Since January 2020 Elsevier has created a COVID-19 resource centre with free information in English and Mandarin on the novel coronavirus COVID-19. The COVID-19 resource centre is hosted on Elsevier Connect, the company's public news and information website.

Annals of Epidemiology 67 (2022) 81–100

Contents lists available at ScienceDirect

Annals of Epidemiology

journal homepage: www.annalsofepidemiology.org

Original article

Design of a population-based longitudinal cohort study of SARS-CoV-2 incidence and prevalence among adults in the San Francisco Bay Area

Christina P. Lindan, MD MS[a,b,1,*], Manisha Desai, PhD[c,1], Derek Boothroyd, PhD[c], Timothy Judson, MD, MPH[d,e], Jenna Bollyky, MD[f], Hannah Sample, BS[a], Yingjie Weng, MPH[c], Yuteh Cheng, MS[g,h], Alex Dahlen, PhD[i], Haley Hedlin, PhD[c], Kevin Grumbach, MD[j,e], Jeff Henne, MA[g], Sergio Garcia, MS, MBA[g], Ralph Gonzales, MD MSPH[d], Charles S. Craik, PhD[k], George Rutherford, MD[a,b,2], Yvonne Maldonado, MD[l,2]

[a] *Department of Epidemiology and Biostatistics, University of California San Francisco, San Francisco, CA, USA*
[b] *Institute for Global Health Sciences, University of California San Francisco, San Francisco, CA, USA*
[c] *Quantitative Sciences Unit, School of Medicine, Stanford University, 1070 Arastradero Road, Palo Alto, CA, USA*
[d] *Department of Medicine, University of California San Francisco, San Francisco, CA, USA*
[e] *Office of Population Health, University of California San Francisco, Brisbane, CA, USA*
[f] *Division of Primary Care & Population Health, School of Medicine, Stanford University, Stanford, CA, USA*
[g] *Institute for Scientific Analysis, Alameda, CA, USA*
[h] *The Henne Group, San Francisco, CA, USA*
[i] *The Chan-Zuckerberg Initiative, Redwood City, CA, USA*
[j] *Department of Family and Community Medicine, University of California San Francisco, San Francisco General Hospital, San Francisco, CA, USA*
[k] *Department of Pharmaceutical Chemistry, University of California San Francisco, San Francisco, CA, USA*
[l] *Department of Pediatrics, School of Medicine, Stanford University, CA, USA*

ARTICLE INFO

Article history:
Received 26 May 2021
Revised 16 September 2021
Accepted 2 November 2021
Available online 17 November 2021

Keywords:
SARS-CoV-2
COVID-19
Surveillance
Population-based survey
Probability sample
SARS-CoV-2 antibody
SARS-CoV-2 viral detection

ABSTRACT

Purpose: We describe the design of a longitudinal cohort study to determine SARS-CoV-2 incidence and prevalence among a population-based sample of adults living in six San Francisco Bay Area counties.
Methods: Using an address-based sample, we stratified households by county and by census-tract risk. Risk strata were determined by using regression models to predict infections by geographic area using census-level sociodemographic and health characteristics. We disproportionately sampled high and medium risk strata, which had smaller population sizes, to improve precision of estimates, and calculated a desired sample size of 3400. Participants were primarily recruited by mail and were followed monthly with PCR testing of nasopharyngeal swabs, testing of venous blood samples for antibodies to SARS-CoV-2 spike and nucleocapsid antigens, and testing of the presence of neutralizing antibodies, with completion of questionnaires about socio-demographics and behavior. Estimates of incidence and prevalence will be weighted by county, risk strata and sociodemographic characteristics of non-responders, and will take into account laboratory test performance.

List of abbreviations: 100py, 100 Person-Years; ABS, Address-Based Sampling; ACE-2, Angiotensin-Converting Enzyme 2; ACS, American Community Survey; CBO, Community-Based Organization; CI, Confidence Intervals; CLIA, Clinical Laboratory Improvement Act; CT, Cycle Threshold; ELISA, Enzyme-Linked Immunosorbent Assay; EUA, Emergency Use Authorization; FDA, Food and Drug Administration; HH, Household; IgG, Immunoglobulin-G; LASSO, Least Absolute Shrinkage and Selection Operator; LDT, Laboratory Developed Test; N-Protein, Nucleocapsid Protein; Rt-PCR, Reverse Transcriptase-Polymerase Chain Reaction; RBD, Receptor Binding Domain; SHC, Stanford Health Center; S1, Spike Protein; THG, The Henne Group; UCSF, University of California, San Francisco; US, United States.

Conflicts of interest: All the authors listed below for the submission of the following manuscript declare that they have no financial interests or personal relationships that may be considered potential competing interests, or that have inappropriately influenced their contribution to this study or the paper.

* Corresponding author. Christina Lindan, Department of Epidemiology and Biostatistics, University of California San Francisco, San Francisco, 550 16th Street, San Francisco, CA 94143. Tel.: +1 650-387-4452.
E-mail address: krysia.lindan@ucsf.edu (C.P. Lindan).

[1] Co-first authors.
[2] Co-last authors.

https://doi.org/10.1016/j.annepidem.2021.11.001

Results: We enrolled 3842 adults from August to December 2020, and completed follow-up March 31, 2021. We reached target sample sizes within most strata.
Conclusions: Our stratified random sampling design will allow us to recruit a robust general population cohort of adults to determine the incidence of SARS-CoV-2 infection. Identifying risk strata was unique to the design and will help ensure precise estimates, and high-performance testing for presence of virus and antibodies will enable accurate ascertainment of infections.

Introduction

By the beginning of May 2021, 32.6 million people in the U.S. had been were reported as infected with SARS-CoV-2, of whom 579,634 had died [1]. These numbers under-represent the total burden of infection due to incomplete testing and the lower likelihood of asymptomatic persons coming to clinical attention. Accurate data on the extent of infection, even as vaccines are rolled out, are critical to understanding continued transmission and informing ongoing mitigation efforts.

Numerous cross-sectional studies aimed at determining population-levels of infection have been conducted in the U.S., including in Chicago, New York, Indiana, Georgia, California [2–9], as well as country-wide and internationally [10–18]. Approaches to determining the prevalence of infection have also involved testing of remnant blood samples [19–21] including from dialysis patients [22,23]. The Centers for Disease Control and Prevention estimated an overall prevalence of infection in the U.S. of 14% based on data from community-based studies and the testing of remnant blood specimens from 10 sites nationwide, coupled with multipliers based on case reports [9]. Seroprevalence estimates have varied widely, however, due to differences in sampling approaches, the target population, and the dates during which surveys were implemented [18,24]. In addition, general population estimates may not take into account the higher rates of infection among subgroups; for example, several studies have demonstrated that Latinx communities in the U.S. are more highly affected by the pandemic, likely due to occupational hazard, higher housing density and other factors [25–30]. Errors in prevalence estimates can also occur because of imperfect antibody test performance, which can under- or overestimate actual infections [24,31,32]. This was problematic earlier in the pandemic when rapid tests with poor test performance were used in surveys [33]. A large effort is underway to obtain nationwide estimates of prevalence and incidence by mailing home-testing kits to a household probability sample in the U.S., although results from this study are not yet published [34].

We utilized a robust epidemiologic and statistical design to enroll a representative population-based sample of adults from 6 counties in the San Francisco Bay Area into a longitudinal surveillance cohort. The original aim was to obtain regional estimates of incidence and prevalence of SARS-CoV-2 to assist local public health departments, which at the time of study conception in late March 2020 were grappling to determine the trajectory of the epidemic, to identify communities most at risk and the most effective prevention methods. Additional aims included determining the association of occupation and behaviors with infection rates, the proportion of infections that were asymptomatic, COVID-19 vaccine acceptability, and the presence of circulating viral strains. This paper describes the design and methods used for sampling, enrollment, ascertainment of infection, and analysis. As this paper is focused on methods, we do not describe study results, or the characteristics of the study sample. The project, called TrackCOVID, is a collaboration of the University of California, San Francisco (UCSF), the Stanford University Health Center, and the Zuckerberg San Francisco General Hospital, with support from the local county Departments of Public Health, and funded by the Chan-Zuckerberg Initiative.

Materials and methods

Summary

We describe the design of a longitudinal cohort study used to enroll and follow a population-based sample of adults to determine the incidence and prevalence of SARS-CoV-2 infection in the Bay Area. We used a two-stage stratification sampling scheme, based on an address-based sampling frame. We first sampled by county proportionate to the number of households (HH), and then by census tract risk strata (high, medium, low) within each county. Risk strata were determined by using regression models to predict the number of cases in each census tract. Participants were primarily recruited by mail; only one randomly selected adult from each participating household was enrolled. Participants were followed monthly with SARS-CoV-2 PCR and antibody testing, and with questionnaires. Recruitment by mail began at the end of July 2020, with enrolment between August 2020 and December 2020; follow-up was completed March 31, 2021.

Study population

The total adult population of the six counties in the Bay Area was 5,321,907 based on 2019 census data, and racially and ethnically diverse (20.4% Hispanic, 31.5% Asian, 6.0% Black, and 3.6% mixed or other) [35]. Slightly less than 1/3 (30.1%) of adults were 18–34 years of age, and 19.2% were 65 years of age or older.

The targeted sample size for this study was 3400 adults, based on estimating an incidence of 5.0 cases of infection per 100 person-years (100py) (total width of 95% confidence interval [CI] = 2.2 cases/100py). Figure 1 shows estimates of precision assuming different incidence rates and sample sizes, assuming a mean follow-up time of 6 months. The target study population included persons 18 years of age or older residing in Alameda, Contra Costa, Marin, San Francisco, San Mateo, and Santa Clara counties, who were not living in congregate settings or prisons, and who did not report a prior confirmed SARS-CoV-2 infection at screening.

