Impact of individual-level characteristics and transmission mitigation behaviors on SARS-CoV-2 infection and seroprevalence in a large Northern California Bay Area cohort

Authors: Cameron Adams, MPH[1], Mary Horton, MPH[1], Olivia Solomon, MPH, Marcus Wong, Sean L. Wu, PhD, Sophia Fuller, MA, Xiaorong Shao, MA, Indro Fedrigo, Hong L. Quach, Diana L. Quach, Michelle Meas, MPH, Luis Lopez, Abigail Broughton, Anna L. Barcellos, MPH, Joan Shim, Yusef Seymens, Samantha Hernandez, Magelda Montoya, Darrell M. Johnson, MS, Kenneth B. Beckman, PhD, Michael P. Busch, MD, Josefina Coloma, PhD, Joseph A. Lewnard, PhD, Eva Harris, PhD[2], Lisa F. Barcellos, PhD[2]

Affiliations: University of California, Berkeley, California, USA (C. Adams, M. Horton, O. Solomon, M. Wong, S. Fuller, X. Shao, I. Fedrigo, H.L. Quach, D.L. Quach, M. Meas, L. Lopez, A. Broughton, A.L. Barcellos, J. Shim, Y. Seymens, S. Hernandez, M. Montoya, J. Coloma, J.A. Lewnard, E. Harris, L.F. Barcellos); Institute for Health Metrics and Evaluation, University of Washington, Seattle, Washington, USA (S.L. Wu); University of Minnesota Genomics Center, Minneapolis, Minnesota, USA (D.M. Johnson, K.B. Beckman); Vitalant Research Institute, San Francisco, California, USA (M.P. Busch)

Abstract

Comprehensive data on transmission mitigation behaviors and SARS-CoV-2 infection and serostatus are needed from large, community-based cohorts to identify SARS-CoV-2 risk factors and impact of public health measures. From July 2020 to March 2021, ≈5,500 adults from the East Bay Area, California were followed over three data collection rounds. We estimated the prevalence of antibodies from SARS-CoV-2 infection and COVID-19 vaccination, and self-reported COVID-19 test positivity. Population-adjusted SARS-CoV-2 seroprevalence was low,

increasing from 1.03% (95% CI: 0.50-1.96) in Round 1 (July-September 2020), to 1.37% (95% CI: 0.75-2.39) in Round 2 (October-December 2020), to 2.18% (95% CI: 1.48-3.17) in Round 3 (February-March 2021). Population-adjusted seroprevalence of COVID-19 vaccination was 21.64% (95% CI: 19.20-24.34) in Round 3. Despite >99% of participants reporting wearing masks, non-Whites, lower-income, and lower-educated individuals had the highest SARS-CoV-2 seroprevalence and lowest vaccination seroprevalence. Our results demonstrate that more effective policies are needed to address these disparities and inequities.

Introduction

The first confirmed severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) infection in the California Bay Area was reported on February 28, 2020.(1) In response, measures to prevent transmission were implemented soon after and included shelter-in-place orders, mask mandates, business and school closures, and social distancing recommendations. Despite these measures, spikes in reported infections occurred from July to September 2020 and December 2020 to February 2021.(2)

Local public health and healthcare systems experienced major challenges in preventing infections, identifying COVID-19 cases, and ensuring adherence to transmission mitigation strategies. Furthermore, disparities and inequities in COVID-19 have been observed in the United States(3) and there are few studies investigating the association of risk behaviors with infection outcomes at the individual level.(4–6) A detailed understanding of the effectiveness of transmission mitigation behaviors and sociodemographic factors that contribute to the dipropionate impact of COVID-19 in vulnerable communities is critical. Public health and policy directives aimed at controlling ongoing transmission, developing future prevention strategies,

and targeting health disparities and inequities must evidence-based. Large population-representative cohorts with individual-level data on social and behavioral factors(7) associated with COVID-19 and SARS-COV-2 infection and serostatus(8–12) are needed.

To address this need, we investigated individual-level characteristics and mitigation behaviors that contributed to SARS-CoV-2 seroprevalence, self-reported infection, and viral infection, and other outcomes in a large, population-based sample of over 5,500 individuals from 12 East Bay cities in Northern California followed longitudinally. During three time periods from July 2020 to April 2021, we estimated the population-adjusted prevalence of SARS-CoV-2 outcomes and differences by age, sex, race/ethnicity, ZIP code, and other demographic strata as well as the effect of transmission mitigation behavior on SARS-CoV-2 prevalence.

Methods

Recruitment and participants

Recruitment and selection of study participants was completed using a screening phase followed by a longitudinal study phase with three rounds of data collection (Figure 1). In the screening phase, all residential addresses within the East Bay cities and communities of Albany, Berkeley, El Cerrito, El Sobrante, Emeryville, Hercules, Kensington, Oakland, Piedmont, Pinole, Richmond, and San Pablo (~307,000 residential households) were mailed an invitation to complete a consent form and screening questionnaire. In a household, the individual aged 18 years or older with the next birthday was eligible to participate. Additional eligibility criteria included living within the study region, willingness to provide biospecimens and questionnaire responses, ability to read and speak English or Spanish, and having internet access and an email address.

The target sample size was 5,500 participants. The distribution of racial and ethnic identification in the screening questionnaire responses was more White and non-Hispanic than the region population. To obtain a sample that resembled the racial and ethnic proportions in the 2018 American Community Survey (ACS) for the study region, we ranked assigned ranks for order of inclusion to all eligible individuals who responded to the screening questionnaire. Black and/or Hispanic individuals were ranked the highest followed by other non-Whites. Order of inclusion for Whites was randomly sampled. In Round 1, individuals ranked between 1 and 5,500 were offered study enrollment. Those who declined to enroll were replaced with next highest ranked individuals who had not been offered study entry. In subsequent rounds, individuals who had participated in the previous round(s) were offered participation in the next round. If participation was declined, individuals from the pool of participants who had not participated in a study round were invited. Approximate dates for each round were July-September 2020, October-December 2020, and February-March 2021.

All participants provided their informed consent for the screening phase. All participants in the study phase provided their informed consent for each study round. The study was approved by the University of California, Berkeley Committee on Protection of Human Subjects (Protocol #2020-03-13121).

Study procedures

At the start of each round, eligible participants were invited to participate. Those who agreed to participate received a kit via FedEx containing materials for self-collection of biospecimens, pre-paid return shipping labels, and instructions to complete an online-administered questionnaire at the same time as biospecimen collection.

Questionnaire

90 The questionnaire addressed sex, gender, age, race/ethnicity, income, employment,
91 physical and mental health, as well as symptoms potentially related to COVID-19 within the
92 previous 2 weeks, and SARS-CoV-2 testing outside of the study. Participants were also asked
93 about behaviors that might affect risk of SARS-CoV-2 infection including physical distancing
94 practices, close contacts with others, and mask wearing. Questionnaires were available in
95 English or Spanish.

96 *SARS-CoV-2 viral and antibody testing*

97 Anterior nasal nare swabs for viral RNA testing and dried blood spots (DBS) for antibody
98 testing were collected from participants at each study round. Quantitative reverse transcription
99 PCR (RT-qPCR) was used to identify SARS-CoV-2 viral infection. Three tests were used to
100 assess anti-SARS-CoV-2 antibodies in DBS: Ortho VITROS® Anti-SARS-CoV-2 Total IgG and
101 spike IgG ELISA targeted antibodies against the SARS-CoV-2 spike protein (indicating prior
102 natural infection or vaccination), and Roche-NC Total IgG targeted antibodies against the
103 nucleocapsid (NC) protein (indicating prior natural infection only).(13) Before COVID-19
104 vaccines were available in the study region during rounds 1 and 2, detection of antibodies against
105 the SARS-CoV-2 spike protein was considered evidence of SARS-CoV-2 infection. During
106 Round 3, vaccinations were widely available, therefore detection of antibodies against the NC
107 protein were considered evidence of SARS-CoV-2 infection, while detection of antibodies
108 against the spike protein were considered evidence of SARS-CoV-2 infection or COVID-19
109 vaccination (Supplement S-4.2, Figure S-1).

