BMJ Open

Predictors of incident SARS-CoV-2 infections in an international prospective cohort study

Anthony Lin [ID],[1] Eric Vittinghoff [ID],[2] Jeffrey Olgin,[1] Noah Peyser [ID],[1] Sidney Aung,[1] Sean Joyce,[1] Vivian Yang,[1] Janet Hwang,[1] Robert Avram,[1] Gregory Nah,[1] Geoffrey H Tison,[1] Alexis Beatty,[1] Ryan Runge,[1] David Wen [ID],[1] Xochitl Butcher,[1] Cathy Horner,[1] Helena Eitel,[1] Mark Pletcher,[2] Gregory M Marcus[1]

[1]Department of Medicine, University of California San Francisco, San Francisco, California, USA
[2]Epidemiology and Biostatistics, University of California San Francisco, San Francisco, California, USA

Correspondence to
Dr Gregory M Marcus;
Greg.Marcus@ucsf.edu

ABSTRACT

Objective Until effective treatments and vaccines are made readily and widely available, preventative behavioural health measures will be central to the SARS-CoV-2 public health response. While current recommendations are grounded in general infectious disease prevention practices, it is still not entirely understood which particular behaviours or exposures meaningfully affect one's own risk of incident SARS-CoV-2 infection. Our objective is to identify individual-level factors associated with one's personal risk of contracting SARS-CoV-2.

Design Prospective cohort study of adult participants from 26 March 2020 to 8 October 2020.

Setting The COVID-19 Citizen Science Study, an international, community and mobile-based study collecting daily, weekly and monthly surveys in a prospective and time-updated manner.

Participants All adult participants over the age of 18 years were eligible for enrolment.

Primary outcome measure The primary outcome was incident SARS-CoV-2 infection confirmed via PCR or antigen testing.

Results 28 575 unique participants contributed 2 479 149 participant-days of data across 99 different countries. Of these participants without a history of SARS-CoV-2 infection at the time of enrolment, 112 developed an incident infection. Pooled logistic regression models showed that increased age was associated with lower risk (OR 0.98 per year, 95% CI 0.97 to 1.00, p=0.019), whereas increased number of non-household contacts (OR 1.10 per 10 contacts, 95% CI 1.01 to 1.20, p=0.024), attending events of at least 10 people (OR 1.26 per 10 events, 95% CI 1.07 to 1.50, p=0.007) and restaurant visits (OR 1.95 per 10 visits, 95% CI 1.42 to 2.68, p<0.001) were associated with significantly higher risk of incident SARS-CoV-2 infection.

Conclusions Our study identified three modifiable health behaviours, namely the number of non-household contacts, attending large gatherings and restaurant visits, which may meaningfully influence individual-level risk of contracting SARS-CoV-2.

STRENGTHS AND LIMITATIONS OF THIS STUDY

⇒ This large international cohort study with 2.4 million participant-days of data from participants in 99 different countries provides unprecedented geographical diversity for a study analysing individual-level factors associated with risk of SARS-CoV-2.

⇒ All participants included in this study were free of SARS-CoV-2 infection early in the pandemic, allowing for real-time ascertainment of significant individual-level behaviours and exposures related to higher risk of incident infection.

⇒ Using PCR or antigen testing as the gold standard for SARS-CoV-2 infections relied on a participant's development of symptoms, index of suspicion and access to testing facilities, but ensured our study identified risk factors associated with true infection and increased specificity over traditional methods of symptom reporting.

INTRODUCTION

The novel coronavirus (SARS-CoV-2) global pandemic has created a major public health crisis for nearly every country and community in the world. Responses to mitigate transmission have varied by government, but have generally been grounded in known respiratory virus disease prevention practices. Current strategies have included a combination of social distancing, limitations to travel and public gatherings, increased handwashing practices, and use of face masks. While these interventions are believed to reduce human-to-human transmission, efforts to study these interventions have been limited as they rely on individual-level behaviours that are dynamic with policy changes and can be difficult to capture at scale. Furthermore, the politicisation of social distancing recommendations[1–3] makes it difficult to fully understand levels of compliance at the individual level and calls for a larger evidence base for recommendations like handwashing, face mask wearing, and limiting human contact, large social gatherings and visits to restaurants. Identifying predictors of infection requires a

longitudinal cohort study. The information gleaned from the longitudinal characterisation of SARS-CoV-2 infection risk factors may be crucial to understanding which strategies are most effective and can further inform public policy. Moreover, such data may help elucidate the individual behaviours directly under one's control to influence one's personal risk of contracting SARS-CoV-2.

While previous prospective studies have focused primarily on symptom detection and the constellation of symptoms associated with SARS-CoV-2 infection,[4–7] mobile technology provides an opportunity to study the effects of various exposures and behaviours that can be ascertained prospectively, repeatedly and in nearly real time. The majority of previous research regarding SARS-CoV-2 has focused on hospitalised individuals, primarily those who already have the disease, and predictors of disease severity as opposed to those pertinent to developing infection. This is not surprising as accumulating sufficient numbers to characterise non-infected individuals at baseline and then follow them over time is generally time-consuming and would require enrolment of particularly large numbers to derive useful results. While systematic reviews and meta-analyses of previous studies have investigated the efficacy of behavioural interventions,[8 9] we are not aware of a longitudinal cohort study in which risk factors have been characterised in detail prior to infection and exposures and behaviours tracked as individuals contracted (or did not contract) SARS-CoV-2 in the community.