Stratification. We used a stratified random sampling scheme. We first sampled by county and then by modeled risk strata within each county. Sampling was based on the number of HHs rather than number of adults. The number of HHs to be sampled within each county was determined proportionate to the number of residential HHs that were listed in the Postal Delivery Sequence file of the U.S. Postal Service [36]. (The total number of HHs within the six counties was listed as 2,442,926). We then sampled by census tract risk strata (high, medium, low) within each county; strata were based on regression model-prediction of infection (described below). Given the smaller population size within medium

Table 1
Number of households listed in the US Postal Service Delivery Sequency File, by county and risk strata, and the sampling fraction.

	Risk Strata					
	Low		Medium		High	
County	Households *N*	Sampling fraction	Households *N*	Sampling fraction	Households *N*	Sampling fraction
Alameda	190,570	1.00	316,127	2.16	95,905	4.36
Contra Costa	187,079	1.00	198,771	2.05	28,299	2.62
Marin	39,767	1.00	59,580	2.29	5025	6.78
San Francisco	115,300	1.00	210,431	2.30	52,148	3.70
San Mateo	18,659	1.00	165,903	2.73	90,605	5.42
Santa Clara	347,258	1.00	281,453	1.89	40,046	2.76

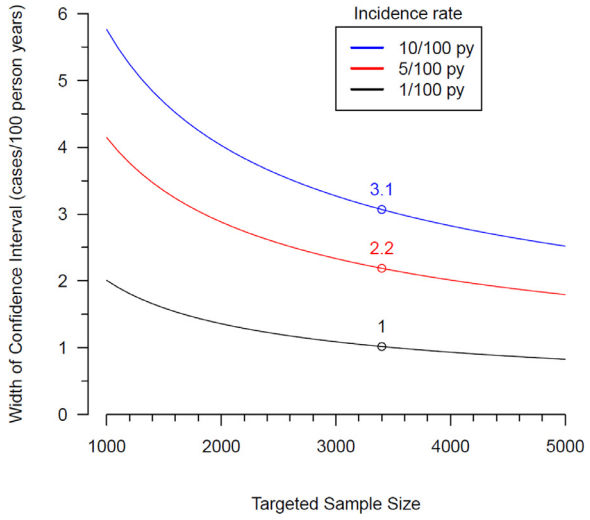

Figure 1. Precision (total width of 95% confidence interval) of estimates of SARS-CoV-2 incidence as a function of sample size and different incidence rates.

and high-risk census tract strata, we oversampled the number of HH adults within those strata, to ensure the precision of our incidence and prevalence estimates (Table 1) [37].

Risk strata were classified based on predicting the number of SARS-CoV-2 cases that could be expected to occur within each census tract. We used predicted number of cases to identify high risk areas, rather than actual cases reported to public health departments, because at the time of study initiation, widespread access to testing was not available, particularly for communities at higher risk. In addition, residents of some communities were often hesitant to seek testing, regardless of availability [25,28]. Therefore, reported infections would not reflect the actual prevalence of infection by geographic area. Likewise, we did not rely upon reported hospitalizations and/or deaths to identify high risk areas. In the Bay Area, the majority of hospitalizations were among LatinX and Black persons, disproportionate to their representation in the general population [38,39]. Thus, reasons for increased morbidity and mortality in these populations were not only related to prevalence of infection, but to co-morbid and other conditions contributing to more severe disease. Therefore, relying on hospitalizations might overestimate the levels of infection within communities.

We classified census tracts into risk strata based on predicting the number of infections using Least Absolute Shrinkage and Selection Operator (LASSO) regression [40]. Factors potentially predictive of SARS-CoV-2 risk were selected based on existing knowledge of socioeconomic and health characteristics among persons more likely to be infected. The distribution of these factors by census-tract was abstracted from data reported in the 2018 American Community Survey (ACS) [41] and the UCSF *HealthAtlas* [42]. We initially included 66 census-level characteristics in the model and from these, identified 27 with the highest coefficients for predicting the cumulative numbers of cases reported by census tract and provided to us by county health departments (model R^2=0.50) (Appendix A). We then applied the model using these selected factors to predict the number of SARS-CoV-2 cases that would exist within each census tract. Based on the Cochrane method [37], census tracts were grouped into strata according to the predicted cases per 100,000 adult population: high risk (>457 cases/100,000 adults), medium risk (114–457 cases/100,000 adults), and low risk (<114 cases/100,000 adults).

Recruitment and enrolment

To determine the number of households that we needed to target for recruitment, we assumed that response to recruitment letters would be 9% in low-risk, 6% in medium risk, and 4% in high-risk strata, based on previous experience with mail-based recruitment. Using these response rates and the desired sample size by strata, we determined the number of HHs to be targeted. We then purchased a stratified random sample of 60,000 HH addresses derived from the US Postal Delivery File, obtained through the Marketing Systems Group (Horsham, PA) (www.m-s-g.com). Participants were primarily recruited by mail, starting in mid-July 2020. We developed letters that described the study and encouraged enrollment, and translated them into the most prevalent languages spoken in the Bay Area (English, Spanish, Chinese, Tagalog and Vietnamese). We also developed postcards in English and Spanish. Letters and postcards were mailed in successive waves to households, with each being sent at least two letters and a postcard. We monitored response by zipcode and strata, and sent additional mailers to HHs in areas where enrolment was low.

Mailers invited the adult with the next birthday to participate, and provided a link to a study -specific website that provided more detailed information and instructions on how to enroll (trackcovidbayarea.com). Mailers listed a telephone number at a health survey research firm employed to assist with the study (The Henne Group [THG] www.thehennegroup.com), which potential participants could call to speak with someone directly. Staff fluent in five languages were available to answer questions and help with screening and enrolment.

We also attempted telephone recruitment. About one-third to one-half of addresses obtained through the Postal Delivery File are linked to telephone numbers. Starting in September 2020, THG began phoning all HHs in our target sample that had an associated phone number. Telephone recruitment was conducted in the preferred language of the prospective participant. Starting in October 2020, to increase response from residents of high-risk strata, we collaborated with local community-based organizations (CBO)

working in the 6 participating counties, as well as a survey team that visited selected households to directly encourage enrollment. Prior to being deployed, teams were trained in how to guide individuals through online enrollment and scheduling; team members were bi-lingual in Spanish and English. Outreach staff also provided printed information about the study, and a gift bag with hand sanitizer and cloth masks; these items were given directly to an adult in the HH, or left outside homes at which no-one answered.

Potential participants could be screened and complete an electronic consent form directly on the study website, verbally through THG on the phone, or at their first visit. They could also schedule their first visit for testing either online or by telephone.

Laboratory testing

Enrolled participants provided samples for viral detection and for antibody testing at one of 13 testing sites that were set up throughout the 6 counties for the purposes of the study. These were co-located at existing testing sites affiliated with UCSF, Stanford Health Center, county public health departments, CBOs, and private hospitals. Sites were supported either by the UCSF or Stanford study teams. All testing platforms had received FDA emergency use authorization (EUA). We performed reverse-transcriptase polymerase chain reaction (rt-PCR) testing of nasopharyngeal swab (NP) samples to identify the presence of virus, indicative of active infection, persistent shedding, or presence of viral particles [43]. We also obtained venous blood samples for testing of antibodies to different viral antigens. Details of testing platforms and performance are provided in Appendix B.

Briefly, PCR testing of swab samples was performed using several different testing platforms, depending on whether tests were performed at the Chan-Zuckerberg BioHub, San Francisco [44], UCSF [45–47] or at Stanford Health Center laboratories [48,49]. Positive PCR samples were sent to the BioHub for whole genome sequencing [50,51].

Plasma from venous blood samples collected at UCSF-supported sites was tested for the presence of IgG antibodies to SARS CoV-2 nucleocapsid (N)-protein [52]. Blood collected at Stanford-supported sites was tested for presence of IgG antibodies to SARS-CoV-2 Spike glycoprotein (S1) and the S1 Receptor Binding Domain (RBD) [49]. Positive antibody samples were cross-tested for presence of IgG at both institutions using the above methods. All samples positive for IgG to either S1, N, or both, were assayed for the presence of neutralizing antibodies at UCSF or the Vitalant Research Institute [22]. Remnant plasma and NP eluent samples from each visit are being stored at specimen banks for confirmation testing if needed and for future research.

All participants were required to register with the electronic health record system of either UCSF or Stanford, to enable processing and reporting of laboratory tests; positive PCR test results were automatically reported to California's electronic disease reporting system [53]. Persons who had a positive PCR or a confirmed antibody test were contacted through their electronic health record system, and were also called by a study physician who counseled them on isolation guidelines, and referred them as necessary for health care and/or support services.

Questionnaire

At baseline, participants completed a detailed questionnaire (Appendix C). Socio-demographic questions included gender identity, age, race, ethnicity, education, income, occupation, household size, and numbers of hours/week working outside the home. Behaviors potentially related to the risk of infection were addressed by asking questions about the proportion of time wearing a mask outside the home in the last month, level of avoidance of people not in the home, travel outside the state, and any known exposure to someone with COVID-19. We also asked about COVID-related symptoms in the previous month and in the last 24 hours, and chronic health conditions including diabetes, obesity, immunologic compromise, among others. Questions about occupation were asked according to the Council of State and Territorial Epidemiologists Occupation Health Subcommittee recommendations [54]. Starting in December 2020, supplemental questions were added inquiring about receipt of COVID-19 vaccination, including date(s) and type of vaccine. The questionnaires were available in the five targeted languages, could be completed electronically through the study website, by phone through THG, or at a testing site with assistance from study staff.