110 *SARS-CoV-2 outcomes*

111 The following outcomes were investigated in each round: (1) cumulative SARS-CoV-2
112 antibody positivity, (2) participants' self-reported history of SARS-CoV-2 positivity from RT-

113 qPCR testing outside the study, (3) and a surveillance definition of "probable COVID-19 case"
114 derived from self-reported symptoms and close contact with infected individuals(14), and (4)
115 viral SARS-CoV-2 positivity from RT-qPCR testing of nasal swabs (Supplement 4.2). We also
116 investigated antibodies induced by COVID-19 vaccination only in Round 3. SARS-CoV-2
117 antibody positivity was defined as having detectable antibodies in the current and/or previous
118 round(s). DBS samples that tested both negative for anti-NC antibodies and positive for anti-
119 spike antibodies in Round 3 were considered to have antibodies from COVID-19 vaccination
120 alone. Self-reported COVID-19 viral positivity prevalence was defined as the proportion of
121 participants reporting a positive viral test among all participants who reported being tested within
122 a study round. Probable COVID-19 case prevalence was defined as the proportion of participants
123 identified as a probable COVID-19 case among all participants who provided valid responses
124 within a study round. Viral positivity prevalence was defined as testing positive by RT-qPCR
125 from nasal swab samples.

126 **Statistical analysis**

127 *Population-adjusted seroprevalence and other SARS-CoV-2 outcomes*

128 Bayesian multilevel regression and poststratification (MRP) was used to estimate
129 population-adjusted cumulative seroprevalence, self-reported SARS-CoV-2 viral positivity
130 prevalence at each study round, and "probable COVID-19" prevalence at each study round. MRP
131 is a regression-based method for estimating population and sub-population averages from survey
132 data that has been shown to perform better than survey weighting, particularly with sparse data
133 (15,16).

134 In addition to estimation of regional prevalence of our outcomes, we estimated
135 prevalence within demographic groups and geographic areas in the study region. Variables of

interest were gender, age, race, Hispanic ethnicity, income, education, household size, and ZIP code. We used a method described by Leeman and colleagues to generate a synthetic population for poststratification using data from the 2018 American Community Survey (ACS) and the Public Use Microdata Sample (PUMS).(16) Poststratification was done using binary sex because gender is not reported by the ACS. Race and ethnicity were combined into a single variable to reduce the number of poststratification strata.

At each study round, binary SARS-CoV-2 outcomes were modeled as a function of geographic and demographic characteristics using multilevel logistic regression models. Participant sex was included as a fixed effect. Vectors of random intercepts were defined for each category of race/ethnicity, age, education, income, household size, and ZIP code and two-way interactions between ZIP code, race/ethnicity, educational attainment, income, and age. To improve estimation of geographic effects we allowed for spatial correlations using the modified Besag-York-Mollié model and included the proportion of Spanish speaking households within the ZIP code of residence from the ACS as a fixed effect (Supplement S-6).(17)

SARS-CoV-2 outcome prevalence and measures of association

We report populated-adjusted prevalence of SARS-CoV-2 outcomes across the study region and within geographic and demographic groups of interest. To calculate prevalence estimates, posterior distributions of the relevant poststratification stratum were aggregated. We also estimated prevalence differences (PD) and prevalence ratios (PR) for the association between populated-adjusted SARS-CoV-2 outcomes and race/ethnicity, education, and sex. For each parameter of interest, the mean of the posterior distribution was the point estimate, and the 95% credible interval (CI) was the 2.5% and 97.5% quantiles of a posterior distribution.

SARS-CoV-2 test-kit bias corrected seroprevalence

We estimated cumulative SARS-CoV-2 seroprevalence at each study round adjusted for the net sensitivity and specificity of the SARS-CoV-2 antibody testing algorithm in each study round (Supplement S-6).

Transmission mitigation behavior analyses

In each round, participants were asked about physical distancing practices, recent close contacts with others, mask wearing, and other behaviors and activities that might affect the risk of SARS-CoV-2 infection. We classified participants into two behavior categories, "high-risk" and "low-risk", with those responses using latent class analysis.(18) Crude associations between behavior categories and characteristics such as sex, age, race/ethnicity, education, and income were assessed with χ^2 tests. Associations between high-risk vs. low-risk behavior and within round SARS-CoV-2 seroprevalence and self-reported test positivity were estimated using the MRP model described above with random intercepts for behavior categories and interactions between the behavior categories and ZIP code, age, race/ethnicity, education, and income (Supplement S-5 and S-6).

Statistical analyses were completed in R 4.0.2. NIMBLE was used to implement MRP models.(19) Detailed descriptions of MRP methods and code are provided in Supplementary Methods and at github.com/adams-cam/ebcovid_prev.

Results

Enrollment and characteristics of study participants

Of the 16,115 residents who consented and completed the screening procedures between May-July 2020, 1,777 did not satisfy inclusion criteria and were excluded (Figure 1). Characteristics of participants are presented in Table 1 and Table S-1. Participation rates were high (Round 1: 76.8%, Round 2: 89.8%, and Round 3: 87.3%), and participants identified predominantly as

182 female (~63%). Those aged 45-64 years were the largest age group of participants across all
183 study rounds (ranging from 37.3% to 39.4%). Most participants identified as White (52.5% to
184 63.3%), followed by Asian/Pacific Islander (13.9% to 15.7%), Hispanic (11.0% to 15.6%), two
185 or more races (6.9% to 9.1%), African American or Black (3.0% to 4.9%), and Native
186 American/Alaska Native or other (1.7% to 2.2%).

187 Of those who completed the questionnaire, 87.3%, 95.3%, and 96.6% provided DBS and
188 93.6%, 98.1% and 98.% provided nasal swabs in rounds 1, 2, and 3, respectively (Table 2).
189 Antibodies against the SARS-CoV-2 spike protein were detected in 29 (0.6%) and 33 (0.6%) of
190 DBS in rounds 1 and 2, respectively. In Round 3, NC antibodies from natural infection alone
191 were detected in 84 participants (1.8% of 4,806) and spike antibodies from natural infection or
192 vaccination were detected in 1,452 participants (31.3% of 4,806). Viral infection was detected in
193 less than three nasal swabs in each round. The proportion of participants reporting both receiving
194 a SARS-CoV-2 test outside the study and testing positive within a study round increased over the
195 study period; 10/1,030 (1.0%) participants reporting a positive COVID-19 result in Round 1,
196 19/2,059 (0.9%) in Round 2, and 53/1892 (2.8%) in Round 3. In all rounds, few participants met
197 the criteria for being a COVID-19 probable case (<0.5%) (Table S-2).

198 *Population adjusted SARS-CoV-2 outcome prevalence in study region*

199 Population adjusted SARS-CoV-2 seroprevalence, self-reported COVID-19 test
200 positivity, and probable COVID-19 cases are reported in Table 3. Overall, populated-adjusted
201 SARS-CoV-2 natural infection seroprevalence was low across the study region: Round 1 (July-
202 September 2020) 1.03% (95% CI 0.50-1.96), Round 2 (October-December 2020 1.37% (0.75-
203 2.39), and Round 3 (February-March 2021) 2.18% (95% CI 1.48-3.17). In Round 3 the
204 populated-adjusted seroprevalence of COVID-19 vaccination was 21.64% (95% CI: 19.2, 24.34).

205 Models incorporating sensitivity and specificity of the antibody assays yielded lower SARS-
206 CoV-2 seroprevalence estimates in Round 1, similar estimates in Rounds 2 and 3, and a higher
207 estimate of COVID-19 vaccine seroprevalence in Round 3. Population adjusted self-reported test
208 positivity to SARS-CoV-2 was similar in Rounds 1 (1.11%, 95 CI: 0.39-2.40) and 2 (1.29%,
209 95% CI: 0.55, 2.17) to seroprevalence estimates and increased to 4.58% (95% CI: 2.56-7.64) in
210 Round 3. Population-adjusted prevalence of being a COVID-19 probable case was <1% across
211 rounds (Figure S-2).

