olgin, Noah D. Peyser, Eric Vittinghoff, Xochitl Butcher, Mark J. Pletcher.

Project administration: Gregory M. Marcus.

Software: David Wen, Xochitl Butcher.

Writing – original draft: Gregory M. Marcus.

Writing – review & editing: Gregory M. Marcus, Jeffrey E. Olgin, Noah D. Peyser, Eric Vittinghoff, Vivian Yang, Sean Joyce, Robert Avram, Geoffrey H. Tison, David Wen, Xochitl Butcher, Helena Eitel, Mark J. Pletcher.

References

1. McAnulty JM, Ward K. Suppressing the Epidemic in New South Wales. *N Engl J Med* 2020; 382:e74. <https://doi.org/10.1056/NEJMc2011592> PMID: 32383832
2. Wang J, Pan L, Tang S, Ji JS, Shi X. Mask use during COVID-19: A risk adjusted strategy. *Environ Polut* 2020; 266:115099. <https://doi.org/10.1016/j.envpol.2020.115099> PMID: 32623270
3. Islam N, Sharp SJ, Chowell G, et al. Physical distancing interventions and incidence of coronavirus disease 2019: natural experiment in 149 countries. *BMJ* 2020; 370:m2743. <https://doi.org/10.1136/bmj.m2743> PMID: 32669358
4. Price-Haywood EG, Burton J, Fort D, Seoane L. Hospitalization and Mortality among Black Patients and White Patients with COVID-19. *N Engl J Med* 2020; 382:2534–43. <https://doi.org/10.1056/NEJMsa2011686> PMID: 32459916
5. Goyal P, Choi JJ, Pinheiro LC, et al. Clinical Characteristics of COVID-19 in New York City. *N Engl J Med* 2020; 382:2372–4. <https://doi.org/10.1056/NEJMc2010419> PMID: 32302078
6. Bhatraju PK, Ghassemieh BJ, Nichols M, et al. COVID-19 in Critically Ill Patients in the Seattle Region—Case Series. *N Engl J Med* 2020; 382:2012–22. <https://doi.org/10.1056/NEJMoa2004500> PMID: 32227758
7. Richardson S, Hirsch JS, Narasimhan M, et al. Presenting Characteristics, Comorbidities, and Outcomes Among 5700 Patients Hospitalized With COVID-19 in the New York City Area. *JAMA* 2020.
8. Williamson EJ, Walker AJ, Bhaskaran K, et al. OpenSAFELY: factors associated with COVID-19 death in 17 million patients. *Nature* 2020.
9. Silverman JD, Hupert N, Washburne AD. Using influenza surveillance networks to estimate state-specific prevalence of SARS-CoV-2 in the United States. *Sci Transl Med* 2020. <https://doi.org/10.1126/scitranslmed.abc1126> PMID: 32571980
10. Martín-Sánchez FJ, Del Toro E, Cardassay E, et al. Clinical presentation and outcome across age categories among patients with COVID-19 admitted to a Spanish Emergency Department. *Eur Geriatr Med* 2020. <https://doi.org/10.1007/s41999-020-00359-2> PMID: 32671732
11. Huang C, Wang Y, Li X, et al. Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China. *Lancet* 2020; 395:497–506. [https://doi.org/10.1016/S0140-6736\(20\)30183-5](https://doi.org/10.1016/S0140-6736(20)30183-5) PMID: 31986264
12. Liu J, Cui M, Yang T, Yao P. Correlation between gastrointestinal symptoms and disease severity in patients with COVID-19: a systematic review and meta-analysis. *BMJ Open Gastroenterol* 2020; 7. <https://doi.org/10.1136/bmjgast-2020-000437> PMID: 32665397
13. Adorni F, Prinelli F, Bianchi F, et al. Self-reported symptoms of SARS-CoV-2 infection in a non-hospitalized population: results from the large Italian web-based EPICOV19 cross-sectional survey. *JMIR Public Health Surveill* 2020.
14. Lan FY, Filler R, Mathew S, et al. COVID-19 symptoms predictive of healthcare workers' SARS-CoV-2 PCR results. *PLoS One* 2020; 15:e0235460. <https://doi.org/10.1371/journal.pone.0235460> PMID: 32589687
15. Gangaputra SS, Patel SN. Ocular Symptoms among Nonhospitalized Patients Who Underwent COVID-19 Testing. *Ophthalmology* 2020. <https://doi.org/10.1016/j.ophtha.2020.06.037> PMID: 32585259
16. 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. <https://doi.org/10.1371/journal.pone.0234765> PMID: 32574165
17. Dawson P, Rabold EM, Laws RL, et al. Loss of Taste and Smell as Distinguishing Symptoms of COVID-19. *Clin Infect Dis* 2020.
 18. Scalinci SZ, Trovato Battagliola E. Conjunctivitis can be the only presenting sign and symptom of COVID-19. *IDCases* 2020; 20:e00774. <https://doi.org/10.1016/j.idcr.2020.e00774> PMID: 32373467
 19. Goodman E, Adler NE, Kawachi I, Frazier AL, Huang B, Colditz GA. Adolescents' perceptions of social status: development and evaluation of a new indicator. *Pediatrics* 2001; 108:E31. <https://doi.org/10.1542/peds.108.2.e31> PMID: 11483841