Reimbursement

A $25 gift card was provided as reimbursement at each visit, with a one-time increase to $100 in November and December 2020 to boost enrollment and improve retention. Assistance with the cost of transportation was provided as requested.

Follow-up visits

Participants were followed monthly and were asked to complete a short questionnaire about behavior, symptoms in the last month, exposure to someone with COVID-19, and any change in health status. An NP swab and venous blood samples were also obtained. COVID-19 vaccinations were rolled out in California in a staggered fashion beginning in late December 2020. We continued to follow vaccinated individuals with PCR and antibody testing to identify vaccine breakthrough infections.

COVID-19 protections

We instituted precautions to reduce the risk of SARS-CoV-2 transmission for participants and study staff. All participants were asked to wear a face mask when arriving at the testing site, except for when an NP swab sample was being obtained. Staff collecting NP and/or venous blood specimens wore face masks, eye shields, gloves and gowns; gloves were changed between participants. Hand sanitizer was available. Most of the testing sites were outside under tents and therefore with adequate ventilation. So that participants could avoid public transportation, reimbursement was provided for travel and/or parking as requested.

Primary outcomes and statistical analysis

Our primary outcomes are prevalence and incidence of SARS-CoV-2 infection. A prevalent case is defined as someone who had either a positive PCR test and/or a confirmed antibody test at their baseline visit. A confirmed positive antibody test indicative of infection is defined as having at least 2 of 3 antibodies detected (anti-S1 anti-N, or neutralizing). An incident case is defined as someone who has a positive PCR test or a confirmed antibody test without evidence of infection at baseline or the prior visit. An infection in a vaccinated or partially vaccinated person is defined as having a positive PCR test, and/or a positive anti-nucleocapsid antibody test. Anti-spike and neutralizing antibodies can be generated in response to the vaccine and were therefore were not considered to confirm a true infection [55].

We will use a weighted binomial approach to estimate baseline prevalence with 95% CI. To estimate incidence (new SARS CoV-2 infections/100py), we will use weighted Poisson regression with person-days in the model as an offset. Persons who have evidence of a prevalent infection at their baseline visit will not be included

Table 2
Enrolment: desired sample size (SS), and the number and proportion of participants enrolled, by county and census tract risk strata.

	Census Tract Risk Strata							
	Low		Medium		High		Total	
County	SS	Enrolled *N* (%)	SS	Enrolledd *N* (%)	SS	Enrolledd *N* (%)	SS	Enrolledd *N* (%)
Alameda*	116	116 (100%)	421	521 (124%)	261	263 (101%)	798	900 (113%)
Contra Costa	152	159 (105%)	334	271 (81%)	61	37 (61%)	547	467 (85%)
Marin	57	61 (107%)	194	295 (152%)	49	76 (155%)	300	432 (144%)
San Francisco	66	74 (112%)	307	407 (133%)	130	185 (142%)	503	666 (132%)
San Mateo	7	10 (143%)	171	249 (146%)	189	251 (133%)	367	510 (139%)
Santa Clara	304	293 (96%)	475	481 (101%)	106	93 (88%)	885	867 (98%)
All	702	713 (102%)	1902	2224 (117%)	796	905 (114%)	3400	3842 (113%)

* Includes the City of Berkeley, which has its own Department of Health.

in incidence calculations. For participants with previously negative test results, and who have a confirmed antibody test on a follow-up visit without a positive PCR test, the date of infection will be imputed as the mid-point between the last negative test and the first positive antibody test. Individuals will be censored if they meet the definition of a new infection, die, withdraw from the study, or are lost to follow-up

Weights will be estimated to account for stratification, the probability of being selected based on the number of adults in the household, and differential non-response and coverage by age, education, gender, and race/ethnicity [56]. The latter relies on raking methods [57] that will be applied after determining key differences in socio-demographic characteristics between the weighted sample and the general population based on 2019 ACS data [41,58]. The standard error used in the confidence interval estimates will be obtained via bootstrapping procedures to account for uncertainty of sample size, weight estimation, and positive percent agreement (PPA) and negative percent agreement (NPA) of testing platforms.

Sensitivity analysis

We will calculate estimates of incidence and prevalence excluding persons after a first dose of a COVID-19 vaccine, and also calculate them including vaccinated individuals; the latter will provide an estimate of the general incidence in the population during vaccine uptake. Estimates of prevalence will be adjusted for the laboratory assay performance (using bootstrapping methods according to Sempos and Tian [59]). These methods will also be applied to incidence estimates. We will use bootstrapping to estimate the variance of incidence and prevalence estimates to account for the uncertainty of weight estimation and sample size.

Ethical considerations

The study was reviewed and classified as public health surveillance by both the UCSF Office of Human Research Protections and the Stanford Medical Center institutional review board, based on the definition of surveillance in the US 2018 Revised Common Rule (45 CFR 46.102(l)(2). Official support from and engagement with the local county health departments was obtained. Participants signed separate consent forms for inclusion in the main study and for banking of remnant samples for future testing. Participants indicated at enrolment whether or not they were willing to be contacted for recruitment into future studies.

Results

We enrolled 3842 participants, continuing recruitment beyond our desired sample size of 3400 to ensure adequate numbers of enrolled adults from high-risk strata. Comparison of the desired sample sizes by county and census tract strata, and actual numbers of enrolled participants, is shown in Table 2. Overall, we enrolled the desired number of participants except from Contra Costa County, due to delays in setting up testing sites. The proportion enrolled from high-risk strata in Santa Clara (88%) was also slightly lower than desired.

The response rate, or the proportion of HHs from which a participant was enrolled, from among the number of targeted HHs, is shown in Table 3. Our overall response rate was 6%– 9% from low-risk, 7% from medium-risk, and 4% from high-risk strata. Retention at the five-month follow-up visit, meaning completion of the questionnaire as well as providing specimens for testing (NP swab and venous blood) was 86.6%. All participants who completed a follow-up survey also agreed to be tested.

THG attempted phone calls to 21,918 residences for which we had associated telephone numbers (36.5% of the 60,000 HH sample). Among the 9258 persons who were reached, 1390 (15.0%) were not associated with the address listed in the sample, 6196 (66.9%) refused participation, and 1014 (10.9%) enrolled on the phone or on the website. Among those who refused, 2095 (33.8%) hung up the phone before indicating why they were not interested, 1413 (18.4%) said they didn't want to participate in a study, and 1697 (27.4%) did not provide a reason. Only 135 refused because they didn't want to be tested; 200 were uninterested because they felt the study required too much time.

CBOs and a survey team approached 1590 HHs in selected high risk census tracts in 5 of the 6 counties, from which 119 (7.5%) eligible adults were enrolled at the time of canvassing. This is nearly twice the response to mailers from persons in high-risk strata, and is likely an underestimate of response, as we could not track the number of persons from these HHs who decided to enroll later.

Discussion

We designed and implemented a longitudinal cohort study to recruit a probability sample of adults in the San Francisco Bay Area to estimate the population-level incidence and prevalence of SARS-CoV-2 infection. One of the main strengths of the study was the use of stratified random sampling that relied on an address-based sampling frame. The U.S. Postal Service Delivery Sequence File provides a nearly complete list of all addresses in the country, and its use in defining our sampling frame will reduce bias compared to other non-representative, but easier to implement, sampling schemes. In our study, we did not enroll persons without housing, and excluded those living in nursing homes, homeless shelters and prisons, where rates of infection were extremely high [25,60–64]. Thus, our estimates will not represent infection among these groups.

We used regression models to predict the number of infections within census tracts, as a means of identifying strata for sampling,

Table 3
Response rate: number of households targeted for recruitment and the response (number and proportion of participants enrolled), by county and census tract risk strata

	Risk Strata							
	Low		Medium		High		Total	
County	Households targeted *N*	Enrolled *N* (%)	Households targeted *N*	Enrolled *N* (%)	Households targeted *N*	Enrolled *N* (%)	Households targeted *N*	Enrolled *N* (%)
Alameda	1300	116 (9%)	7000	521 (7%)	6500	263 (4%)	14,800	900 (6%)
Contra Costa	1700	159 (9%)	5600	271 (5%)	1600	37 (2%)	8900	467 (5%)
Marin	700	61 (9%)	3300	295 (9%)	1300	76 (6%)	5300	432 (8%)
San Francisco	1019	74 (7%)	5329	407 (8%)	3247	185 (6%)	9595	666 (7%)
San Mateo	159	10 (6%)	3080	249 (8%)	4678	251 (5%)	7917	510 (6%)
Santa Clara	3122	293 (9%)	7491	481 (6%)	2875	93 (3%)	13,488	867 (6%)
All Counties	8000	713 (9%)	31,800	2224 (7%)	20,200	905 (4%)	60,000	3842 (6%)

which was a unique feature of our study. The goal of using models was not to estimate the prevalence or incidence in particular regions, but rather to identify correlates of infection to categorize strata such that a weighted stratified sample would provide more precise estimates than a strictly random sample. Predicting cases, or the 'risk' within geographic areas, had the advantage of not being sensitive to short-term fluctuations in the local pandemic, such as a contained outbreak of infections. On the other hand, predicted risk strata would not reflect overall shifts in infection rates among different communities as the pandemic expanded. Use of a prediction model was based on several assumptions, however, one of which was that the risk level of all HH adults living within a census tract was the same. To evaluate this, we estimated precision based on different probabilities of misclassifying HH risk, and confirmed that even with moderate misclassification, stratification would improve precision. We also assumed that the risk, or at least the comparative risk between strata, would remain constant during the study period. Finally, we assumed that the socio-demographic characteristics we included in the model were reflective of risk. Several of the predictors included in the model have empirically been shown to be associated with higher rates of infection, including being LatinX and having low income [25,30,38,39,60].