Given the widespread use of smartphones and associated mobile apps, the technology is now available to regularly query large populations to assess patterns in SARS-CoV-2 infection rates based on individual-level exposures and behaviours. We have previously demonstrated the utility of this technology in characterising ambulatory cardiovascular risk factors.[10–14] In this study, we sought to use prospectively collected information from the COVID-19 Citizen Science Study to identify individual characteristics, exposures or behaviours associated with an increased risk of contracting SARS-CoV-2.

METHODS
Study design
The COVID-19 Citizen Science Study is a mobile application that enables longitudinal and time-updated collection of health survey and location data from thousands of global participants. The application was developed by a team of investigators at the University of California, San Francisco using the Eureka Digital Research Platform. Enrolment began on 26 March 2020 and is ongoing (https://covid19.eurekaplatform.org/). The current analysis included participant information collected until 8 October 2020. Enrolment is available to all adults over the age of 18 years and has been facilitated by press releases, social media and word of mouth.

Informed electronic consent was obtained remotely using mobile application at the time of study enrolment.

Data collection
Surveys collected information about demographics, medical comorbidities, SARS-CoV-2 infection status, daily behaviours, environmental or social exposures, and symptoms. Surveys were written in English and met the Flesch-Kincaid criteria for eighth-grade reading level (https://readabilityformulas.com). Participants received a baseline survey at the time of enrolment ascertaining general demographic information such as age, race/ethnicity, sex, education level, MacArthur subjective social status, occupation, smoking patterns, presence of children or pets at home, and pre-existing medical comorbidities. After completing the baseline survey, participants then received daily surveys that enquired about current symptoms, household contacts and non-household contacts; weekly surveys that assessed changes to individual-level behaviours such as sleep, exercise, social distancing efforts, hand hygiene and use of face masks while out in public; and monthly surveys that collected information regarding employment, mood and alcohol consumption (online supplemental appendix 1).

The MacArthur Subjective Social Status Ladder was used as a previously validated single-item question to capture the socioeconomic status of study participants, with higher point ratings indicating higher subjective social status.[15 16] Occupation was dichotomised based on working in healthcare or not. Exercise was defined as self-reported physical activity lasting for at least 20 min and resulted in heavy breathing or 'break[ing] a sweat' and was categorised into never/rarely, <1 time/month, <1 time/week, approximately weekly, 2–4 days/week and >4 days/week. Alcohol use was categorised into none, >0–7 standard drinks per week, >7–14 standard drinks per week and >14 standard drinks per week. Smoking activity was differentiated by use of cigarettes, e-cigarettes or marijuana and then dichotomised by any use in the last 30 days or not. Daily contacts were defined as any non-household individual with which the participant was within 1.83 meters (6 feet) of during the course of the day.

Participants were queried regarding PCR or antigen testing at baseline and during the weekly survey. Using triggered logic, related questions distinguished between evidence of active infection with the PCR test from other tests, such as antibody tests (the latter were not considered sufficient to constitute incident infection). All participants who reported a positive PCR or antibody test for SARS-CoV-2 prior to enrolling in the study were excluded from this analysis. Self-reported positive PCR tests for SARS-CoV-2 were validated by contacting a sample of participants and obtaining documentation of test results (online supplemental appendix 2).

Patient and public involvement
The COVID-19 Citizen Science Study, which remains open to any interested adult with a smartphone, was designed to answer questions most relevant to patients and the lay public, with an emphasis on identifying clinically

relevant behaviours and exposures that can be modified or influenced by any individual. The study was launched using the National Institutes of Health-supported Eureka Digital Research Platform, which was heavily influenced by prior work designing and implementing the Health eHeart Study[17]—from the beginning, these studies have included patients as key stakeholders, such as the Patient-Centered Outcomes Research Institute-supported Health eHeart Alliance,[18] to assure that the user experience was relatable and understandable to interested participants around the world. Modifications of questions and the basic content of some research questions were derived from participant feedback received ad hoc and as a result of campaigns to solicit novel research questions from participants for incorporation into the study. All participants in the COVID-19 Citizen Science Study are encouraged to help with recruitment, with regular reminders via text messages, push notifications and newsletters to share the link and/or 'text back' with friends and family members. Results are disseminated back to COVID-19 Citizen Scientists in the form of data visualisations and text shared via newsletters, the study website and links sent via text message or app-based push notification.

Statistical analyses

Baseline continuous variables are presented using mean and SD or median and IQR, while categorical variables are presented as frequencies (percentages), and compared between participants who reported incident infection and those remaining infection-free using t-tests for continuous variables and χ^2 tests for categorical variables. Pooled logistic regression models for repeated SARS-CoV-2 test results self-reported on the weekly surveys were used to identify factors, obtained from the baseline and earlier weekly and daily surveys, associated with incident infection. We considered demographics; pre-existing medical conditions; behavioural contributors such as mask wearing, hand hygiene and social distancing efforts; and individual exposures such as number of non-household contacts, large gatherings, and visits to gyms,

restaurants and movie theatres. Exposures from earlier weekly and daily surveys were averaged over measurements obtained 4–21 days prior to the weekly survey providing the SARS-CoV-2 test result. All variables associated with SARS-CoV-2 infection with p values <0.1 in the pooled logistic regression models adjusting for only a three-knot restricted cubic spline in calendar date were included in a fully adjusted pooled logistic regression model. In a sensitivity analysis, backward deletion was used to select a more parsimonious pooled logistic regression model retaining covariates with p values <0.05. These models all used robust SEs to account for clustering of the repeated weekly SARS-CoV-2 test results by participant. Additionally, recognising the importance of geographical location, sensitivity analyses restricted to US participants were performed accounting for clustering by county-based Federal Information Processing System (FIPS) and zip codes. All analyses used complete case data. Two-tailed p values <0.05 were considered statistically significant. All statistical analyses were performed using Stata V.16.