212 *SARS-CoV-2 seroprevalence and infection vary by geographic area*

213 There was evidence for spatial differences in both populated-adjusted seroprevalence and
214 self-reported test positivity (Figure 2). The northern areas (Richmond, San Pablo, Pinole, and
215 Hercules) and southern areas (East Oakland) of the study region had higher seroprevalence than
216 Berkeley, El Cerrito, and North/Downtown Oakland. Self-reported test positivity was also higher
217 in the northern and southern areas. These trends were consistent across study rounds.

218 *Population adjusted seroprevalence and self-reported test positivity prevalence within subgroups*

219 Differences in populated-adjusted seroprevalence were observed among demographic
220 groups (Figure 3). In Round 1, those aged less than 45 years had higher likelihood of antibody
221 positivity than those aged 45 or older, but this relationship reversed over time (Figure 3B). In
222 general, non-Whites consistently had a higher likelihood of antibody positivity than Whites
223 (Figure 3C). The likelihood of antibody positivity also differed by household income and
224 educational attainment, with those reporting a household income <$100,000 and less than a
225 college degree having a higher likelihood of antibody positivity than their comparison groups
226 (Figures 3D-E). Those in households with more than four people had a higher likelihood of
227 antibody positivity than those in households with four or less people (Figure 3F). There was no

228 clear relationship between sex and populated-adjusted seroprevalence (Figure 3A). Similar
229 relationships were seen between demographic groups and self-reported test prevalence (Figure S-
230 3).

231 *COVID-19 vaccination seroprevalence differs by race/ethnicity and age*

232 In Round 3, populated-adjusted COVID-19 vaccination seroprevalence was lower in all
233 non-White groups compared to Whites, (White vs. Non-White, PD: -4.35%, 95% CI: (-9.32, -
234 0.35), Figure 4, Table 4). The difference between racial/ethnic groups was largest among older
235 participants; non-Whites had a lower COVID-19 vaccination seroprevalence compared to Whites
236 among those aged between 65-74 (PD: -11.67%, 95% CI: (-20.98, -2.49)) and those aged 75 or
237 older (PD: -10.15%, 95% CI: (-22.25, 0.43)).

238 *Mask-wearing and association between high-risk vs. low-risk behavior and seroprevalence*

239 More than 99% of participants reported ever wearing a mask 99% reported wearing a
240 mask during leisure and exercise activities, >91% reported wearing a mask at work, and >88%
241 reported wearing a mask while shopping (Table S-3). After clustering participants into "low-
242 risk" and "high-risk" groups according to self-reported mitigation behaviors, most participants
243 were considered low-risk across the study rounds (70%, 82%, and 77%, respectively; Figure
244 5A). Behaviors with the largest differences in high- versus low-risk behavior were reporting
245 "yes" to: left the house for work, medical/healthcare, care of relative, or other, worked with
246 potential COVID-19 contact, attended a gathering, and traveled to county outside of residence
247 within last two weeks (Figure 5B). Age, race/ethnicity, and education were all associated
248 ($p<0.001$) with high- vs. low-risk groups (Table S-4), but we did not observe strong evidence for
249 an association between mitigation behavior and either seroprevalence or self-reported test

positivity (Table 5), although the point estimates indicate that "low-risk" behavior was associated with lower seroprevalence.

Discussion

In the current study, we investigated individual-level characteristics and behaviors that contributed to SARS-CoV-2 related outcomes, including seroprevalence and self-reported infection, in a large, population-based sample of over 5,500 participants from 12 East Bay (Northern California) cities. During three time periods from July 2020 to March 2021, we estimated the population-adjusted prevalence of SARS-CoV-2 outcomes across the study region and within strata of age, sex, race/ethnicity, ZIP code, and household size. We then characterized behaviors to mitigate transmission of SARS-CoV-2 and their effects on related outcomes. Overall, prevalence of SARS-CoV-2 outcomes for natural infection were low which may be attributable to the high percentage of mask-wearing and other risk-mitigating behaviors among our participants. COVID-19 vaccination seroprevalence estimates in Round 3 was greater than 20%, with non-Whites having much lower seroprevalence than Whites.

Despite the low overall SARS-CoV-2 seroprevalence and infection observed in our study, ZIP code of residence, age, racial/ethnic identification, education, and income all contributed to differences in seroprevalence. Specifically, non-Whites, those without a college degree, and lower income households, had higher seroprevalence. Further, ZIP codes with higher proportions of Spanish speakers had higher populated-adjusted seroprevalence estimates (Figure S-4). These differences persisted despite the high rates of mask wearing reported by our study sample, further adding to strong evidence that the risk of COVID-19 is distributed unequally and that structural inequities play an important role in COVID-19 risk.(20–23) Moreover, during Round 3 (February-March 2021), COVID-19 vaccines were widely available across the study region. We

273 found that Whites had much higher prevalence of antibodies against the SARS-CoV-2 spike
274 protein in Round 3, indicative of vaccination status, than non-Whites, as reported
275 elsewhere.(3,24) Notably, this difference was largest among those aged 65 or older. These
276 findings demonstrate that in the first few months of vaccine availability in the Bay Area, large
277 disparities in vaccination rates by race/ethnicity were observed among older persons, and Whites,
278 the group with the lowest prevalence of SARS-CoV-2 infection, were more likely to be
279 vaccinated, further underscoring the inequities that exist surrounding the coronavirus pandemic.
280 Another key finding was that almost all participants reported wearing masks. This
281 contrasts with models of mask usage reported by the Institute for Health Metrics and Evaluation
282 which reported mask usage ranging between 75-82% from December 2020 through March 2021
283 in California.(25,26) Mask wearing is one the most effective behaviors for controlling
284 community spread of SARS-CoV-2 infection.(27) The high rate of mask usage by study
285 participants may partially explain why we did not detect a differences between high-risk and
286 low-risk mitigation behavior and SARS-CoV-2 prevalence, and partially explain why our
287 estimates of SARS-CoV-2 seroprevalence and self-report test positivity were lower than public
288 case reports. However, a recent study estimating SARS-CoV-2 seroprevalence from blood
289 donors reported seroprevalence in the Bay Area region of California that were similar to our
290 estimates through December 2020.(28)
291 A major strength of this study was the longitudinal design and collection of individual-
292 level data, including biospecimens for antibody and virus testing, which is challenging but much
293 needed in current studies of the pandemic. Comprehensive data on social distancing, self-
294 quarantine, mask wearing, working from home, and other public health and SARS-CoV-2
295 mitigation efforts are needed to inform current and future strategies of prevention and for whom

296 to target prevention efforts. At-home collection of biospecimens, including dried blood spots for
297 antibody testing, made it possible to regularly test participants without any in-person interaction.
298 This was a critical feature, particularly early in the pandemic, when recommendations were to
299 travel only for essential purposes and to limit in-person interactions. At-home sample collection
300 was used to obtain more than 30,000 biospecimens and is a feasible approach for even larger
301 populations and geographic regions.

302 One limitation of this study was the under-representation of certain demographics in our
303 sample. Supplementary mailings of recruitment postcards in Spanish were sent to residences in
304 ZIP codes with high proportions of Spanish speaking households. We also placed recruitment
305 flyers in local grocery stores and conducted outreach to community organizations, local
306 government officials, and school districts in the study region. Despite these efforts, non-Whites,
307 males, lower income households, and individuals with lower education and from lower
308 socioeconomic ZIP codes were underrepresented in our sample. Unhoused individuals were also
309 not captured in our sample. This was important given evidence that individuals who identify as
310 Hispanic or Black, and other underrepresented groups, are at the highest risk for COVID-19.(23)
311 These discrepancies may have led to underestimation of overall SARS-CoV-2 seroprevalence
312 and self-reported infection, although the MRP models may have somewhat minimized this effect
313 by pooling information across similar observed characteristics in the sample data. Our
314 seroprevalence estimates were consistent with weekly case prevalence reported by public health
315 agencies in ZIP codes with high response rates (Figure S-5).(29) Additionally, there may be
316 unmeasured confounding from variables not included in the analyses and self-selection bias from
317 some participants in COVID-19 research who may be more fervent adherers to COVID-19

public health measures. Finally, some individuals do not seroconvert after infection or vaccination.