 20. Cooper DC, Milic MS, Mills PJ, Bardwell WA, Ziegler MG, Dimsdale JE. Endothelial function: the impact of objective and subjective socioeconomic status on flow-mediated dilation. *Ann Behav Med* 2010; 39:222–31. <https://doi.org/10.1007/s12160-010-9181-9> PMID: 20376585
 21. Melvin RL, Xiao J, Godwin RC, Berenhaut KS, Salsbury FR Jr. Visualizing correlated motion with HDBSCAN clustering. *Protein Sci* 2018; 27:62–75. <https://doi.org/10.1002/pro.3268> PMID: 28799290
 22. Castro Gertrudes J, Zimek A, Sander J, Campello R. A unified view of density-based methods for semi-supervised clustering and classification. *Data Min Knowl Discov* 2019; 33:1894–952. <https://doi.org/10.1007/s10618-019-00651-1> PMID: 32831623
 23. Alene M, Yismaw L, Assemie MA, Ketema DB, Gietaneh W, Birhan TY. Serial interval and incubation period of COVID-19: a systematic review and meta-analysis. *BMC Infect Dis* 2021; 21:257. <https://doi.org/10.1186/s12879-021-05950-x> PMID: 33706702
 24. Lessler J, Brookmeyer R, Reich NG, Nelson KE, Cummings DA, Perl TM. Identifying the probable timing and setting of respiratory virus infections. *Infect Control Hosp Epidemiol* 2010; 31:809–15. <https://doi.org/10.1086/655023> PMID: 20569117
 25. Tostmann A, Bradley J, Bousema T, et al. Strong associations and moderate predictive value of early symptoms for SARS-CoV-2 test positivity among healthcare workers, the Netherlands, March 2020. *Euro Surveill* 2020; 25. <https://doi.org/10.2807/1560-7917.ES.2020.25.16.2000508> PMID: 32347200
 26. Center for Systems Science and Engineering at Johns Hopkins University. COVID-19 dashboard. (<https://coronavirus.jhu.edu/map.html>). 2020.
 27. Tromberg BJ, Schwetz TA, Pérez-Stable EJ, et al. Rapid Scaling Up of COVID-19 Diagnostic Testing in the United States—The NIH RADx Initiative. *N Engl J Med* 2020. <https://doi.org/10.1056/NEJMs2022263> PMID: 32706958
 28. Wang D, Li Z, Liu Y. An overview of the safety, clinical application and antiviral research of the COVID-19 therapeutics. *J Infect Public Health* 2020.
 29. COVID-19 therapies and vaccine landscape. *Nature materials* 2020; 19:809. <https://doi.org/10.1038/s41563-020-0758-9> PMID: 32704138
 30. Kutter JS, Spronken MI, Fraaij PL, Fouchier RA, Herfst S. Transmission routes of respiratory viruses among humans. *Curr Opin Virol* 2018; 28:142–51. <https://doi.org/10.1016/j.coviro.2018.01.001> PMID: 29452994
 31. Rouse BT, Sehrawat S. Immunity and immunopathology to viruses: what decides the outcome? *Nat Rev Immunol* 2010; 10:514–26. <https://doi.org/10.1038/nri2802> PMID: 20577268
 32. Wynants L, Van Calster B, Collins GS, et al. Prediction models for diagnosis and prognosis of covid-19 infection: systematic review and critical appraisal. *BMJ* 2020; 369:m1328. <https://doi.org/10.1136/bmj.m1328> PMID: 32265220
 33. Menni C, Valdes AM, Freidin MB, et al. Real-time tracking of self-reported symptoms to predict potential COVID-19. *Nat Med* 2020; 26:1037–40. <https://doi.org/10.1038/s41591-020-0916-2> PMID: 32393804
 34. Williamson EJ, Walker AJ, Bhaskaran K, et al. Factors associated with COVID-19-related death using OpenSAFELY. *Nature* 2020; 584:430–6. <https://doi.org/10.1038/s41586-020-2521-4> PMID: 32640463
 35. Peckham H, de Groot NM, Raine C, et al. Male sex identified by global COVID-19 meta-analysis as a risk factor for death and ICU admission. *Nat Commun* 2020; 11:6317. <https://doi.org/10.1038/s41467-020-19741-6> PMID: 33298944
 36. Pagani G, Conti F, Giacomelli A, et al. Seroprevalence of SARS-CoV-2 significantly varies with age: Preliminary results from a mass population screening. *J Infect* 2020; 81:e10–e2. <https://doi.org/10.1016/j.jinf.2020.09.021> PMID: 32961253
 37. Pollan M, Perez-Gomez B, Pastor-Barriuso R, et al. Prevalence of SARS-CoV-2 in Spain (ENE-COVID): a nationwide, population-based seroepidemiological study. *Lancet* 2020; 396:535–44. [https://doi.org/10.1016/S0140-6736\(20\)31483-5](https://doi.org/10.1016/S0140-6736(20)31483-5) PMID: 32645347
 38. Cai H. Sex difference and smoking predisposition in patients with COVID-19. *The Lancet Respiratory medicine* 2020; 8:e20. [https://doi.org/10.1016/S2213-2600\(20\)30117-X](https://doi.org/10.1016/S2213-2600(20)30117-X) PMID: 32171067