RESULTS

After excluding 628 participants with prevalent SARS-CoV-2 infection, 28 575 individuals without a history of SARS-CoV-2 infection at baseline contributed 2 479 149 participant-days of data to the COVID-19 Citizen Science Study across 99 different countries, including all 50 states in the USA (figure 1). The mean proportion of participants who completed at least one health survey during a study week was 88.6%±5.0% and the mean proportion of participants who completed at least one health survey during a study month was 98.1%±1.6% (online supplemental tables 1 and 2). Of the total study population, 112 participants (0.4%) developed a SARS-CoV-2 infection during the study period. Differences in participant demographics, baseline comorbidities, behaviours and exposures between participants who became infected during
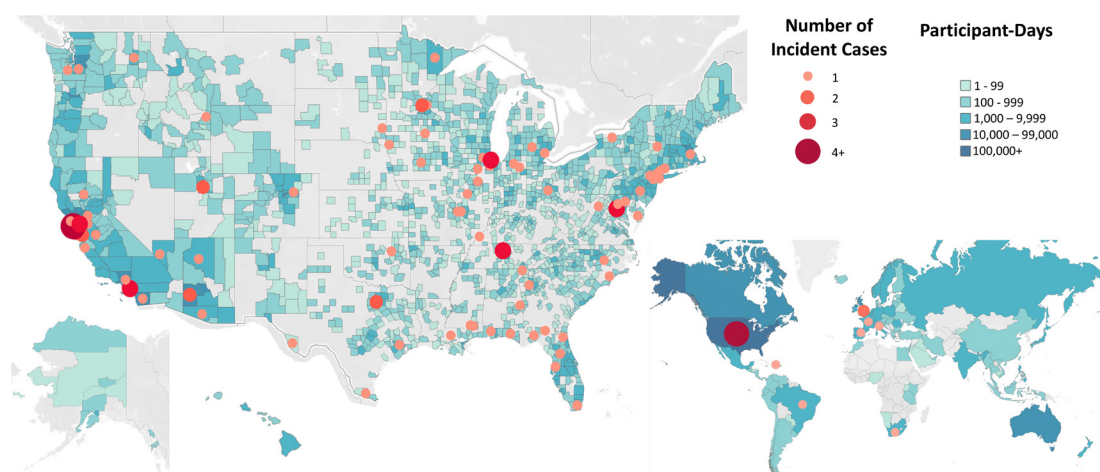
Figure 1 Location of all study participants. The blue shading represents the number of participant-days by county within the USA and by nation in the world. The red shading illustrates all participants infected by SARS-CoV-2 during the study period.

the study period and those who did not are displayed in table 1.

After adjusting only for age, sex, race/ethnicity and calendar date, older age, higher education level, higher subjective social status and increased alcohol use were associated with lower risk, while working in healthcare, a history of HIV, e-cigarette use, less exercise frequency, increased number of recent contacts, attending gatherings with at least 10 people, and visiting movie theatres and restaurants were each associated with a higher risk of incident SARS-CoV-2 infection (table 2). Importantly, pertinent factors that failed to exhibit statistically significant relationships included common medical comorbidities like hypertension, diabetes, coronary artery disease, congestive heart failure, atrial fibrillation, asthma or chronic obstructive pulmonary disease, as well as handwashing practices and mask wearing frequency. Pooled logistic regression models that incorporated all eligible predictors showed that increased age was associated with lower risk of developing a SARS-CoV-2 infection, whereas increased number of contacts, attending events of at least 10 people and visits to restaurants were associated with significantly higher risk of later testing positive for SARS-CoV-2 (figure 2). Backward stepwise deletion did not change any of the statically significant relationships (online supplemental table 3). Similarly, the sensitivity analysis using county-based FIPS and zip codes as random effects in USA-based data did not meaningfully change the results (online supplemental tables 4 and 5).

DISCUSSION

Among an international cohort free of SARS-CoV-2 at baseline and tracked longitudinally, prospectively and in a time-updated manner, increased number of daily non-household contacts within 1.83 meters (6 feet), events of 10 or more individuals and restaurant visits each independently predicted a higher risk of developing SARS-CoV-2 infection. Increased age was associated with a lower risk of subsequently developing SARS-CoV-2 infection.

As of 22 March 2021, there have been over 123 million confirmed cases of SARS-CoV-2 and over 2.7 million SARS-CoV-2-related deaths worldwide.[19] The pandemic has been exacerbated by a recent resurgence of a 'second wave' of SARS-CoV-2 cases and confirmation of new strains with potentially increased transmissibility. The pandemic has spurred international efforts to improve testing capabilities,[20] identify therapies to treat the novel coronavirus[21] and develop vaccines designed to prevent it.[22 23] Even as vaccines from biopharmaceutical companies like Pfizer and Moderna are being delivered, distribution to members of the public has been slow in nearly every country and community, with only countries like Israel, the United Arab Emirates, Chile and the UK managing to administer at least 40 vaccine doses per 100 people.[24] Until and if production, distribution, administration and acceptability of approved vaccines can satisfy the overwhelming need throughout the international community, the identification of preventative health behaviours under an individual's control is crucial to the SARS-CoV-2 public health response.