Our results underscore the substantial and persistent inequities that exist surrounding the coronavirus pandemic. Non-Whites, lower-income, and lower-educated individuals had the highest SARS-CoV-2 seroprevalence. This disparity in seroprevalence was observed despite near universal rates of mask wearing in our sample. We also observed large differences in COVID-19 vaccination seroprevalence between racial and ethnic groups. More work must be done to address these disparities and inequities, such as allocation of resources for high-risk communities and strategies to mitigate the structural barriers posed by social and structural determinants of health.

Acknowledgments

We want to acknowledge the contributions of the following individuals to this study: Ella Parsons, Jordan Keen, Janine Solomon, Jose Salinas, Kevin Duong, Joseph Egbunikeokye, Maya Talavera, Riya Shrestha, Colin Warnes, José Victor Zambrana, Nicholas Lo, Parnal Narvekar, Fausto Bustos, Gregorio Dias, Reinaldo Mercado-Hernandez, Julia Huffaker, Raymond Montes, Alexandra Zermeno, Alejandra Zeiger, William Dow, Michael Lu, Lila Krop, Kelly Lam, Yan Zhang, Sarah Folkmanis, Sophie Zhai, Dingjun Chen, Ruben Vargas Ethan Garcia, Oliver Li, Manisha Sahoo, Raina Walencewicz, Sophia Wang, Antonia Gibbs, Amrita Ramanathan, Catherine Livelo, Taylor Worley, Amanda Tanaka, Savinnie Ho, Jane Liu Ryan Allen, Sofia Soltero, Victoria Van Metter, Madeleine Fraix, Allie Coyne, Subeksha Sharma Lydia Yu, Shreeya Garg, Sanjeet Paluru, Malika Saxena, Talia Panadero, Ayra Rahman, Joshua Calangian, Dharaa Upadhyaya, Sophia Kemp, Ruhi Parikh, Amy Rich, Sophie Manoukian, Nola Vu, Crystal Nguyen, Jordyn Pinochi, Alma Kuc, Siri Ylenduri, Manvir Kaur, Angikaar Chana, and Sannidhi

Sarvadhavabhatla, Benjamin T. Auch, Dinesha Walek, Evan Forsberg, Jerry Daniel, Veronica Tonnell, Ji Hyun (Jay) Kim, Mary Nieuwenhuis. Creative Testing Solutions: Valerie Green, Sherri Cyrus, Phillip Willamson, Brett Hirsch, Paul Contestable, Mars Stone, Joe Derisi, Emily Crawford, Emily Ahlvin, Armando Diaz, and Favianna Rodriquez.

References

1. Eby K. Coronavirus Timeline: Tracking major moments of COVID-19 pandemic in San Francisco Bay Area [Internet]. ABC7 San Francisco. 2020 [cited 2021 Nov 29]. Available from: https://abc7news.com/6047519/

2. United States. Agency for Toxic Substances and Disease Registry. Geospatial Research A and SP, editor. CDC COVID data tracker. 2020 Sep 15; Available from: https://stacks.cdc.gov/view/cdc/96118

3. Njoku A, Joseph M, Felix R. Changing the Narrative: Structural Barriers and Racial and Ethnic Inequities in COVID-19 Vaccination. International Journal of Environmental Research and Public Health. 2021 Jan;18(18):9904.

4. Qualls N. Community Mitigation Guidelines to Prevent Pandemic Influenza — United States, 2017. MMWR Recomm Rep [Internet]. 2017 [cited 2021 Nov 10];66. Available from: https://www.facebook.com/CDCMMWR

5. CDC. Interim pre-pandemic planning guidance : community strategy for pandemic influenza mitigation in the United States : early, targeted, layered use of nonpharmaceutical interventions [Internet]. [cited 2021 Nov 10]. Available from: https://stacks.cdc.gov/view/cdc/11425

362 6. Bell DM, World Health Organization Working Group on International and Community
363 Transmission of SARS. Public health interventions and SARS spread, 2003. Emerg Infect
364 Dis. 2004 Nov;10(11):1900–6.

365 7. Hutchins HJ. COVID-19 Mitigation Behaviors by Age Group — United States, April–June
366 2020. MMWR Morb Mortal Wkly Rep [Internet]. 2020 [cited 2021 Nov 10];69. Available
367 from: https://www.cdc.gov/mmwr/volumes/69/wr/mm6943e4.htm

368 8. Havers FP, Reed C, Lim T, Montgomery JM, Klena JD, Hall AJ, et al. Seroprevalence of
369 Antibodies to SARS-CoV-2 in 10 Sites in the United States, March 23-May 12, 2020.
370 JAMA Internal Medicine. 2020 Dec 1;180(12):1576–86.

371 9. Bendavid E, Mulaney B, Sood N, Shah S, Ling E, Bromley-Dulfano R, et al. COVID-19
372 Antibody Seroprevalence in Santa Clara County, California. medRxiv. 2020
373 Apr;2020.04.14.20062463-2020.04.14.20062463.

374 10. Sood N, Simon P, Ebner P, Eichner D, Reynolds J, Bendavid E, et al. Seroprevalence of
375 SARS-CoV-2–Specific Antibodies Among Adults in Los Angeles County, California, on
376 April 10-11, 2020. JAMA. 2020 Jun 16;323(23):2425–7.

377 11. Self WH. Seroprevalence of SARS-CoV-2 Among Frontline Health Care Personnel in a
378 Multistate Hospital Network — 13 Academic Medical Centers, April–June 2020. MMWR
379 Morb Mortal Wkly Rep [Internet]. 2020 [cited 2021 Nov 10];69. Available from:
380 https://www.cdc.gov/mmwr/volumes/69/wr/mm6935e2.htm

12. Rosenberg ES, Tesoriero JM, Rosenthal EM, Chung R, Barranco MA, Styer LM, et al. Cumulative incidence and diagnosis of SARS-CoV-2 infection in New York. Ann Epidemiol. 2020 Aug;48:23-29.e4.

13. Wong MP, Meas MA, Adams C, Hernandez S, Green V, Montoya M, et al. Development and Implementation of Dried Blood Spot-based COVID-19 Serological Assays for Epidemiologic Studies. medRxiv. 2021 Jan 1;2021.11.25.21266786.

14. CSTE Interim Position Statement: Update to COVID-19 Case Definition - Council of State and Territorial Epidemiologists [Internet]. [cited 2021 Oct 13]. Available from: https://www.cste.org/news/520707/CSTE-Interim-Position-Statement-Update-to-COVID-19-Case-Definition.htm

15. Gelman A, Little TC. Poststratification Into Many Categories Using Hierarchical Logistic Regression. Survey Research. 1997 Dec;23:127–35.

16. Leemann L, Wasserfallen F. Extending the Use and Prediction Precision of Subnational Public Opinion Estimation: EXTENDING USE AND PRECISION OF MrP. American Journal of Political Science. 2017 Oct;61(4):1003–22.

17. Simpson D, Rue H, Riebler A, Martins TG, Sørbye SH. Penalising Model Component Complexity: A Principled, Practical Approach to Constructing Priors. Statist Sci [Internet]. 2017 Feb 1 [cited 2021 Sep 15];32(1). Available from: https://projecteuclid.org/journals/statistical-science/volume-32/issue-1/Penalising-Model-Component-Complexity--A-Principled-Practical-Approach-to/10.1214/16-STS576.full

401 18. Linzer DA, Lewis JB. poLCA: An R Package for Polytomous Variable Latent Class
402 Analysis. Journal of Statistical Software. 2011 Jun 14;42:1–29.