Our overall goal was to estimate incidence and prevalence among the 'general adult population'. The Bay Area, however, is highly heterogeneous, and includes many first- and second-generation immigrants from around the globe. Due to logistical constraints and available funding, the study was not designed to determine the incidence or prevalence by risk strata, county, or race/ethnicity with precision; therefore, outcome estimates will reflect an average across communities. We will also not be able to determine precise outcomes at specific points in time; this limits the interpretation of results, as the trajectory of the local pandemic changed during the study period, with a significant surge in reported cases in November and December 2020 [65,66]. Rates of infection have also been influenced by masking, social-distancing requirements, and the roll-out of COVID-19 vaccines.

A limitation of this study, as well as of other similar surveys, is non-response bias. Although weighting can be used to account for socio-demographic differences between the enrolled sample and the general population, the validity of results relies on the assumption that those who respond are similar to those who do not. Evaluating characteristics of non-responders requires reaching and surveying them, which is often impractical. Our overall response rate was 6%, which is what we assumed when designing the study. We also attempted recruitment by phone, but only one-third of HH addresses in our sampling frame were linked to a phone number. And although phone calls increased enrolment slightly, this approach required significant staff effort. Finally, we collaborated with CBOs to increase inclusion of participants from communities with the most barriers to participation. Other study design features, however, may have negatively affected response, such as the requirement to visit sites for sample collection. We tried to reduce this barrier by placing study sites throughout the 6 counties, by including reimbursement for transportation, and by making evening and weekend appointments available. We also increased reimbursement from $25 to $100 during the last 2 months of enrolment. The combination of these methods allowed us to reach our sample size goals. Despite these efforts, determining the sampling scheme, and recruiting and following a population-based cohort were logistically complicated, time intensive and costly. We began designing the study in April 2020, and recruited our first participants in August of that year. Enrolment of the cohort itself took 5 months, which was longer than we anticipated.

An additional strength of our study was the use of multiple antibody tests and viral PCR detection which will increase our ability to identify SARS-CoV-2 infections. We used tests able to detect antibodies to both nucleocapsid and spike RBD proteins, and positive samples were additionally evaluated for neutralizing and ACE-2 receptor-binding antibodies [48]. A variety of antibodies can be generated in response to SARS-CoV-2 that may not be detected by testing for antibodies to only one antigen [67,68]. In addition, the antibody tests we used were of relatively high test performance so that even with a low population prevalence, the likelihood that a positive test indicated a true infection will be improved, and the possibility of missing an infection reduced. Because COVID-19 vaccination began at the end of the enrolment period and during follow-up, tests that detect the presence of anti-nucleocapsid antibodies can help identify vaccine breakthrough infections or those that occur before vaccine immunity has developed, whereas antibodies to spike-protein can develop in response to immunization and therefore may not indicate a true infection [55,69]. Viral detection in combination with antibody testing and monthly specimen collection will allow us to assess the relationship of antibody production to viral shedding, the frequency of asymptomatic infections, and short-term persistence of antibodies. Neutralizing antibody tests provide additional information about humoral immunity in response to infection [22].

Although undergoing repeated NP swabs and venous blood draws can be uncomfortable, we chose these sample collection methods because at the start of the study, PCR testing of other specimens (such as anterior nasal swabs), as well as rapid tests for antigen and antibody detection had not been fully developed [70,71]. Since then, rapid antigen testing and PCR testing of self-collected nasal swabs [72] and saliva [73] have been shown to be fairly accurate, and are being used in various settings; additionally, some studies are using finger-prick capillary blood samples [34,74] to test for presence of antibodies. Self-collection of samples at our testing sites or by using mail-in home test kits would likely have

increased response. However, we decided not to change our testing platforms and algorithm midway through the study, to avoid accounting for potential differences in test performance. And despite the discomfort of testing, those who enrolled in the study continued with follow-up, enhanced by the personal and ongoing interaction with site staff, including physicians and nurses, and telephone calls from staff or THG whenever a participant missed a visit.

Assembly of a longitudinal general population cohort such as TrackCOVID can be used as a platform for evaluating a variety of questions. Almost all participants agreed to be contacted for further studies. We administered a supplemental questionnaire in December 2020, just prior to vaccine roll-out, that inquired about attitudes, beliefs, and willingness to receive a COVID-19 immunization[75]. The response was high and results indicated disparities in vaccine intention by race/ethnicity, even among persons working in health care. In addition, participants are being enrolled in a follow-up study to identify breakthrough infections among those who have been vaccinated, and re-infections among whose had previously had COVID-19.

One of the aims of the study was to inform and collaborate with the public health departments in participating counties. We developed a real-time dashboard of study results that was available to counties [76]. The dashboard contained information on study recruitment, incidence and prevalence of infection, retention, sociodemographic characteristics of infected participants compared to the overall cohort, and vaccine uptake. Data were presented for the overall cohort as well as by county. Monthly meetings with county health departments were instituted to obtain their feedback and share information that could inform policy.

Conclusions

The design of this study can provide guidance for other surveys, while acknowledging the inherent difficulties in recruiting a population-based sample and the restrictions on interpretation of results. The project was enabled by collaboration with public health departments that were significantly invested in our findings and provided ongoing resources and feedback during study planning and implementation. Overall, designing and implementing a study to enroll a representative sample of the general population is challenging and requires a strong multi-disciplinary team. Employing multiple methods of recruitment, including through involvement of CBOs trusted by the local population, can also be helpful.

Acknowledgments

We would like to acknowledge the following for their invaluable assistance with the study: all of the study participants; the Departments of Public Health in the counties of San Francisco, Contra Costa, Alameda, Marin, San Mateo and Santa Clara, and in the City of Berkeley; the TrackCOVID clinical research coordinators; the staff of the The Henne Group; Mansour Fahimi, PhD, of the Marketing Systems Group for assistance with determining the sampling frame; community-based organizations for their help with recruitment (the Community Health Partnership, the Canal Alliance, and The Bay Area Community Health Advisory Council); Joe DeRisi, PhD and the Chan-Zuckerberg BioHub for PCR testing, viral sequencing and storage of samples; the clinical research lab at the Stanford Health Center; the UCSF Clinical Laboratory; Jing Jin and Graham Simmons (Vitalant Research Institute), and Rodolfo Villa (UCSF) for assistance with pseudovirus neutralization assays; David Glidden, PhD, UCSF for assistance with initial study design; Sara Covin, UCSF for assistance with references.

This study was funded by a grant from the Chan-Zuckerberg Initiative (CZI). The content of this article is solely the responsibility of the authors. CZI was not involved in the study design, the collection, analysis or reporting of results, the writing of the manuscript or the decision to submit the article for publication.

Appendix A

Socio-demographic variables considered for inclusion in the LASSO Regression model, and coefficients of variables included in the final model, used to predict SARS-CoV-2 cases within census tracts.

Family	Variable Name	LASSO coef.
Race/Ethnicity		
	% Hispanic (overall)	0.23
	% Central American	0.08
	% Mexican	0.02
	% Black	
	% Foreign-born	0.05
	% Native American	-0.01
	% Native Hawaiian/Pacific Islander	0.04
	% Southeast Asian	
	% South American	0.02
	% Asian (overall)	
	% East Asian	-0.04
	% South Asian	-0.14
	% White, non-Hispanic	-0.12
Age / Gender		
	% 18 - 40 years old	-0.02
	% Male	0.02
	% < 5 years of age	
	% < 18 years old	
	% Households with a resident younger than 18	
	% Households with a resident older that 65	
	% > 65 years old	
Education		
	% Less than high school	
	% College-educated	
Socio-Economic		
	Teen-birth rate (% of women who gave birth before 20)*	0.07
	% Households with more occupants than bedrooms	
	% Households on food stamp / SNAP benefits in the last year	
	% Households without internet	
	% Households classified as “extremely low income” (making less than 30% of the HUD Area Median Family Income)*	
	% Not fluent in English	
	% Households that spend > 50% of income on rent*	
	% Households that are single-family homes	
	% Households earning below 1.25 the poverty line	0.06
	Incarceration rate* (% of children who grew up in this census tract who were in jail on April 1, 2010)	0.01
	% Households without vehicle access	
	Overall population density (per square mile)	
	Average number of occupants/household	
	Unemployment rate (% of 16+ population without a job)	
	Food desert (binary variable: is there grocery store access within 0.5 miles for urban areas and 10 miles for rural ones?)*	-0.03
	Eviction-filing rate* (% of renter occupied housing units that have evictions filed)	-0.02
	Gini index (measure of income inequality)	
	Traffic density (vehicle-kms / hour / road length within 150 m of census tract boundary: percentile)*	

Family	Variable Name	LASSO coef.
	% Families moved in the last year	-0.02
	% with limited public transit (no stops within half a mile)*	
	% Households that own (vs rent)	
	Median rent	
	Median house price	0.01
	Median household income	0.06
Job		
	% Employed population working in service	
	% Employed population working in production / transportation	0.04
	% Employed population working in construction / natural resources	
	% Employed population working in sales / office work	
	% Employed population working in military	
	% Employed population working in management	
Commute		
	% Commute by carpool	
	% Commute by public transit	
	% Commute by bike or walk	
	% Commute lasts <15min	-0.02
	% Commute lasts > 1hr	
	Avg commute time	-0.07
	% Commute by car (solo)	
	% Work from home	-0.03
Health		
	% Without health insurance	
	ER visits for asthma/capita*	0.11
	% Adults with poor physical health*	0.06
	% Population with a disability*	0.04
	% Adults with poor mental health*	
	% Adults who get annual checkup*	

*Data for variable obtained from the UCSF *Health Atlas* (36). Data for all other variables were taken from the ACS 2018 (37).