The COVID-19 Citizen Science Study launched on 26 March 2020 and has been ongoing while recommendations to limit disease transmission continue to evolve at variable rates across the globe. The study has been prospectively collecting data through the initial shelter-in-place recommendations in early 2020 and continues to capture changes in behavioural health patterns as the second spike of SARS-CoV-2 infections surmounts. Our study observed an increased association of SARS-CoV-2 infection in individuals who reported higher numbers of recent contacts. In a similar vein, increased attendance of events of 10 or more people and restaurant visits were associated with increased odds of developing SARS-CoV-2 infection. Given our general understanding of disease transmission for respiratory viruses and recent research characterising the asymptomatic transmission of SARS-CoV-2,[25 26] these findings are bolstered by biological plausibility. They add to previous research supporting the use of government-mandated physical distancing policies to reduce SARS-CoV-2 incidence[27 28] and suggest that behaviours to minimise human-to-human interaction could be effective means to lower one's individual risk of contracting SARS-CoV-2. To our knowledge, this is the first longitudinal cohort to determine that such behaviours among individuals prior to infection actually influence risk.

While the lower risk among older individuals may at first glance appear counterintuitive, this may be consistent with similar protective behaviours and compliance with social distancing behaviours, especially given data reporting high incidence of SARS-CoV-2 in nursing homes[29] as well as disproportionately higher rates of hospitalisation and death in older populations infected with SARS-CoV-2.[30 31] If such phenomena were operative, the fact that we were unable to detect differences in such behaviours (such as significant relationships between hand hygiene or mask wearing) may be due to collinearity with age and/or suboptimal ascertainment of the actual protective approaches used by older individuals. Also contrary to most reports, medical comorbidities thought to increase one's risk of morbidity and mortality from SARS-CoV-2,[32 33] such as hypertension, diabetes, congestive heart failure, chronic obstructive pulmonary disease, cancer and history of myocardial infarctions, were not retained predictors in our multivariate models, suggesting that prior comorbidities may affect one's response to SARS-CoV-2, but may not play a large role in an individual's risk of contracting SARS-CoV-2.

While previous studies have observed benefits in universal masking at the community level,[34 35] our study did not reveal a clear association between an individual's mask wearing behaviour and their risk for SARS-CoV-2 infection. Similarly, self-reported frequency of handwashing did not seem to consistently correlate with SARS-CoV-2 incidence as well. Simple frequencies of mask wearing and handwashing behaviours may be too confounded or measured

Table 1 Demographics, comorbidities and behavioural risk factors of participants in the COVID-19 Citizen Science Study assessed at the time of enrolment, divided by participants who later tested positive for COVID-19 during the study period and participants who did not

	Participants with incident SARS-CoV-2 (n=112)	Participants without incident SARS-CoV-2 (n=28 463)	P value
Age, median (IQR)	46.0 (37.0–55.5)	44.0 (36.0–55.0)	0.84
Age category, n (%)			0.71
18–29	12 (10.7)	2594 (9.2)	
30–39	26 (23.2)	7832 (27.7)	
40–49	31 (27.7)	7121 (25.2)	
50–59	27 (24.1)	6041 (21.3)	
60+	16 (14.3)	4711 (16.6)	
Female biological sex, n (%)	71 (65.1)	18 908 (67.7)	0.79
Race/ethnicity, n (%)			0.25
White	94 (86.2)	23 814 (85.2)	
Black	2 (1.8)	229 (0.8)	
Hispanic (any race)	9 (8.3)	1902 (6.8)	
Asian or Pacific Islander	1 (0.9)	1384 (5.0)	
Other (including multiracial)	3 (2.8)	618 (2.2)	
Highest level of education, median, n (%)			<0.001
Less than high school	1 (0.9)	101 (0.4)	
High school graduate	8 (7.3)	882 (3.2)	
Some college	24 (22.0)	4091 (14.7)	
College graduate	40 (36.7)	9891 (35.4)	
Postgraduate	33 (30.3)	12 690 (45.5)	
Other	3 (2.8)	247 (0.9)	
MacArthur Subjective Social Status Ladder, mean (SD)	6.6 (1.5)	6.9 (1.6)	0.054
Working in healthcare, n (%)	31 (27.7)	5719 (20.1)	0.046
Medical comorbidities, n (%)			
Atrial fibrillation	2 (1.8)	835 (3.0)	0.47
Anaemia	8 (7.2)	2957 (10.5)	0.26
AnaemiaAsthma	9 (8.0)	2815 (9.9)	0.5
Coronary artery disease	2 (1.8)	693 (2.4)	0.65
Cancer	5 (4.5)	908 (3.2)	0.45
Congestive heart failure	1 (0.9)	174 (0.6)	0.71
COPD	2 (1.8)	444 (1.6)	0.84
Diabetes	8 (7.1)	1163 (4.1)	0.11
Hypertension	31 (27.7)	5675 (20.1)	0.045
HIV	3 (2.7)	108 (0.4)	<0.001
Other immunodeficiency	4 (3.6)	542 (1.9)	0.21
History of heart attack	2 (1.8)	283 (1.0)	0.4
Sleep apnoea	13 (11.7)	3019 (10.8)	0.75
History of stroke	2 (1.8)	355 (1.3)	0.6
Alcohol use			0.1
None	26 (24.8)	6541 (25.7)	
>0–7 drinks per week	60 (57.1)	13 362 (52.6)	
>7–14 drinks per week	18 (17.1)	3764 (14.8)	