403 19. de Valpine P, Turek D, Paciorek CJ, Anderson-Bergman C, Lang DT, Bodik R.
404 Programming with models: writing statistical algorithms for general model structures with
405 NIMBLE. Journal of Computational and Graphical Statistics. 2017 Apr 3;26(2):403–13.

406 20. Reitsma MB, Claypool AL, Vargo J, Shete PB, McCorvie R, Wheeler WH, et al.
407 Racial/Ethnic Disparities In COVID-19 Exposure Risk, Testing, And Cases At The
408 Subcounty Level In California. Health Affairs. 2021 Jun 1;40(6):870–8.

409 21. Paremoer L, Nandi S, Serag H, Baum F. Covid-19 pandemic and the social determinants of
410 health. BMJ. 2021 Jan 29;372:n129.

411 22. Abrams EM, Szefler SJ. COVID-19 and the impact of social determinants of health. Lancet
412 Respir Med. 2020 Jul;8(7):659–61.

413 23. Magesh S, John D, Li WT, Li Y, Mattingly-app A, Jain S, et al. Disparities in COVID-19
414 Outcomes by Race, Ethnicity, and Socioeconomic Status: A Systematic-Review and Meta-
415 analysis. JAMA Network Open. 2021 Nov 11;4(11):e2134147.

416 24. Agarwal R, Dugas M, Ramaprasad J, Luo J, Li G, Gao G (Gordon). Socioeconomic privilege
417 and political ideology are associated with racial disparity in COVID-19 vaccination. PNAS
418 [Internet]. 2021 Aug 17 [cited 2021 Nov 10];118(33). Available from:
419 https://www.pnas.org/content/118/33/e2107873118

420 25. Reiner RC, Barber RM, Collins JK, Zheng P, Adolph C, Albright J, et al. Modeling COVID-
421 19 scenarios for the United States. Nat Med. 2021 Jan;27(1):94–105.

422 26. IHME | COVID-19 Projections [Internet]. Institute for Health Metrics and Evaluation. [cited
423 2021 Oct 1]. Available from: https://covid19.healthdata.org/

424 27. Brooks JT, Butler JC. Effectiveness of Mask Wearing to Control Community Spread of
425 SARS-CoV-2. JAMA. 2021 Mar 9;325(10):998–9.

426 28. Jones JM, Stone M, Sulaeman H, Fink RV, Dave H, Levy ME, et al. Estimated US Infection-
427 and Vaccine-Induced SARS-CoV-2 Seroprevalence Based on Blood Donations, July 2020-
428 May 2021. JAMA [Internet]. 2021 Sep 2 [cited 2021 Sep 30]; Available from:
429 https://doi.org/10.1001/jama.2021.15161

430 29. The Los Angeles Times' open-source archive of California coronavirus data [Internet]. Los
431 Angeles Times Data and Graphics Department; 2021 [cited 2021 Oct 12]. Available from:
432 https://github.com/datadesk/california-coronavirus-data

433

434 Address for correspondence: Lisa F. Barcellos, 324 Stanley Hall, Berkeley, CA 94720,
435 lbarcellos@berkeley.edu, 510-642-7814

436

437

438 Table 1. Characteristics of participants at each round of the study compared to study region

439 population.

	PS Strata† %	Round 1 no. (%)	Round 2 no. (%)	Round 3 no. (%)
Invited to round		7166	6242	5506
Participated in study round		5501 (76.8)	5603 (89.8)	4806 (87.3)
Gender				
Female		3451 (62.7)	3509 (62.6)	3040 (63.3)
Male		1996 (36.3)	2036 (36.3)	1721 (35.8)
Other		51 (0.9)	52 (0.9)	40 (0.8)
Missing		3 (0.1)	6 (0.1)	5 (0.1)
Sex				
Female	51.8	3502 (63.7)	3559 (63.5)	3080 (64.1)
Male	48.2	1996 (36.3)	2042 (36.4)	1725 (35.9)
Missing		3 (0.1)	2 (0)	1 (0)
Age				
18-29	24.2	476 (8.7)	378 (6.7)	286 (6)
30-44	29.0	1768 (32.1)	1656 (29.6)	1330 (27.7)
45-64	30.0	2052 (37.3)	2141 (38.2)	1895 (39.4)
65-74	10.2	928 (16.9)	1077 (19.2)	977 (20.3)
75+	6.6	275 (5)	345 (6.2)	313 (6.5)
Missing		2 (0)	6 (0.1)	5 (0.1)
Race/ethnicity				
African American/Black	17.4	268 (4.9)	180 (3.2)	146 (3)
American Indian/Alaska Native/Other	0.5	123 (2.2)	114 (2)	84 (1.7)
Asian/Pacific Islander	20.3	862 (15.7)	763 (13.6)	668 (13.9)
Hispanic	22.4	860 (15.6)	637 (11.4)	527 (11)
Two or more races	3.8	500 (9.1)	400 (7.1)	334 (6.9)
White	35.6	2886 (52.5)	3503 (62.5)	3041 (63.3)
Missing		2 (0)	6 (0.1)	6 (0.1)
Educational attainment				
College degree or more	47.0	4912 (89.3)	5101 (91)	4404 (91.6)
Less than college degree	53.0	552 (10)	496 (8.9)	399 (8.3)
Missing		37 (0.7)	6 (0.1)	3 (0.1)
Current annual household income, USD				
<$100K	62.9	2081 (37.8)	2075 (37)	1779 (37)
≥$100K	37.1	3192 (58)	3347 (59.7)	2866 (59.6)
Missing		228 (4.1)	181 (3.2)	161 (3.3)

*PS, Poststratification; USD, United States Dollar.
†Population percentages from synthetic poststratification tables generated from American Community Survey (ACS) and Public Use Microdata Sample data. Each cell contains the percentage of population age >18. Gender is not available in ACS data.

440

441

442 Table 2. Prevalence of SARS-CoV-2 related outcomes among participants per study round.

443

	Round 1	Round 2	Round 3
Biospecimen collection period	Jul. 4, 2020 – Sep. 1, 2020	Oct. 10, 2020 – Dec. 16, 2020	Feb. 11, 2021 – Mar. 20, 2021
N	5501	5603	4806
Provided DBS, no. (%)	4801 (87.3)	5340 (95.3)	4641 (96.6)
Provided nasal swab, no. (%)	5148 (93.6)	5499 (98.1)	4735 (98.5)
SARS-CoV-2 spike antibody status†, no. (%)			
Non-reactive	4641 (96.7)	5275 (98.8)	3157 (68.0)
QNS	131 (2.7)	31 (0.6)	32 (0.7)
Reactive	29 (0.6)	33 (0.6)	1452 (31.3)
SARS-CoV-2 nucleocapsid antibody status‡, no. (%)			
Non-reactive			4181 (90.1)
QNS			376 (8.1)
Reactive			84 (1.8)
SARS-CoV-2 viral status, no. (%)			
Invalid/Inconclusive	4 (0.1)	5 (0.1)	7 (0.1)
Not Detected	5143 (99.9)	5491 (99.9)	4725 (99.8)
Positive 2019-nCoV	1 (0.0)	3 (0.1)	3 (0.1)
Self-reported positive COVID-19 test, no. (%)			
No	1020 (18.5)	2059 (36.7)	1892 (39.4)
Yes	10 (0.2)	19 (0.3)	53 (1.1)
Not tested	4471 (81.3)	3525 (62.9)	2861 (59.5)
Probable COVID-19 case§, no. (%)			
No	4800 (87.3)	4938 (88.1)	4419 (91.9)
Yes	21 (0.4)	10 (0.2)	8 (0.2)
Not reported	680 (12.4)	655 (11.7)	379 (7.9)

*DBS, dried blood spot; QNS, quantity not sufficient.
†Antibody assays specific to spike antibodies detect presence of antibodies induced by SARS-CoV-2 natural infection or vaccination.
‡Antibody assays specific to nucleocapsid antibodies detect presence of antibodies induced by SARS-CoV-2 natural infection.
§As defined by the Council for State and Territorial Epidemiologists.