Appendix B. Description of laboratory assays

rt-PCR assays

Viral detection was performed by reverse transcriptase polymerase chain reaction (rt-PCR) on eluent from nasopharyngeal swab samples. Samples were collected in RNA/DNA shield, viral transport media, or phosphate-buffered saline depending on the assay to be used.

NP swabs collected at UCSF-supported sites were processed at the UCSF Clinical Microbiology Laboratory and eluent tested using the M2000 Abbott RealTime Sars-CoV-2 assay [45] amplifying the RdRP and N genes (positive cycle threshold [Ct] value $\leq$31.5) [46-47], or the Luminex NxTag assay (Hayward, California) amplifying the N, Orf1ab and E genes [47]. The positive percent agreement (PPA) and the negative percent agreement (NPA) for both assays are reported as 100%.

Some samples from UCSF-supported sites were also processed at the Chan-Zuckerberg BioHub, using a CLIA-validated laboratory developed test (LDT) amplifying the N and E genes, with a positive Ct $<$ 40 [44].

Samples collected at Stanford-supported sites were processed using a Stanford Health Center (SHC) Emergency Use Authorization (EUA) LDT amplifying the E gene; tests were considered positive with a Ct value $<$ 40. This test and has been shown to have 100% PPA and 100% NPA with a comparable rt-PCR test [48]. Some samples were tested using the Panther Fusion SARS-CoV-2 assay (Hologic, Massachusetts) [49]. Among symptomatic persons, the PPA for this test was 100%, and the NPA was 100%; among aymtpomatic persons the PPA was 95.5%, and the NPA was 98.9%.

Genome sequencing

Positive PCR samples were sent to the BioHub for whole genome sequencing using the NOVASeq (Illumina, Inc., San Diego, California), analyzed with IDseq (Chan Zuckerberg BioHub, San Francisco, California) [50], and visualized using the COVID Tracker [51].

Serological assays

Venous blood samples were collected in sodium heparin-coated vacutainers and processed at either the UCSF Clinical Microbiology Laboratory or at the Stanford Anatomic Pathology and Clinical Laboratory.

At UCSF, plasma samples were tested for the presence of IgG antibodies to SARS-CoV-2 nucleocapsid (N) protein using the Abbot Architect (Abbott Park, Illinois). When tested against rt-PCR-confirmed positive and negative samples, this method had a 93.8% PPA and a 99.4% NPA.

Samples processed at Stanford were tested for the presence of IgG antibodies to SARS-CoV-2 spike glycoprotein (S1) using the Euroimmun SARS-CoV-2 IgG Enzyme-linked Immunoassay (ELISA) [52] (Lübeck, Germany). When compared against rt-PCR-confirmed positive and negative samples, this assay had an 85.4% PPA, and a 96.7% NPA [52]. Values were considered positive with a signal-to-cutoff ratio greater than 2.5. Samples with a ratio between 0.8 and 2.5 were considered indeterminate and were subsequently tested for the presence of IgG antibodies to SARS-CoV-2 S1 Receptor Binding Domain (RBD) by an SHC LDT run on the Inova ESP600 QuantaLyser 2 (Inova Diagnostics, San Diego, CA). When evaluating using pre-pandemic samples, this test had a 99.75% NPA [49].

Samples with antibodies identified using one institution's assays, were cross-tested for the presence of antibodies at the other institution, using the above methods. All samples positive for IgG to either S1, N, or both, were assayed for the presence of SARS-CoV-2 neutralizing antibodies at UCSF or the Vitalant Research Institute, San Francisco, using a lentivirus-based pseudo-type neutralization assay [22].

Appendix C

Baseline Survey

Please complete the survey below.

Thank you!

Hello, [screening_arm_1][first_name] [screening_arm_1][last_name]. We look forward to seeing you for your appointment soon! Please complete the follow-up survey below as close to your appointment as possible (ideally within 24 hours prior to your visit).	
Date	_____
Demographic Information	
Have you moved out of [screening_arm_1][cntyid] County since you joined the study?	◯ Yes ◯ No ([screening_arm_1][cntyid] County)
What is your current address?	
Street Number	_____
Apartment number (if applicable)	_____
Street Name	_____
City	_____
State (if not California)	_____
Zip Code	_____
Are you currently covered by any of the following types of health insurance or health coverage plans?	◯ Insurance through a current or former employer ◯ Insurance purchased directly from an insurance company ◯ Medicare (for people 65 and older, or people with certain disabilities ◯ MediCal (CA government assistance ◯ TRICARE or other military health care ◯ VA ◯ Indian Health Service ◯ None ◯ Other
Please specify other insurance coverage	_____

Which of the following best describes your current living situation?	◯ Permanent housing with other people (e.g. family, roommates) ◯ Permanent housing alone ◯ Unstable housing (couch surfing; temporarily staying with friends/family) ◯ Group home ◯ Other
Please describe your current living situation	_____
How many separate rooms (e.g. living room, kitchen) are in the home where you live?	_____ (Do NOT include bathrooms, porches, hallways or unfinished basements.)
How many children under the age of 18 currently stay in your household?	_____ (Number)
In the past week, how many of these children have attended in-person school or day care?	_____ (Number)
How many adults between the ages of 18-64 years currently stay in your household (not including yourself)	_____ (Number)
How many adults 65 years or older currently stay in your household (not including yourself)?	_____ (Number)
Are you currently employed?	◯ Yes, full time ◯ Yes, part time ◯ No
Are you self-employed?	◯ Yes ◯ No
What kind of work do you do?	_____ (e.g. registered nurse, janitor, cashier, auto mechanic)
What kind of business or industry do you work in?	_____ (e.g. hospital, elementary school, clothing manufacturing, restaurant)
Are you a primary earner of income for your household?	◯ Yes, the primary earner ◯ Yes, split with one or more other people in the household ◯ An income earner, but not a primary earner ◯ No
Coronavirus Exposure History	
For the following questions, "close contact" means being within 6 feet of a person for more than 15 minutes, or physical contact like hand-shaking, hugging, or kissing, whether or not a mask is worn.	

In the last month, have you had close contact with a person who tested positive for COVID-19?	◯ Yes ◯ No ◯ I don't know/unknown
Which of the following best describes your relationship with the person who tested positive for COVID-19? If more than one contact, answer for the person with whom you have had the most contact.	◯ Living in the same home ◯ Close regular contact (ie working together or frequently spending time together outside of the home) ◯ Rare contact (ie one or two meetings with someone from outside the home) ◯ Had contact with another person who had contact with a known COVID+ individual (secondary contact) ◯ Unknown
In the past month, have you been to a medical facility as a patient?	◯ Yes, I went to an emergency department or stayed overnight in a hospital ◯ Yes, I went to a doctor's office or clinic ◯ No
In the past month, to what extent have you avoided contact with people who live outside of your home?	◯ All of the time ◯ Most of the time. I only leave my home to buy food or other essentials, or to walk/exercise ◯ Some of the time. I have reduced the amount of time I am in public spaces, social gatherings or at work ◯ None of the time

In the past 24 hours, with how many people have you had close contact, not including those inside your household?

Your best estimate is fine.

At work	_____ (Number)
Shopping for groceries and other essentials	_____ (Number)
At social gatherings (including restaurants or bars)	_____ (Number)
Other	_____ (Number)
In the past month, how often have you worn a mask when you left your home?	◯ Always ◯ Most of the time ◯ Sometimes ◯ Never

In the past month, have you worked or volunteered in any of these settings? Check all that apply	☐ Grocery store ☐ Restaurant, hotel ☐ Pharmacy/drug store ☐ Other retail (e.g. pet store, hardware store) ☐ Transportation ☐ Housekeeping/janitorial services ☐ Child day care ☐ Dormitory ☐ School with in-person classes ☐ Senior care ☐ Construction/utilities/maintenance/landscaping ☐ Hair or nail salon ☐ Delivery services (post office or home delivery) ☐ Law-enforcement/firefighter ☐ Gas station ☐ Other essential business ☐ None of these (Select all that apply or select "None of these")
In the past month, have you worked or volunteered in a hospital, medical office, first responder service (e.g. ambulance), or other health care setting?	◯ Yes ◯ No
Which of the following categories best describes your work in healthcare?	◯ Direct patient care involved in intubating or suctioning patient airways ◯ Direct patient care but NOT performing any airway procedures ◯ Ancillary staff with indirect patient contact (e.g. reception, janitorial services, etc.) ◯ Laboratory ◯ Work in healthcare but not in a setting with patients or their biological samples ◯ Other
Please specify other healthcare work	_____
In the past month, have you primarily worked from home?	◯ Yes ◯ No
In the past month, how much time per week did you spend at a work site outside of your home?	◯ < 10 hrs/wk ◯ 10-20 hrs/wk ◯ 21-30 hrs/wk ◯ 31-40 hrs/wk ◯ > 40 hrs/wk
In the past month, have you worked at or visited a long-term care facility, nursing home, assisted living, correctional facility, homeless shelter or dialysis center?	☐ Long-term care facility ☐ Nursing home ☐ Assisted living ☐ Correctional facility ☐ Homeless shelter ☐ Dialysis center ☐ None of these (Select all that apply or select "None of these")