Continued

Table 1 Continued

	Participants with incident SARS-CoV-2 (n=112)	Participants without incident SARS-CoV-2 (n=28 463)	P value
>14 drinks per week	1 (1.0)	1743 (6.9)	
Smoking			
Cigarette use in the last 30 days	8 (7.2)	1421 (5.0)	0.29
E-cigarette use in the last 30 days	5 (4.5)	723 (2.6)	0.19
Marijuana use in the last 30 days	10 (9.0)	2650 (9.5)	0.87
Sleep duration, median (IQR)	7.0 (6.0–8.0)	7.0 (6.0–8.0)	
Living with children at home, n (%)	34 (30.4)	8926 (31.6)	0.78
Living with pets at home, n (%)	73 (65.8)	18 442 (64.9)	0.86
Use of face masks, n (%)			0.29
'Never'	10 (9.3)	1650 (6.0)	
'Sometimes'	17 (15.7)	3359 (12.2)	
'Most times'	75 (69.4)	20 591 (74.8)	
'Always'	6 (5.6)	1910 (6.9)	
Handwashing frequency, n (%)			0.32
<1 time/day	0 (0)	55 (0.2)	
~1 time/day	1 (0.9)	341 (1.2)	
2–4 times/day	16 (14.3)	4699 (16.5)	
5–10 times/day	47 (42.0)	13 866 (48.7)	
>10 times/day	48 (42.9)	9502 (33.4)	
Exercise frequency, n (%)			<0.001
Never/rarely	5 (4.5)	1591 (5.6)	
<1 time/month	16 (14.3)	2369 (8.3)	
<1 time/week	23 (20.5)	3678 (12.9)	
Approximately weekly	12 (10.7)	3668 (12.9)	
2–4 days/week	30 (26.8)	8956 (31.5)	
>4 days/week	23 (20.5)	8107 (28.5)	
Number of contacts in the past 24 hours, mean (SD)	3.8 (6.2)	3.1 (7.3)	0.36
Number of events with 10 or more people in the past week, mean (SD)	3.8 (14.0)	1.9 (9.8)	0.035
Number of gym visits in the past week, mean (SD)	0.4 (3.4)	0.9 (6.6)	0.5
Number of visits to movie theatres in the past week, mean (SD)	0.1 (0.9)	0.1 (1.6)	0.81
Number of visits to restaurants in the past week, mean (SD)	3.4 (9.3)	2.2 (7.7)	0.095

COPD, chronic obstructive pulmonary disease.

too imprecisely to observe a consistent trend in our data. Additionally, the higher prevalence of healthcare workers in the study population may have resulted in participants having higher rates of mask wearing and handwashing, but also higher risk of infection, thereby degrading any associations between predictor and outcome. As such, these negative results should be interpreted cautiously in the context of the study design, and insufficient power may render negative results (or lack of associations) less informative than the statistically significant relationships (positive results)

that have been observed thus far (even if in the absence of a longitudinal cohort with time-updated assessments as described here).

Our study has a number of important limitations to note. While focusing on individual-level behaviours mitigated issues involving compliance compared with studies examining state-level or country-level government mandates, self-report is still a subjective process and still prone to bias based on differing definitions of qualitative words (ie, 'sometimes' vs 'most times'). However, health survey

Table 2 Minimally adjusted odds of incident SARS-CoV-2 infection

	OR	95% CI	P value	Group p value
Age (years)	0.98	0.96 to 0.99	<0.001	
Female biological sex	0.94	0.63 to 1.39	0.76	
Race/ethnicity				
White	Reference			0.35*
Black	2.04	0.50 to 8.27	0.32	0.24†
Hispanic (any race)	1.2	0.61 to 2.39	0.59	
Asian or Pacific Islander	0.18	0.02 to 1.26	0.08	
Other (including multiracial)	1.22	0.39 to 3.85	0.73	
Highest level of education				
Less than high school	Reference			<0.001*
High school graduate	0.91	0.11 to 7.44	0.93	<0.001†
Some college	0.51	0.07 to 3.87	0.52	0.42‡
College graduate	0.34	0.05 to 2.56	0.3	
Postgraduate	0.2	0.03 to 1.51	0.12	
Other	1.02	0.10 to 10.02	0.99	
MacArthur Subjective Social Status Ladder (per point on scale)	0.87	0.79 to 0.96	0.004	
Working in healthcare	1.66	1.09 to 2.50	0.017	
Medical comorbidities				
Atrial fibrillation	0.38	0.09 to 1.55	0.18	
Anaemia	0.65	0.32 to 1.34	0.24	
Asthma	0.78	0.40 to 1.55	0.48	
Coronary artery disease	0.46	0.11 to 1.89	0.28	
Cancer	0.96	0.39 to 2.34	0.92	
Congestive heart failure	0.99	0.14 to 7.09	0.99	
COPD	0.84	0.21 to 3.44	0.81	
Diabetes	1.37	0.67 to 2.83	0.39	
Hypertension	1.13	0.75 to 1.71	0.56	
HIV	5.31	1.65 to 17.12	0.005	
Other immunodeficiency	1.57	0.58 to 4.25	0.37	
History of heart attack	1.16	0.28 to 4.73	0.84	
Sleep apnoea	0.91	0.51 to 1.62	0.74	
History of stroke	1	0.25 to 4.08	1	
Alcohol use				
None	Reference			0.25*
>0–7 drinks per week	0.95	0.60 to 1.51	0.83	0.13†
>7–14 drinks per week	1.01	0.55 to 1.84	0.97	0.047‡
>14 drinks per week	0.13	0.02 to 0.95	0.044	
Smoking				
Cigarette use in the last 30 days	1.91	0.94 to 3.88	0.07	
E-cigarette use in the last 30 days	2.98	1.64 to 5.41	<0.001	
Marijuana use in the last 30 days	1.03	0.56 to 1.84	0.93	
Mean sleep duration (per hour of sleep)	1.13	0.86 to 1.49	0.37	
Living with children at home	1.23	0.89 to 1.71	0.21	
Living with pets at home	1.35	0.88 to 2.07	0.17	
Use of face masks, last 4–21 days				
'Never'	Reference			