456

457

458

459 Table 3. Estimated population-adjusted prevalence (%) and 95% credible intervals within study

460 region of SARS-CoV-2 outcomes.

SARS-CoV-2 Outcome	Round 1 Prev.% (95% CI)	Round 2 Prev. % (95% CI)	Round 3 Prev. % (95% CI)
SARS-CoV2 Ab+ NI†			
Crude	29/4670 = 0.62%	48/5949 = 0.81%	110/5991 = 1.83%
Pop-adjusted‡	1.03 (0.50, 1.96)	1.37 (0.75, 2.39)	2.18 (1.48, 3.17)
Pop-adjusted + Test-bias adjusted§	0.63 (0.04, 2.00)	1.02 (0.17, 2.40)	2.19 (1.49, 3.18)
SARS-CoV2 Ab+ V¶			
Crude			1369/4608 = 29.7%
Pop-adjusted‡			21.64 (19.2, 24.34)
Pop-adjusted + Test-bias adjusted§			27.56 (22.84, 33.55)
Self-reported test positivity#			
Crude	10/1030 = 0.97%	19/2078 = 0.91%	53/1945 = 2.73%
Pop-adjusted‡	1.11 (0.39, 2.40)	1.29 (0.55, 2.71)	4.58 (2.56, 7.64)
Probable COVID-19 case¶			
Crude	21/4821 = 0.44%	10/4948 = 0.2%	8/4427 = 0.18%
Pop-adjusted‡	0.59 (0.26, 1.19)	0.64 (0.17, 1.75)	0.65 (0.17, 1.77)

*Ab, antibody; NI, natural infection; Prev., prevalence; V, vaccination.
†Estimate of cumulative antibody prevalence from natural infection by SARS-CoV-2.
‡Population-adjusted prevalence estimated using multilevel regression and poststratification models.
§Population-adjusted prevalence estimated using multilevel regression and poststratification models that included expressions for antibody assay sensitivity and specificity.
¶Estimate of cumulative prevalence of antibodies from natural infection by SARS-CoV-2 or COVID-19 vaccination.
#Estimated within round prevalence of self-reported COVID-19 test positivity.
††Estimated within round prevalence of COVID-19 probable case.

461

462

463 Table 4. Round 3 seroprevalence of antibodies from COVID-19 vaccination within race/ethnicity

464 and age groups and prevalence differences between non-Whites and Whites.

	Within group Prev. % (95% CI)†	Non-whites compared to whites PD % (95% CI)‡
Race/ethnicity		
White	25.33 (23.14, 27.65)	Ref
African American/Black	19.69 (13.47, 26.13)	-5.65 (-12.17, 0.76)
American Indian/Alaska Native/Other	18.87 (12.32, 25.95)	-6.46 (-13.3, 0.51)
Asian/Pacific Islander	21.34 (17.24, 26.13)	-4 (-8.4, 0.93)
Hispanic	18.64 (13.74, 24.4)	-6.69 (-11.83, -0.81)
Two or more races	15.74 (11.85, 19.94)	-9.6 (-13.88, -5.28)
All non-white	19.6 (16.16, 23.33)	-4.35 (-8.32, -0.35)
Age group, y		
18-29	15.04 (10.92, 19.81)	-3.2 (-9.14, 2.23)
30-44	16.53 (13.84, 19.57)	-1.94 (-6, 2.74)
45-64	16.6 (13.99, 19.61)	-1.04 (-5.44, 3.89)
65-74	45.74 (40.72, 50.75)	-11.67 (-20.98, -2.49)
75 +	54.01 (46.12, 61.79)	-10.15 (-22.25, 0.43)

*CI, credible interval; PD, prevalence difference; Prev, prevalence.
†Population adjusted seroprevalence in round 3 estimated using multilevel regression and poststratification (MRP); Seroprevalence and 95% CI are the mean and 2.5% and 97.5% quantiles of a posterior distribution respectively. Assay used detects antibodies from natural SARS-CoV-2 infection or from COVID-19 vaccination.
‡Population adjusted COVID-19 seroprevalence difference in round 3 between each race/ethnicity group and Whites among race/ethnicity and between non-Whites and Whites within each age group. PDs estimated using MRP, and the PDs and 95% CIs are the mean and 2.5% and 97.5% quantiles of a posterior distribution, respectively. A

465

466

467 Table 5. Association between seroprevalence and self-reported test positivity and high-risk vs.

468 low-risk mitigation behavior within each study round.

	MOA†	Round 1	Round 2	Round 3
SARS-CoV-2 antibody prevalence‡	PD % (95% CI)	-0.21 (-1.4, 0.79)	0.02 (-1.59, 1.03)	-0.46 (-2.23, 1.00)
	PR (95% CI)	0.93 (0.34, 2.03)	1.4 (0.32, 3.93)	0.88 (0.45, 1.57)
Self-reported test positivity	PD % (95% CI)	0.53 (-0.73, 2.24)	0.11 (-2.12, 1.6)	-1.11 (-5.26, 2.18)
	PR (95% CI)	2.17 (0.54, 6.72)	1.56 (0.33, 4.65)	0.87 (0.41, 1.64)

*CI, credible interval; MOA, measure of association; PD, prevalence difference; PR, prevalence ratio
†Populated adjusted prevalence estimated using multilevel regression and poststratification models; point estimate is the mean of a posterior distribution of the parameter, 95% credible interval estimated from 2.5% and 97.5% quantiles of a posterior distribution.
‡Defined as having detectable SARS-CoV-2 antibodies within a study round.