In the past month, have any of the people who you lived with been working in any of these settings? Check all that apply	☐ Grocery store ☐ Restaurant ☐ Hotel ☐ Pharmacy/drug store ☐ Other retail ☐ Landscaping ☐ Transportation ☐ Housekeeping/janitorial services ☐ Child daycare ☐ Dormitory ☐ School with in-person classes ☐ Senior care facility ☐ Construction/utilities/maintenance ☐ Hair or nail salon ☐ Delivery services (post office or home delivery) ☐ Law-enforcement/firefighter ☐ Health care with direct patient contact ☐ Health care without direct patient contact ☐ None of these (Select all that apply or select "None of these")
Do you live with anyone who	☐ Has a weakened immune system due to a medical condition or medication they take ☐ Is pregnant ☐ Is chronically sick with another disease, like heart disease, lung disease or cancer ☐ None of these (Select all that apply or select "None of these")
In the past month, have you spent time in another state?	◯ Yes ◯ No

Please select which state(s). Check all that apply	☐ Alabama ☐ Alaska ☐ Arizona ☐ Arkansas ☐ California ☐ Colorado ☐ Connecticut ☐ Delaware ☐ Florida ☐ Georgia ☐ Hawaii ☐ Idaho ☐ Illinois ☐ Indiana ☐ Iowa ☐ Kansas ☐ Kentucky ☐ Louisiana ☐ Maine ☐ Maryland ☐ Massachusetts ☐ Michigan ☐ Minnesota ☐ Mississippi ☐ Missouri ☐ Montana ☐ Nebraska ☐ Nevada ☐ New Hampshire ☐ New Jersey ☐ New Mexico ☐ New York ☐ North Carolina ☐ North Dakota ☐ Ohio ☐ Oklahoma ☐ Oregon ☐ Pennsylvania ☐ Rhode Island ☐ South Carolina ☐ South Dakota ☐ Tennessee ☐ Texas ☐ Utah ☐ Vermont ☐ Virginia ☐ Washington ☐ West Virginia ☐ Wisconsin ☐ Wyoming
In the past month, have you travelled to another country?	◯ Yes ◯ No
What country/countries have you travelled to in the past month?	_____ (List all)
Medical History	
Have you been diagnosed with COVID-19 and/or tested positive for SARS-CoV-2 since entering the study?	◯ Yes ◯ No

What type of test was the positive COVID-19 test?	☐ Antibody test (usually a blood sample that is processed in a lab) ☐ Viral test (usually a swab or saliva sample that is processed in a lab) ☐ Antigen test (usually a swab or saliva sample that provides rapid results without a lab) ☐ Other/Unknown (If more than one type of test was positive since you started in the study, please select all positive test types)
Please describe other COVID-19 test	_____
What date was the positive antibody test collected?	_____ (OK to give best guess if exact date is unknown)
Please upload a digital copy of your positive antibody test result (if available)	
What date was the positive viral test collected?	_____ (OK to give best guess if exact date is unknown)
Please upload a digital copy of your positive viral test result (if available)	
What date was the positive antigen test collected?	_____ (OK to give best guess if exact date is unknown)
Please upload a digital copy of your positive antigen test result (if available)	
What date was the positive test collected? test type: other/unknown	_____ (OK to give best guess if exact date is unknown)
Please upload a digital copy of your positive test result (if available) test type: other/unknown	(test type: other/unknown)
We have record that you took the first dose of the vaccine.	
We have record that you took the Second dose of the vaccine.	
Have you been given a vaccination for COVID-19 (including enrolling in a clinical trial where you could be randomized to receive a vaccine)?	◯ Yes ◯ No

Which vaccine did you receive?	◯ AztraZeneca ◯ Janssen ◯ Moderna ◯ Novavax ◯ Pfizer ◯ Unknown ◯ Other
What vaccine did you receive?	_____
How many doses of the vaccine have you received?	◯ 1 dose ◯ 2 doses
When did you receive the first dose of this vaccine?	_____
When did you receive the second dose of this vaccine?	_____
In the past month, have you had any of the following symptoms that were new for you? Check all that apply	☐ Fever ☐ Chills ☐ Dry cough ☐ Cough with sputum production (i.e. cough up phlegm) ☐ Shortness of breath or difficulty breathing ☐ Fatigue or feeling more tired than usual ☐ Muscle or joint aches ☐ Headache ☐ Sore throat ☐ Nasal congestion ☐ Runny or stuffy nose ☐ Nausea ☐ Vomiting ☐ Diarrhea ☐ Persistent pain or pressure in your chest ☐ Decreased sense of taste or smell ☐ Rash on hands or feet ☐ Rash elsewhere ☐ Conjunctivitis (pink eye with discharge) ☐ Other symptoms ☐ None of these (Select all that apply or select "None of these")
Please describe other new symptom	_____
Do you know what your temperature was?	◯ Yes ◯ No
What was your highest recorded temperature?	_____ (degrees Fahrenheit)
When did these symptoms start? If you have had multiple symptoms that started at different times, tell us about the most severe one.	_____ (If the exact date is unknown, your best guess is fine.)

How many days did these symptoms last? If you have had multiple symptoms that started at different times, tell us about the most severe one.	_____ (If the exact length is unknown, please give your best guess.)
In the past 24 hours, have you personally experienced any of the following symptoms that are not explained by pre-existing conditions? Check all that apply	☐ Fever ☐ Chills ☐ Dry cough ☐ Cough with sputum production (i.e. cough up phlegm) ☐ Shortness of breath or difficulty breathing ☐ Fatigue or feeling more tired than usual ☐ Muscle or joint aches ☐ Headache ☐ Sore throat ☐ Nasal congestion ☐ Runny or stuffy nose ☐ Nausea ☐ Vomiting ☐ Diarrhea ☐ Persistent pain or pressure in your chest ☐ Decreased sense of taste or smell ☐ Rash on hands or feet ☐ Rash elsewhere ☐ Conjunctivitis (pink eye with discharge) ☐ Other symptoms ☐ None of these (Select all that apply or select "None of these")
Please list any other symptoms you have had during the last 24 hours that are not explained by pre-existing conditions.	_____ (List all)
How long, in days, have you been experiencing these symptoms? If you have had multiple symptoms that started at different times, tell us about the most severe one.	_____ (If the exact number of days is unknown, please give your best guess.)
Do you know what your temperature was?	◯ Yes ◯ No
What was your highest recorded temperature?	_____ (degrees Fahrenheit)
Are you currently pregnant or were you pregnant within the last month?	◯ Yes ◯ No ◯ Unknown

Have you ever been told by a doctor, nurse or other health professional that you have any of the following medical conditions? Check all that apply	☐ Diabetes ☐ Cancer (other than skin cancer) ☐ Heart disease ☐ High blood pressure ☐ Asthma ☐ Chronic lung disease such as COPD or emphysema ☐ Kidney disease on dialysis ☐ Autoimmune disorder such as rheumatoid arthritis or Crohn's disease ☐ Severe obesity (BMI > 40) ☐ Liver disease ☐ Neurologic disease ☐ HIV or AIDS ☐ Organ transplant ☐ None of these (Select all that apply or select "None of these")
In the last month, has anyone come into your home to help you with activities like walking, getting dressed, bathing or eating?	◯ Yes ◯ No
Social History	
Have you ever smoked? (cigarettes, cigars, pipes) Not including vaping	◯ Never smoker ◯ Former smoker ◯ Current smoker ◯ Decline to respond ◯ Unknown
Have you vaped in the last 30 days? (e-cigarettes) Not including marijuana	◯ Yes ◯ No ◯ Decline to respond ◯ Unknown
Which of the following best describes your total combined household income for the past 12 months, before taxes?	◯ Less than $5,000 ◯ $5,000 through $11,999 ◯ $12,000 through $15,999 ◯ $16,000 through $24,999 ◯ $25,000 through $34,999 ◯ $35,000 through $49,999 ◯ $50,000 through $74,999 ◯ $75,000 through $99,999 ◯ $100,000 through $124,999 ◯ $125,000 through $149,999 ◯ $150,000 or more ◯ Don't know ◯ Decline to respond
Sexual Orientation/Identity	◯ Bisexual ◯ Gay/Lesbian/Same-Gender Loving ◯ Questioning/Unsure ◯ Straight/Heterosexual ◯ Other ◯ Decline to respond
Please specify other sexual orientation	_____

References

[1] The New York Times. Coronavirus in the U.S.: Latest Map and Case Count 2021 https://www.nytimes.com/interactive/2021/us/covid-cases.html; Accessed 1.5.2021.

[2] Demonbreun AR, McDade TW, Pesce L, Vaught LA, Reiser NL, Bogdanovic E, et al. Patterns and persistence of SARS-CoV-2 IgG antibodies in Chicago to monitor COVID-19 exposure. JCI Insight [Internet] 2021;6(9). https://insight.jci.org/articles/view/146148. doi:10.1172/jci.insight.146148.

[3] Rosenberg ES, Dufort EM, Blog DS, Hall EW, Hoefer D, Backenson BP, et al. COVID-19 testing, epidemic features, hospital outcomes, and household prevalence, New York State—March 2020. Clin Infect Dis 2020;71(8):1953–9. doi:10.1093/cid/ciaa549.

[4] Menachemi N. Population point prevalence of SARS-CoV-2 infection based on a statewide random sample — Indiana, April 25-29, 2020. MMWR 2020;69:960–4. doi:10.15585/mmwr.mm6929e1.

[5] Biggs HM. Estimated community seroprevalence of SARS-CoV-2 antibodies — two Georgia Counties, April 28-May 3, 2020. MMWR 2020;69:965–70. doi:10.15585/mmwr.mm6929e2.

[6] Bendavid E, Mulaney B, Sood N, Shah S, Ling E, Bromley-Dulfano R, et al. COVID-19 antibody seroprevalence in Santa Clara County, California. Int J Epidemiol 2021;50:410–19. doi:10.1093/ije/dyab010.