Continued

Table 2 Continued

	OR	95% CI	P value	Group p value
'Sometimes'	1.15	0.50 to 2.61	0.74	
'Most times' or 'Always'	1.11	0.45 to 2.72	0.82	
Handwashing frequency, last 4–21 days				
<2 times/day	Reference			
2–4 times/day	1.36	0.65 to 2.81	0.41	
5–10 times/day	1.08	0.59 to 1.95	0.8	
>10 times/day	1.5	0.81 to 2.77	0.2	
Exercise frequency, last 4–21 days				
<1 time/month	Reference			
<1 time/week	2.21	1.31 to 3.76	0.003	
Approximately weekly	1.25	0.76 to 2.04	0.38	
2–4 days/week	1.18	0.73 to 1.92	0.5	
>4 days/week	0.91	0.51 to 1.64	0.76	
Number of contacts (per 10), last 4–21 days	1.17	1.09 to 1.26	<0.001	
Number of events with 10 or more people (per 10), last 4–21 days	1.04	1.03 to 1.05	<0.001	
Number of gym visits (per 10), last 4–21 days	0.59	0.15 to 2.35	0.45	
Number of visits to movie theatres (per 10), last 4–21 days	2.17	1.10 to 4.27	0.025	
Number of visits to restaurants (per 10), last 4–21 days	2.06	1.57 to 2.70	<0.001	

Models were adjusted for age, sex, race/ethnicity and calendar date.
*Overall heterogeneity.
†Heterogeneity of non-reference levels.
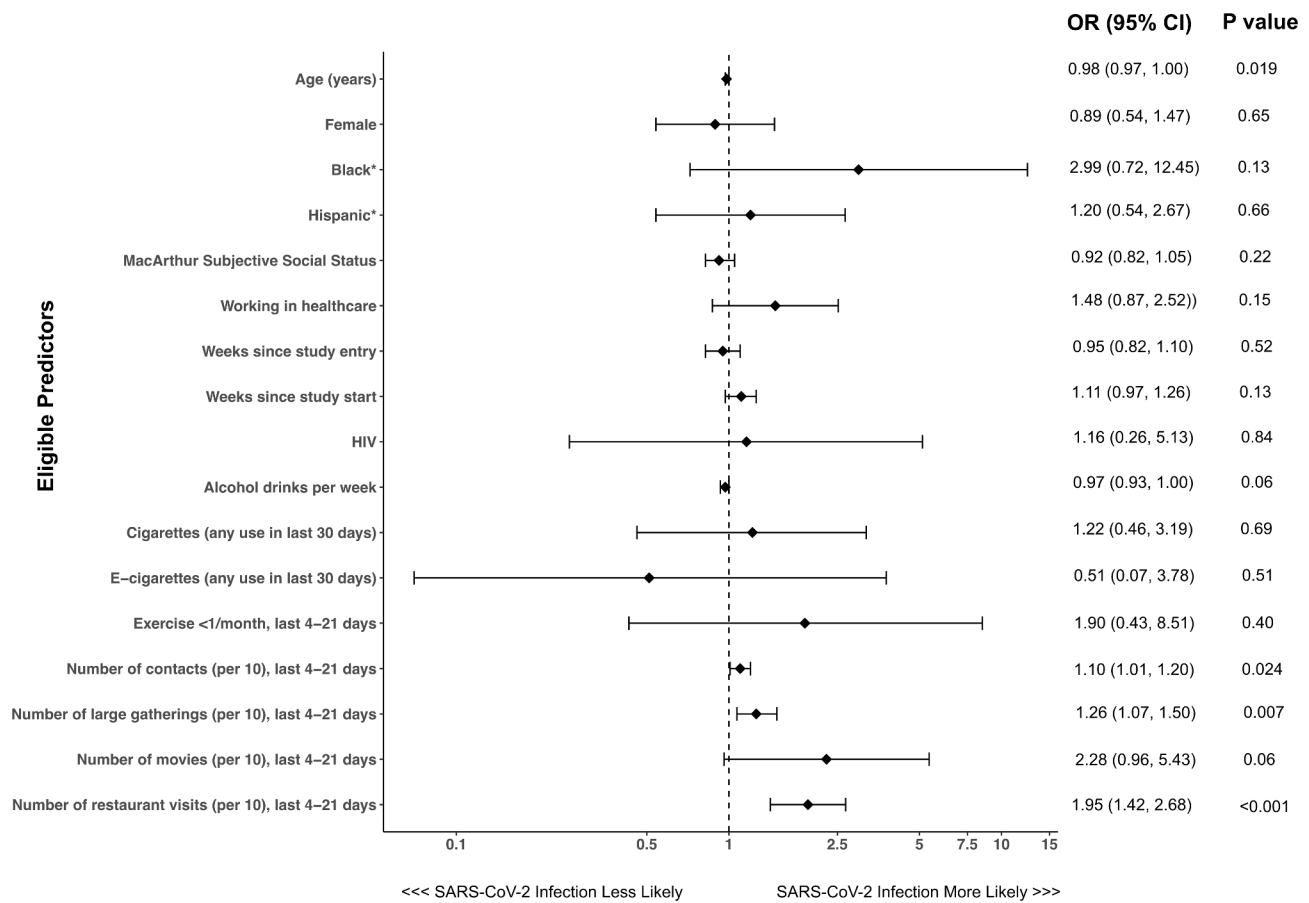‡Linear trend.
COPD, chronic obstructive pulmonary disease.