469

470

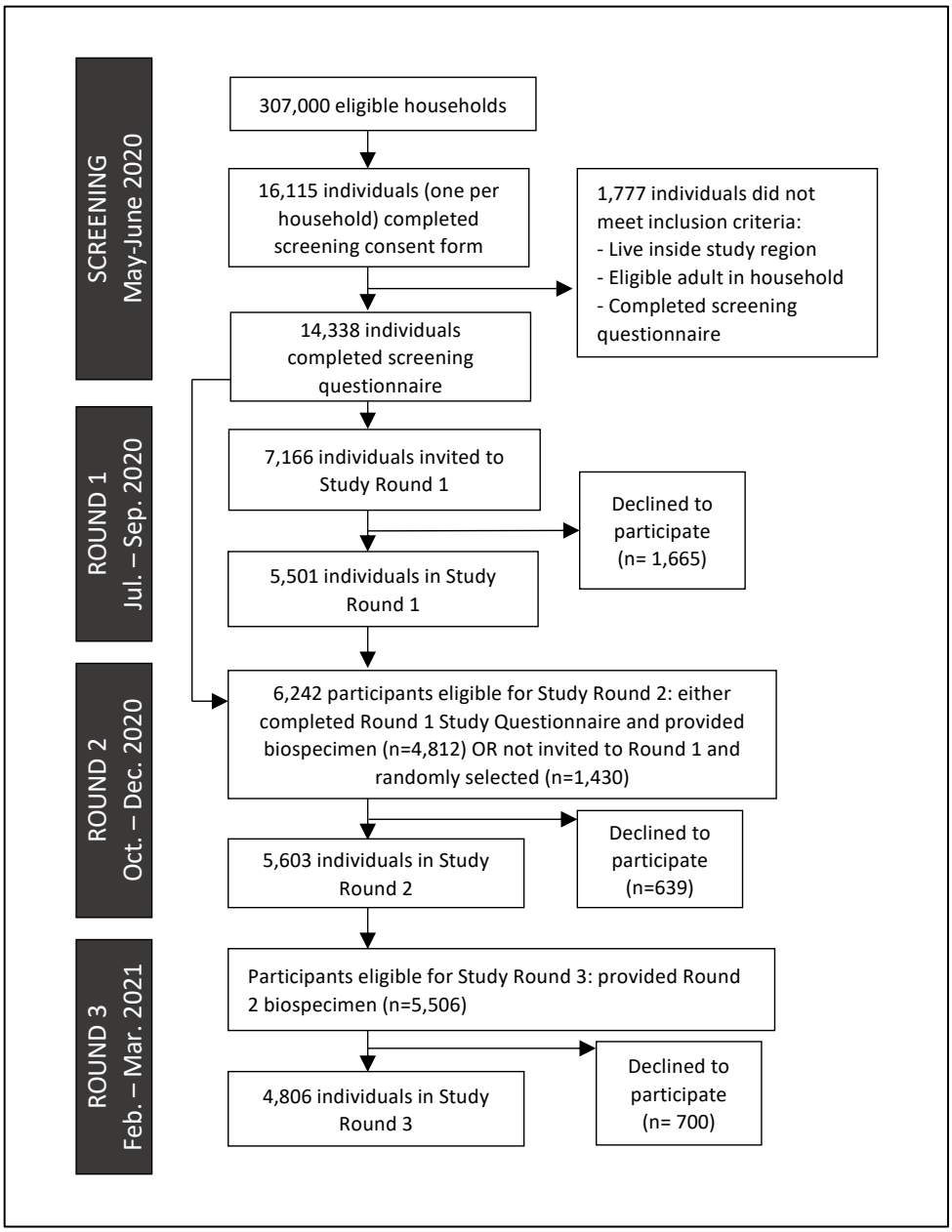

Figure 1. Study flow chart.

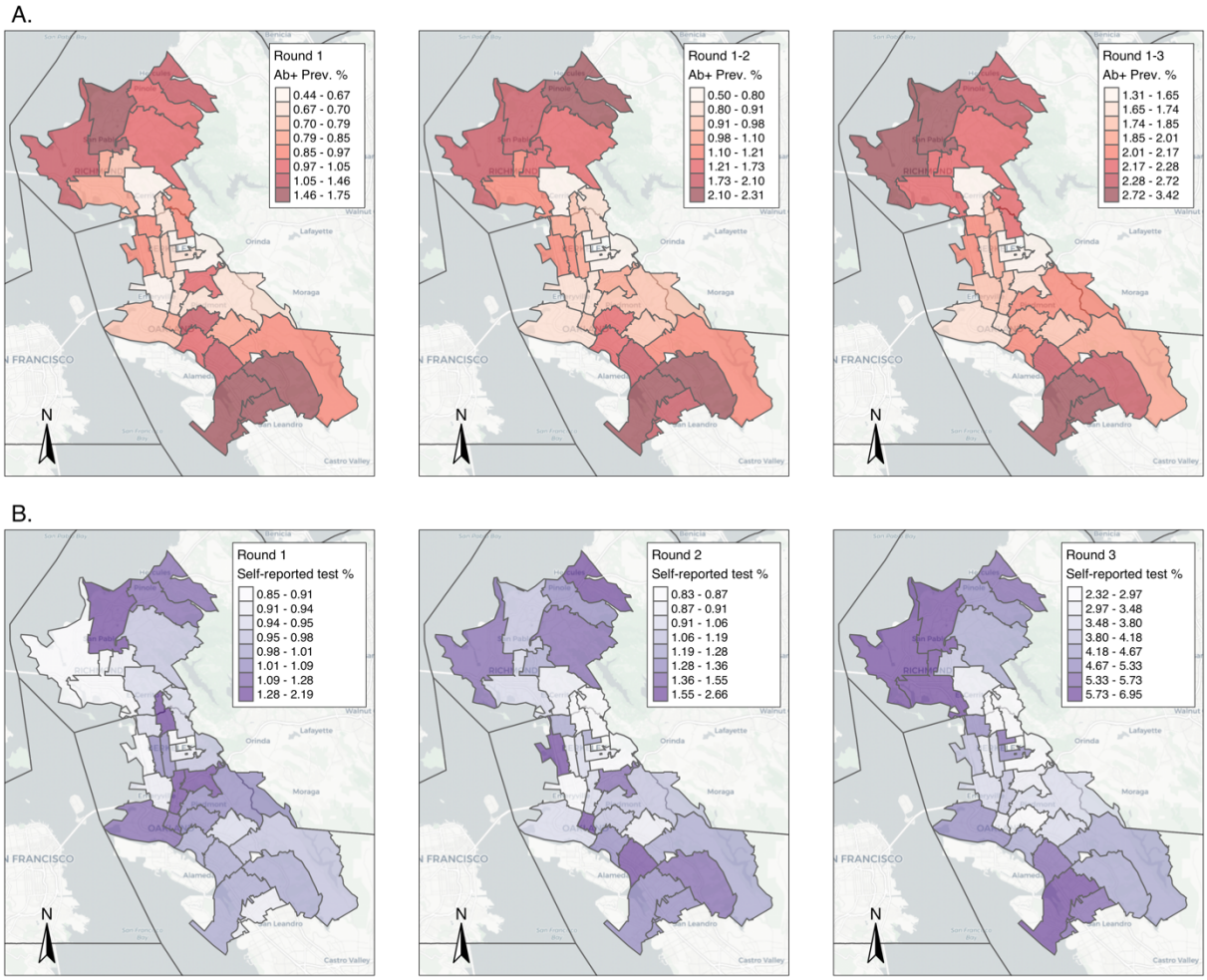

Figure 2. Population adjusted, A) cumulative seroprevalence of SARS-CoV-2 antibodies to natural infection, and B) self-reported test positivity prevalence for (July 2020-April 2021) across the study region.