[7] Sood N, Simon P, Ebner P, Eichner D, Reynolds J, Bendavid E, et al. Seroprevalence of SARS-CoV-2-specific antibodies among adults in Los Angeles County, California. JAMA 2020;323:2425–7. doi:10.1001/jama.2020.8279.

[8] Appa A, Takahashi S, Rodriguez Barraquer I, Chamie G, Sawyer A, Duarte E, et al. Universal polymerase chain reaction and antibody testing demonstrate little to no transmission of severe acute respiratory syndrome coronavirus 2 in a rural community. Open Forum Infect Dis 2021;8(1):ofaa531. doi:10.1093/ofid/ofaa531.

[9] Angulo FJ, Finelli L, Swerdlow DL. Estimation of US SARS-CoV-2 infections, symptomatic infections, hospitalizations, and deaths using seroprevalence surveys. JAMA Netw Open 2021;4:e2033706. doi:10.1001/jamanetworkopen.2020.33706.

[10] Eckerle I, Meyer B. SARS-CoV-2 seroprevalence in COVID-19 hotspots. Lancet 2020;396:514–15. doi:10.1016/S0140-6736(20)31482-3.

[11] Gudbjartsson DF, Helgason A, Jonsson H, Magnusson OT, Melsted P, Norddahl GL, et al. Spread of SARS-CoV-2 in the Icelandic Population. N Engl J Med 2020;382:2302–15. doi:10.1056/NEJMoa2006100.

[12] Malani A, Shah D, Kang G, Lobo GN, Shastri J, Mohanan M, et al. Seroprevalence of SARS-CoV-2 in slums versus non-slums in Mumbai, India. The Lancet Global Health 2021;9:e110–11. doi:10.1016/S2214-109X(20)30467-8.

[13] Poustchi H, Darvishian M, Mohammadi Z, Shayanrad A, Delavari A, Bahadorimonfared A, et al. SARS-CoV-2 antibody seroprevalence in the general population and high-risk occupational groups across 18 cities in Iran: a population-based cross-sectional study. The Lancet Infectious Diseases 2021;21:473–81. doi:10.1016/S1473-3099(20)30858-6.

[14] Stringhini S, Wisniak A, Piumatti G, Azman AS, Lauer SA, Baysson H, et al. Seroprevalence of anti-SARS-CoV-2 IgG antibodies in Geneva, Switzerland (SEROCoV-POP): a population-based study. The Lancet 2020;396:313–19. doi:10.1016/S0140-6736(20)31304-0.

[15] Buss LF, Prete CA, Abrahim CMM, Mendrone A, Salomon T, de Almeida-Neto C, et al. Three-quarters attack rate of SARS-CoV-2 in the Brazilian Amazon during a largely unmitigated epidemic. Science 2021;371:288–92. doi:10.1126/science.abe9728.

[16] Paulino-Ramirez R, Báez AA, Vallejo Degaudenzi A, Tapia L. Seroprevalence of specific antibodies against SARS-CoV-2 from hotspot communities in the dominican republic. Am J Trop Med Hyg 2020;103:2343–6. doi:10.4269/ajtmh.20-0907.

[17] Santos-Hövener C, Neuhauser HK, Rosario AS, Busch M, Schlaud M, Hoffmann R, et al. Serology- and PCR-based cumulative incidence of SARS-CoV-2 infection in adults in a successfully contained early hotspot (CoMoLo study), Germany, May to June 2020. Euro Surveill 2020;25:1–8. doi:10.2807/1560-7917.ES.2020.25.47.2001752.

[18] Lai CC, Wang JH, Hsueh PR. Population-based seroprevalence surveys of anti-SARS-CoV-2 antibody: an up-to-date review. Int J InfectDis 2020;101:314–22. doi:10.1016/j.ijid.2020.10.011.

[19] US Centers for Disease Control and Prevention.Cases, data, and surveillance. https://www.cdc.gov/coronavirus/2019-ncov/cases-updates/commercial-lab-surveys.html (accessed 21.06.2021).

[20] Basavaraju SV, Patton ME, Grimm K, Rasheed MAU, Lester S, Mills L, et al. Serologic testing of US blood donations to identify severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2)-reactive antibodies: december 2019-January 2020. Clin Infect Dis 2020;72:e1004–9 doi:1093/cid/ciaa1785.

[21] Havers FP, Reed C, Lim T, Montgomery JM, Klena JD, Hall AJ, et al. Seroprevalence of Antibodies to SARS-CoV-2 in 10 Sites in the United States, March 23-May 12, 2020. JAMA Intern Med 2020;180:1576–86. doi:10.1001/jamainternmed.2020.4130.

[22] Ng DL, Goldgof GM, Shy BR, Levine AG, Balcerek J, Bapat SP, et al. SARS-CoV-2 seroprevalence and neutralizing activity in donor and patient blood. Nat Commun 2020;11:4698. doi:10.1038/s41467-020-18468-8.

[23] Anand S, Montez-Rath M, Han J, Bozeman J, Kerschmann R, Beyer P, et al. Prevalence of SARS-CoV-2 antibodies in a large nationwide sample of patients on dialysis in the USA: a cross-sectional study. The Lancet 2020;396:1335–44. doi:10.1016/S0140-6736(20)32009-2.

[24] Flower B, Brown JC, Simmons B, Moshe M, Frise R, Penn R, et al. Clinical and laboratory evaluation of SARS-CoV-2 lateral flow assays for use in a national COVID-19 seroprevalence survey. Thorax 2020;75:1082–8. doi:10.1136/thoraxjnl-2020-215732.

[25] Chamie G, Marquez C, Crawford E, Peng J, Petersen M, Schwab D, et al. community transmission of severe acute respiratory syndrome coronavirus 2 disproportionately affects the latinx population during shelter-in-place in san francisco. Clin Infect Dis 2020;73:S127–35. doi:10.1093/cid/ciaa1234.

[26] Center KE, Da Silva J, Hernandez AL, Vang K, Martin DW, Mazurek J, et al. Multidisciplinary community-based investigation of a COVID-19 outbreak among marshallese and hispanic/Latino Communities — Benton and Washington Counties, Arkansas, March-June 2020. MMWR 2020;69:1807–11. doi:10.15585/mmwr.mm6948a2.

[27] Figueroa JF, Wadhera RK, Mehtsun WT, Riley K, Phelan J, Jha AK. Association of race, ethnicity, and community-level factors with COVID-19 cases and deaths across U.S. counties. Healthcare 2021;9:100495. doi:10.1016/j.hjdsi.2020.100495.

[28] Martinez DA, Hinson JS, Klein EY, Irvin NA, Saheed M, Page KR, et al. SARS-CoV-2 Positivity Rate for Latinos in the Baltimore-Washington, DC Region. JAMA 2020;324(4):392. doi:10.1001/jama.2020.11374.

[29] Podewils LJ, Burket TL, Mettenbrink C, Steiner A, Seidel A, Scott K, et al. Disproportionate Incidence of COVID-19 infection, hospitalizations, and deaths among persons identifying as hispanic or latino — Denver, Colorado March-October 2020. MMWR Morb Mortal Wkly Rep 2020;69:1812–16. doi:10.15585/mmwr.mm6948a3.

[30] Esaryk EE, Wesson P, Fields J, Rios-Fetchko F, Lindan C, Bern C, et al. Variation in SARS-CoV-2 infection risk and socioeconomic disadvantage among a mayan-latinx population in Oakland, California. JAMA Netw Open 2021;4(5):e2110789.. doi:10.1001/jama.2020.11374.

[31] Accorsi E, Qiu X, Rumpler E, Kennedy-Shaffer L, Kahn R, Joshi K, et al. How to detect and reduce potential sources of biases in epidemiologic studies of SARS-CoV-2. https://nrs.harvard.edu/URN-3:HUL.INSTREPOS:3736619210.31219/osf.io/46am5. Eur J Epidemiol 2021;36:179–96. doi:10.31219/osf/io/46am5.

[32] Takahashi S, Greenhouse B, Rodríguez BI. Are seroprevalence estimates for severe acute respiratory syndrome coronavirus 2 biased? J Infect. Dis 2020;222:1772–5. doi:10.1093/infdis/jiaa523.

[33] Vogel G. First antibody surveys draw fire for quality, bias. Science 2020;368:350–1. doi:10.1126/science.368.6489.350.

[34] Siegler AJ, Sullivan PS, Sanchez T, Lopman B, Fahimi M, Sailey C, et al. Protocol for a national probability survey using home specimen collection methods to assess prevalence and incidence of SARS-CoV-2 infection and antibody response. Ann Epidemiol 2020;49:50–60. doi:10.1016/j.annepidem.2020.07.015.

[35] The United States Census Bureau. American Community Survey 1-Year Data (2005-2019). 2020. https://www.census.gov/data/developers/data-sets/acs-1year.html (accessed 10. 05. 2021).

[36] American Association for Public Opinion Research. Address-based Sampling - AAPOR. 2016. https://www.aapor.org/Education-Resources/Reports/Address-based-Sampling.aspx (accessed 21. 05. 2021).

[37] Cochran WG. Sampling techniques. 3rd ed. New York: Wiley; 1977.

[38] Riley AR, Chen YH, Matthay EC, Glymour MM, Torres JM, Fernandez A, et al. Excess mortality among Latino people in California during the COVID-19 pandemic. SSM-Population Health 2021;15:100860. doi:10.1016/j.ssmph.2021.100860.

[39] Rodriguez-Diaz CE, Guilamo-Ramos V, Mena L, Hall E, Honermann B, Crowley JS, et al. Risk for COVID-19 infection and death among Latinos in the United States: examining heterogeneity in transmission dynamics. Ann Epidemiol 2020;52:46–53. doi:10.1016/j.annepidem.2020.07.007.