data were ascertained prospectively and time-updated daily and weekly to minimise recall bias, and self-report remains likely the most effective method to ascertain individual-level behaviours. As the study required smartphone ownership and use, it is possible that the COVID-19 Citizen Science Study participants represent a more affluent and more technologically savvy population compared with the general population. Although this would limit generalisability instead of internal validity, our diverse recruitment methods were meant to mitigate risks of sampling bias. The distribution of study participants throughout nearly 100 different countries and every state in the USA provides fairly unprecedented geographical diversity for a study that also ascertains participant-reported behaviours. There are an innumerable number of behaviours that could have been asked on surveys; we limited our questioning to behaviours previously identified by national and international health organisations and/or those with some biological plausibility as effective means of prevention, such as social distancing, handwashing and use of face masks. While PCR testing for SARS-CoV-2 relies on a participant's development of symptoms, index of suspicion and available access to a testing facility, all factors that may have led to under-reporting of all SARS-CoV-2 infections in the study population, the use of these tests to identify SARS-CoV-2 infections ensured that our analyses identified risk factors associated with true infection and increased specificity over traditional methods of symptom reporting. Because identification of predictors was determined by testing for statistical significance, we acknowledge that the effect sizes for some of the identified covariates may be small and of questionable clinical relevance. However, this approach enabled us to be as inclusive as possible without constraining potentially relevant predictors based on preconceived assumptions. Finally, all data in the COVID-19 Citizen Science Study were collected prospectively as an observational study. While this allows for diverse and rapid sampling of a large population to inform global efforts combating the SARS-CoV-2 pandemic, it remains prone to residual and unmeasured confounding.

In conclusion, the COVID-19 Citizen Science Study, in its prospective and time-updated collection of health data, has identified readily modifiable behaviours that may increase one's individual risk of contracting SARS-CoV-2. Increased number of contacts within 1.83 meters (6 feet), events of 10 or more people and visits to restaurants each independently predicted higher risk of contracting SARS-CoV-2 during the pandemic, while one's demographics, prior medical comorbidities, and adherence to handwashing and face mask wearing were not significant predictors of SARS-CoV-2. During a resurgence of SARS-CoV-2 and continued strain on local governments to balance transmission risk with restrictions on daily life, our study

Predictors of SARS-CoV-2 Infection

Eligible Predictors	OR (95% CI)	P value
Age (years)	0.98 (0.97, 1.00)	0.019
Female	0.89 (0.54, 1.47)	0.65
Black*	2.99 (0.72, 12.45)	0.13
Hispanic*	1.20 (0.54, 2.67)	0.66
MacArthur Subjective Social Status	0.92 (0.82, 1.05)	0.22
Working in healthcare	1.48 (0.87, 2.52))	0.15
Weeks since study entry	0.95 (0.82, 1.10)	0.52
Weeks since study start	1.11 (0.97, 1.26)	0.13
HIV	1.16 (0.26, 5.13)	0.84
Alcohol drinks per week	0.97 (0.93, 1.00)	0.06
Cigarettes (any use in last 30 days)	1.22 (0.46, 3.19)	0.69
E–cigarettes (any use in last 30 days)	0.51 (0.07, 3.78)	0.51
Exercise <1/month, last 4–21 days	1.90 (0.43, 8.51)	0.40
Number of contacts (per 10), last 4–21 days	1.10 (1.01, 1.20)	0.024
Number of large gatherings (per 10), last 4–21 days	1.26 (1.07, 1.50)	0.007
Number of movies (per 10), last 4–21 days	2.28 (0.96, 5.43)	0.06
Number of restaurant visits (per 10), last 4–21 days	1.95 (1.42, 2.68)	<0.001

<<< SARS-CoV-2 Infection Less Likely   SARS-CoV-2 Infection More Likely >>>

Figure 2 Forest plot of all eligible predictors in pooled logistic regression models. Higher scores in the MacArthur Subjective Social Status Ladder reflect participants with self-reported higher socioeconomic standing. Large gathering was defined as any gathering in which 10 or more people were present. The reference group for predictors marked with an asterisk (*) was compared with non-Hispanic white participants.

provides community leaders and members of the public with at least three modifiable health behaviours within an individual's control that may lower one's personal risk of contracting SARS-CoV-2 during this pandemic.

ORCID iDs
Anthony Lin http://orcid.org/0000-0001-9548-6194
Eric Vittinghoff http://orcid.org/0000-0001-8535-0920

Noah Peyser http://orcid.org/0000-0003-4529-8325
David Wen http://orcid.org/0000-0003-1986-2164