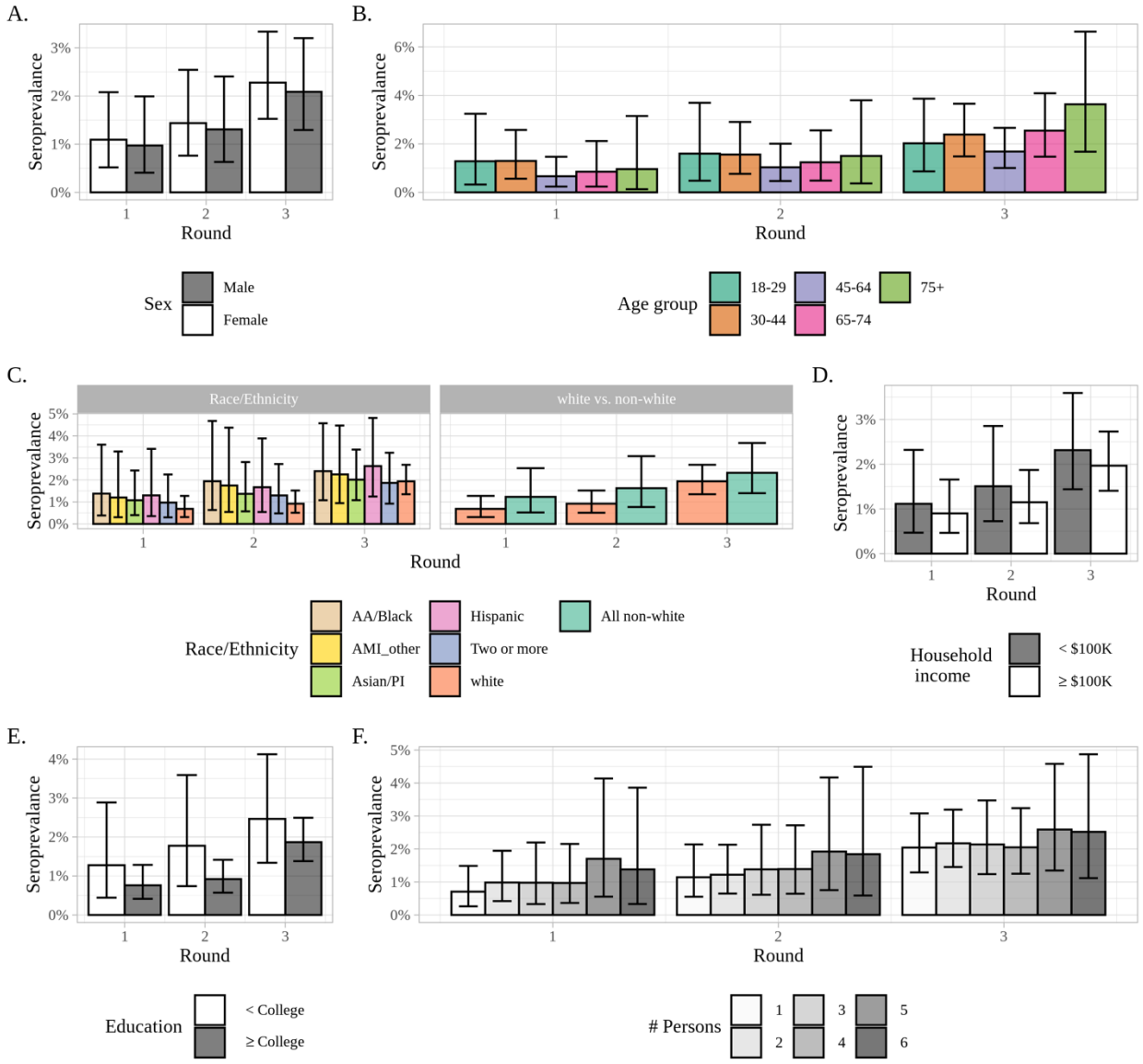

Figure 3. Cumulative seroprevalence of SARS-CoV-2 anti-S antibodies (Rounds 1 and 2) and anti-NC antibodies (Round 3) (July 2020-April 2021) among demographic subgroups, A) sex, B) age, C) race/ethnicity, D) income, E) education, and F) household size. Abbreviations: AA, African American; AMI, American Indian; PI, Pacific Islander.

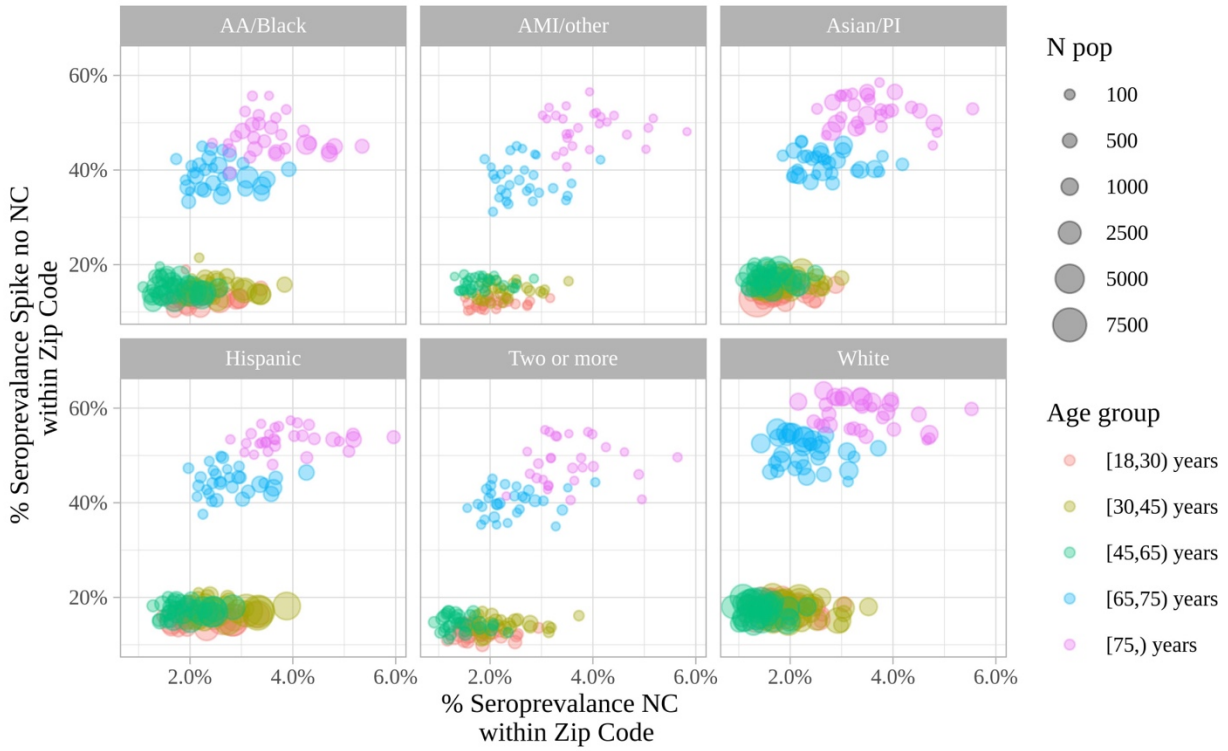

Figure 4. Round 3 population-adjusted nucleocapsid (natural infection) seroprevalence compared to spike (natural infection or vaccination) seroprevalence within strata of race/ethnicity and age. Each point represents a stratum of race/ethnicity and age within a ZIP code, the size of the point is proportional to American Community Survey population size.

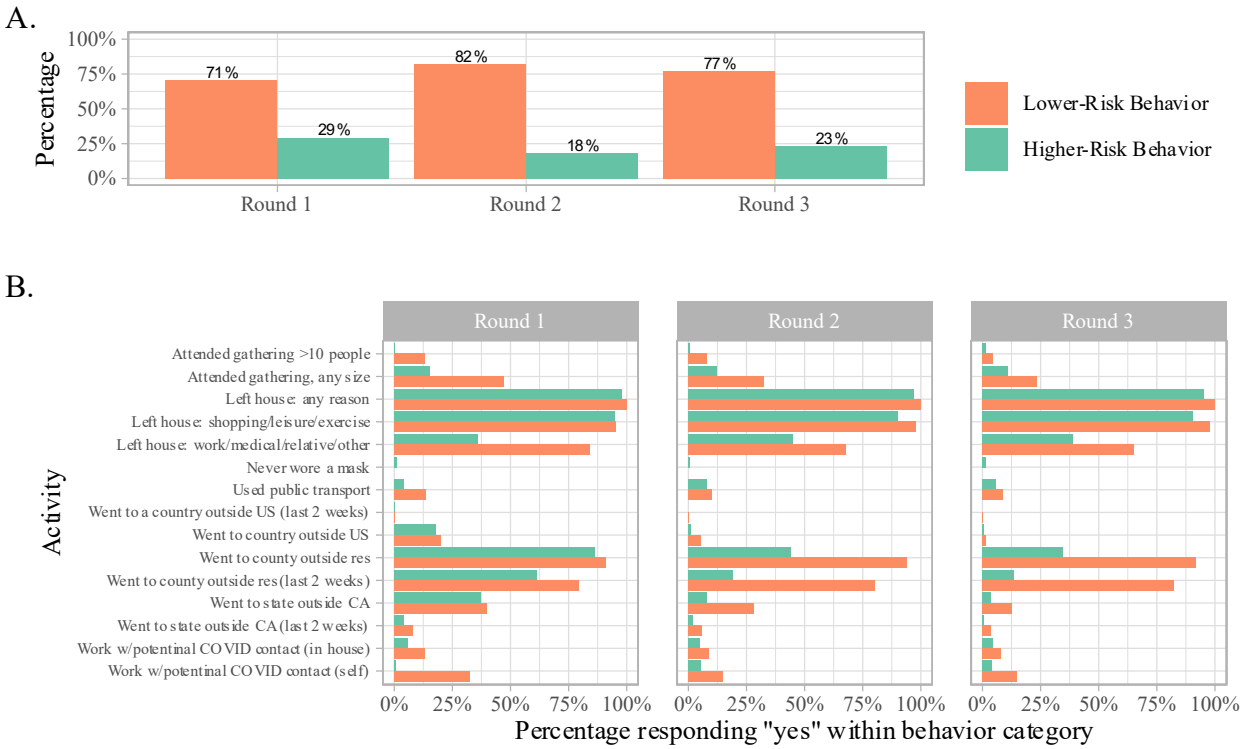

Figure 5. Proportion of participants stratified by high- and low-risk mitigation behaviors in each study round, A) overall, and B) according to specific behavior.