39. Menachemi N, Yiannoutsos CT, Dixon BE, et al. Population Point Prevalence of SARS-CoV-2 Infection Based on a Statewide Random Sample—Indiana, April 25–29, 2020. *MMWR Morb Mortal Wkly Rep* 2020; 69:960–4. <https://doi.org/10.15585/mmwr.mm6929e1> PMID: 32701938
40. Gold MS, Sehayek D, Gabrielli S, Zhang X, McCusker C, Ben-Shoshan M. COVID-19 and comorbidities: a systematic review and meta-analysis. *Postgrad Med* 2020:1–7. <https://doi.org/10.1080/00325481.2020.1786964> PMID: 32573311
41. Liu PP, Blet A, Smyth D, Li H. The Science Underlying COVID-19: Implications for the Cardiovascular System. *Circulation* 2020; 142:68–78. <https://doi.org/10.1161/CIRCULATIONAHA.120.047549> PMID: 32293910
42. Benjamin EJ, Blaha MJ, Chiuve SE, et al. Heart Disease and Stroke Statistics-2017 Update: A Report From the American Heart Association. *Circulation* 2017; 135:e146–e603. <https://doi.org/10.1161/CIR.0000000000000485> PMID: 28122885
43. Haddad C, Bou Malhab S, Sacre H, Salameh P. Smoking and COVID-19: A Scoping Review. *Tob Use Insights* 2021; 14:1179173X21994612. <https://doi.org/10.1177/1179173X21994612> PMID: 33642886
44. Farsalinos K, Barbouni A, Poulas K, Polosa R, Caponnetto P, Niaura R. Current smoking, former smoking, and adverse outcome among hospitalized COVID-19 patients: a systematic review and meta-analysis. *Ther Adv Chronic Dis* 2020; 11:2040622320935765. <https://doi.org/10.1177/2040622320935765> PMID: 32637059
45. Juster RP, Sindi S, Marin MF, et al. A clinical allostatic load index is associated with burnout symptoms and hypocortisolemic profiles in healthy workers. *Psychoneuroendocrinology* 2011; 36:797–805. <https://doi.org/10.1016/j.psyneuen.2010.11.001> PMID: 21129851
46. Simpson RJ, Kunz H, Agha N, Graff R. Exercise and the Regulation of Immune Functions. *Prog Mol Biol Transl Sci* 2015; 135:355–80. <https://doi.org/10.1016/bs.pmbts.2015.08.001> PMID: 26477922
47. Bastiani L, Fortunato L, Pieroni S, et al. Rapid COVID-19 Screening Based on Self-Reported Symptoms: Psychometric Assessment and Validation of the EPICOV19 Short Diagnostic Scale. *J Med Internet Res* 2021; 23:e23897. <https://doi.org/10.2196/23897> PMID: 33320825
48. Lan FY, Filler R, Mathew S, et al. Evolving virulence? Decreasing COVID-19 complications among Massachusetts healthcare workers: a cohort study. *Pathog Glob Health* 2021; 115:4–6. <https://doi.org/10.1080/20477724.2020.1847778> PMID: 33191880