[40] Tibshirani R. Regression shrinkage and selection via the lasso. J R Stat Soc Series B Methodol 1996;58:267–88. doi:10.1111/j.2517-6161.1996.tb02080.x.

[41] The United States Census Bureau. Data Release New and Notable. 2018. https://www.census.gov/programs-surveys/acs/news/data-releases/2018/release.html (accessed 22.05.2021).

[42] UCSF *HealthAtlas*. https://healthatlas.ucsf.edu (accessed 21.05.2021).

[43] Long DR, Gombar S, Hogan CA, Greninger AL, OReilly SV, Bryson-Cahn C, et al. Occurrence and timing of subsequent SARS-CoV-2 RT-PCR positivity among initially negative patients. Clinical Infectious Diseases 2021;72:323–6. doi:10.1093/cid/ciaa722.

[44] Crawford ED, Acosta I, Ahyong V, Anderson EC, Arevalo S, Asarnow D, et al. Rapid deployment of SARS-CoV-2 testing: the CLIAHUB. PLoS Pathog 2020;16:e1008966. doi:10.1371/journal.ppat.1008966.

[45] Abbott Molecular. Abbott RealTime SARS-CoV-2 2020. https://www.molecular.abbott/sal/9N77-095_SARS-CoV-2_US_EUA_Amp_Pl.pdf (accessed 10.05.2021).

[46] Arnaout R, Lee RA, Lee GR, Callahan C, Yen CF, Smith KP, et al. SARS-CoV2 testing: the limit of detection matters. BioRxiv 2020:131144. doi:10.1101/2020.06.02.131144.

[47] Chen JHK, Yip CCY, Chan JFW, Poon RWS, To KKW, Chan KH, et al. Clinical performance of the Luminex NxTAG CoV extended panel for SARS-CoV-2 detection in nasopharyngeal specimens from COVID-19 patients in Hong Kong. J Clin Microbiol 2020;58:e00936. doi:10.1128/JCM.00936-20.

[48] US Food and Drug Administration. Stanford Health Care clinical virology laboratory SARS-CoV-2 test: EUA summary 2020. https://www.fda.gov/media/136818/download (accessed 10.05.2021).

[49] Röltgen K, Powell AE, Wirz OF, Stevens BA, Hogan CA, Najeeb J, et al. Defining the features and duration of antibody responses to SARS-CoV-2 infection associated with disease severity and outcome. Sci Immunol 2020;5:eabe0240. doi:10.1126/sciimmunol.abe0240.

[50] Kalantar KL, Carvalho T, de Bourcy CFA, Dimitrov B, Dingle G, Egger R, et al. IDseq—An open source cloud-based pipeline and analysis service for metagenomic pathogen detection and monitoring. Gigascience 2020;9:giaa111. doi:10.1093/gigascience/giaa111.
[51] Chan-Zuckerberg Biohub. COVID-19 sequencing for California public health. COVID Tracker Dashboard. https://covidtracker.czbiohub.org/ (accessed 10. 05. 2021).
[52] Tang MS, Hock KG, Logsdon NM, Hayes JE, Gronowski AM, Anderson NW, et al. Clinical performance of two SARS-CoV-2 serologic assays. Clin Chem 2020;66:1055–62. doi:10.1093/clinchem/hvaa120.
[53] California reportable disease information exchange (CalREDIE). California Department of Public Health; 2019. https://www.cdph.ca.gov/Programs/CID/DCDC/Pages/CalREDIE.aspx (accessed 21. 05. 21)..
[54] Council of State and Territorial Epidemiologists Occupational Health Subcommittee. Recommended Interim Guidance for Collecting Employment Information about COVID-19 cases 2020. https://cdn.ymaws.com/www.cste.org/resource/resmgr/publications/Guidance_collecting_io_covid.pdf (accessed 10.05.2021).
[55] Bradley T, Grundberg E, Selvarangan R, LeMaster C, Fraley E, Banerjee D, et al. Antibody responses after a single dose of SARS-CoV-2 mRNA vaccine. N Engl J Med 2021;384:1959–61. doi:10.1056/NEJMc2102051.
[56] Piazza T. Fundamentals of Applied Sampling. In: Marsden PV, Wright JD, editors. Handbook of survey research. Bingley, UK: Emerald; 2010. p. 163–8.
[57] Battaglia MP, Hoaglin DC, Frankel MR. Practical considerations in raking survey data. Surv Pract 2009;2:1–10. doi:10.29115/SP-2009-0019.
[58] Rossi PH, Wright JD, Anderson AB. Sampling theory. handbook of survey research. New York: Academic Press; 1983. p. 125–6.
[59] Sempos CT, Tian L. Adjusting coronavirus prevalence estimates for laboratory test kit error. Am J Epidemiol 2020;190:109–15. doi:10.1093/aje/kwaa174.
[60] Tan AX, Hinman JA, Abdel Magid HS, Nelson LM, Odden MC. Association between income inequality and county-level COVID-19 cases and deaths in the US. JAMA Netw Open 2021;4(5):e218799. doi:10.1001/jamanetworkopen.2021.8799.
[61] Mosites E, Parker EM, Clarke KEN, Gaeta JM, Baggett TP, Imbert E, et al. Assessment of SARS-CoV-2 infection prevalence in homeless shelters — Four U.S. Cities, March 27-April 15, 2020. MMWR 2020;69:521–2. doi:10.15585/mmwr.mm6917e1.
[62] Tobolowsky FA, Gonzales E, Self JL, Rao CY, Keating R, Marx GE, et al. COVID-19 outbreak among three affiliated homeless service sites — King County, Washington, 2020. MMWR 2020;69:523–6. doi:10.15585/mmwr.mm6917e2.
[63] Bagchi S. Rates of COVID-19 among residents and staff members in nursing homes — United States, May 25-November 22, 2020. MMWR 2021;70:1–5. doi:10.15585/mmwr.mm7002e2.
[64] Hershow RB, Segaloff HE, Shockey AC, Florek KR, Murphy SK, DuBose W, et al. Rapid Spread of SARS-CoV-2 in a State prison after introduction by newly transferred incarcerated persons — Wisconsin, August 14-October 22, 2020. MMWR 2021;70:478–82. doi:10.15585/mmwr.mm7013a4.
[65] Los Angeles Times. California coronavirus cases: Tracking the outbreak. https://www.latimes.com/projects/california-coronavirus-cases-tracking-outbreak/ (accessed 10.05.2021).
[66] CDPH news releases. California Department of Public Health; 2020. https://www.cdph.ca.gov/Programs/OPA/Pages/New-Release-2020.aspx (accessed 10. 05. 21)..
[67] US Centers for Disease Control and Prevention. Interim guidelines for COVID-19 antibody testing. 2021. https://www.cdc.gov/coronavirus/2019-ncov/lab/resources/antibody-tests-guidelines.html (accessed 10.05.2021).
[68] Yüce M, Filiztekin E, Özkaya KG. COVID-19 diagnosis —A review of current methods. Biosens Bioelectron 2021;172:112752. doi:10.1016/j.bios.2020.112752.
[69] Moyo-Gwete T, Madzivhandila M, Makhado Z, Ayres F, Mhlanga D, Oosthuysen B, et al. Cross-reactive neutralizing antibody responses elicited by SARS-CoV-2 501Y.V2 (B.1.351). N Engl J Med 2021;384:2161–3. doi:10.1056/NEJMc2104192.
[70] Peeling RW, Olliaro PL, Boeras DI, Fongwen N. Scaling up COVID-19 rapid antigen tests: promises and challenges. Lancet Infect Dis 2021;0:e290–5. doi:10.1016/S1473-3099(21)00048-7.
[71] Pray IW, Ford L, Cole D, Lee C, Bigouette JP, Abedi GR, et al. Performance of an antigen-based test for asymptomatic and symptomatic SARS-CoV-2 testing at two university campuses — Wisconsin, September-October 2020. MMWR 2021;69:1642–7. doi:10.15585/mmwr.mm695152a3.
[72] Nny T, Hc S, Ky N, Bj C, Gm L, Dkm I. Diagnostic performance of different sampling approaches for SARS-CoV-2 RT-PCR testing: a systematic review and meta-analysis. Lancet Infect Dis 2021;21:1233–45. doi:10.1016/s1473-3099(21)00146-8.
[73] Butler-Laporte G, Lawandi A, Schiller I, Yao M, Dendukuri N, McDonald EG, et al. Comparison of saliva and nasopharyngeal swab nucleic acid amplification testing for detection of SARS-CoV-2: a systematic review and meta-analysis. JAMA Intern Med 2021;181:353–60. doi:10.1001/jamainternmed.2020.8876.
[74] Prazuck T, Colin M, Giachè S, Gubavu C, Seve A, Rzepecki V, et al. Evaluation of performance of two SARS-CoV-2 Rapid IgM-IgG combined antibody tests on capillary whole blood samples from the fingertip. PLoS ONE 2020;15:e0237694. doi:10.1371/journal.pone.0237694.
[75] Grumbach K, Judson T, Desai M, Jain V, Lindan C, Doernberg SB, et al. Association of race/ethnicity with likeliness of COVID-19 vaccine uptake among health workers and the general population in the San Francisco Bay Area. JAMA Intern Med 2021;181:1008–11. doi:10.1001/jamainternmed.2021.1445.
[76] TrackCOVID study (General Population). University of California, San Francisco, Zuckerberg San Francisco General, Stanford Medicine; 2021. Google Data Studio http://datastudio.google.com/reporting/352813d0-3021-4b6d-9a9f-5bc64076a810 (accessed 10.05.2021).