REFERENCES

1 Gollwitzer A, Martel C, Brady WJ, *et al*. Partisan differences in physical distancing are linked to health outcomes during the COVID-19 pandemic. *Nat Hum Behav* 2020;4:1186–97.
2 Painter M, Qiu T. Political beliefs affect compliance with COVID-19 social distancing orders. *SSRN Electronic Journal* 2020;360.
3 Barrios JM, Hochberg YV. Risk perception through the lens of politics in the time of the COVID-19 pandemic. *SSRN Journal* 2020;24.
4 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.
5 Adorni F, Prinelli F, Bianchi F. Self-reported symptoms of SARS-CoV-2 infection in a non-hospitalized population: results from the large Italian web-based EPICOVID19 cross-sectional survey. *JMIR Public Heal Surveill* 2020.
6 Nguyen LH, Drew DA, Graham MS, *et al*. Risk of COVID-19 among front-line health-care workers and the general community: a prospective cohort study. *Lancet Public Health* 2020;5:e475–83.
7 Salmon Ceron D, Bartier S, Hautefort C, *et al*. Self-reported loss of smell without nasal obstruction to identify COVID-19. The multicenter Coranosmia cohort study. *J Infect* 2020;81:614-620.
8 Chu DK, Akl EA, Duda S, *et al*. Physical distancing, face masks, and eye protection to prevent person-to-person transmission of SARS-CoV-2 and COVID-19: a systematic review and meta-analysis. *Lancet* 2020;395:1973–87.
9 Greenhalgh T, Schmid MB, Czypionka T, *et al*. Face masks for the public during the covid-19 crisis. *BMJ* 2020;369:m1435.
10 Avram R, Tison GH, Aschbacher K, *et al*. Real-World heart rate norms in the health eHeart study. *NPJ Digit Med* 2019;2:58.
11 Christensen MA, Dixit S, Dewland TA, *et al*. Sleep characteristics that predict atrial fibrillation. *Heart Rhythm* 2018;15:1289–95.
12 Dixit S, Pletcher MJ, Vittinghoff E, *et al*. Secondhand smoke and atrial fibrillation: data from the health eHeart study. *Heart Rhythm* 2016;13:3–9.
13 Wang JB, Olgin JE, Nah G, *et al*. Cigarette and e-cigarette dual use and risk of cardiopulmonary symptoms in the health eHeart study. *PLoS One* 2018;13:e0198681–14.
14 Tison GH, Avram R, Kuhar P, *et al*. Worldwide effect of COVID-19 on physical activity: a descriptive study. *Ann Intern Med* 2020;173:767–70.
15 Goodman E, Adler NE, Kawachi I, *et al*. Adolescents' perceptions of social status: development and evaluation of a new indicator. *Pediatrics* 2001;108:e31.
16 Cooper DC, Milic MS, Mills PJ, *et al*. Endothelial function: the impact of objective and subjective socioeconomic status on flow-mediated dilation. *Ann Behav Med* 2010;39:222–31.
17 Olgin JE, Pletcher MJ, Marcus GM. The health eHeart study., 2021. Available: https://www.health-eheartstudy.org/ [Accessed 19 Apr 2021].
18 Patient-Centered Outcomes Research Institute. The heart eHeart alliance, 2021. Available: https://www.pcori.org/research-results/2015/health-eheart-alliance [Accessed 19 Apr 2021].
19 Center for Systems Science and Engineering at Johns Hopkins University. COVID-19 Dashboard, 2020. Available: https://coronavirus.jhu.edu/map.html [Accessed 22 Mar 2021].
20 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;383:1071–7.
21 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;13:1405–14.
22 COVID-19 therapies and vaccine landscape. *Nat Mater* 2020;19:809.
23 Kaur SP, Gupta V. COVID-19 vaccine: a comprehensive status report. *Virus Res* 1981;2020:14.
24 Our World in Data. Coronavirus (COVID-19) vaccinations, 2020. Available: https://ourworldindata.org/covid-vaccinations [Accessed 22 Mar 2021].
25 Mizumoto K, Kagaya K, Zarebski A, *et al*. Estimating the asymptomatic proportion of coronavirus disease 2019 (COVID-19) cases on board the diamond Princess cruise SHIP, Yokohama, Japan, 2020. *Eurosurveillance* 2020;25:1–5.
26 Furukawa NW, Brooks JT, Sobel J. Evidence supporting transmission of severe acute respiratory syndrome coronavirus 2 while presymptomatic or asymptomatic. *Emerg Infect Dis* 2020;26:E1–6.
27 Courtemanche C, Garuccio J, Le A, *et al*. Strong social distancing measures in the United States reduced the COVID-19 growth rate. *Health Aff* 2020;39:1237–46.
28 Hsiang S, Allen D, Annan-Phan S, *et al*. The effect of large-scale anti-contagion policies on the COVID-19 pandemic. *Nature* 2020;584:262–7.
29 Bagchi S, Mak J, Li Q, *et al*. Rates of COVID-19 Among Residents and Staff Members in Nursing Homes - United States, May 25-November 22, 2020. *MMWR Morb Mortal Wkly Rep* 2021;70:52–5.
30 Centers for Disease Control and Prevention. COVID-NET, 2021. Available: https://www.cdc.gov/coronavirus/2019-ncov/covid-data/covidview/index.html#hospitalizations [Accessed 16 Jan 2021].
31 National Center for Health Statistics. Daily updates of Totals by week and state: provisional death counts for coronavirus disease 2019 (COVID-19), 2021. Available: https://www.cdc.gov/nchs/nvss/vsrr/COVID19/index.htm [Accessed 16 Jan 2021].
32 Sanyaolu A, Okorie C, Marinkovic A, *et al*. Comorbidity and its impact on patients with COVID-19. *SN Compr. Clin. Med.* 2020;2:1069–76.
33 Harrison SL, Fazio-Eynullayeva E, Lane DA, *et al*. Comorbidities associated with mortality in 31,461 adults with COVID-19 in the United States: a federated electronic medical record analysis. *PLoS Med* 2020;17:e1003321–11.
34 Lyu W, Wehby GL. Community use of face masks and COVID-19: evidence from a natural experiment of state mandates in the US. *Health Aff* 2020;39:1419–25.
35 Chernozhukov V, Kasahara H, Schrimpf P. Causal impact of masks, policies, behavior on early covid-19 pandemic in the U.S. *J Econom* 2020:1–40.