



OPEN

Case–control study of the association of chronic acid suppression and social determinants of health with COVID-19 infection

Bing Zhang^{1,8,13}, Anna L. Silverman^{2,9,13}, Saroja Bangaru^{3,13}, Douglas Arneson⁴, Sonya Dasharathy³, Nghia Nguyen⁵, Diane Rodden², Jonathan Shih⁶, Atul J. Butte^{4,7}, Wael Noor El-Nachef^{3,12}✉, Brigid S. Boland^{5,11}✉ & Vivek Ashok Rudrapatna^{1,4,10}✉

Acid suppressants are widely-used classes of medications linked to increased risks of aerodigestive infections. Prior studies of these medications as potentially reversible risk factors for COVID-19 have been conflicting. We aimed to determine the impact of chronic acid suppression use on COVID-19 infection risk while simultaneously evaluating the influence of social determinants of health to validate known and discover novel risk factors. We assessed the association of chronic acid suppression with incident COVID-19 in a 1:1 case–control study of 900 patients tested across three academic medical centers in California, USA. Medical comorbidities and history of chronic acid suppression use were manually extracted from health records by physicians following a pre-specified protocol. Socio-behavioral factors by geomapping publicly-available data to patient zip codes were incorporated. We identified no evidence to support an association between chronic acid suppression and COVID-19 (adjusted odds ratio 1.04, 95% CI 0.92–1.17, $P = 0.515$). However, several medical and social features were positive (Latinx ethnicity, BMI ≥ 30 , dementia, public transportation use, month of the pandemic) and negative (female sex, concurrent solid tumor, alcohol use disorder) predictors of new infection. These findings demonstrate the value of integrating publicly-available databases with medical data to identify critical features of communicable diseases.

The COVID-19 pandemic, caused by the novel severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), has affected over 200 million individuals across the globe and killed over 4.4 million as of August 2021¹. Previous studies have examined risk factors for infection and predictors of disease severity. Established medical risk factors predicting infection and severity encompass age ≥ 65 , diabetes mellitus, obesity, smoking, chronic pulmonary

¹Division of Gastroenterology and Hepatology, Department of Medicine, University of California, San Francisco, San Francisco, CA 94143, USA. ²Department of Medicine, University of California, San Diego, La Jolla, CA 92093, USA. ³Vatche and Tamar Manoukian Division of Digestive Diseases, Department of Medicine, University of California, Los Angeles, Los Angeles, CA 90095, USA. ⁴Bakar Computational Health Sciences Institute, University of California, San Francisco, San Francisco, CA 94143, USA. ⁵Division of Gastroenterology and Hepatology, Department of Medicine, University of California, San Diego, La Jolla, CA 92093, USA. ⁶Department of Neurology, University of California, San Francisco, San Francisco, CA 94143, USA. ⁷Center for Data-driven Insights and Innovation, University of California Health, Oakland, CA, USA. ⁸Division of Gastrointestinal and Liver Disease, Department of Medicine, Keck School of Medicine, University of Southern California, Los Angeles, CA 90033, USA. ⁹Division of Gastroenterology and Hepatology, Department of Medicine, Mayo Clinic, Phoenix, AZ 85054, USA. ¹⁰Bakar Computational Health Sciences Institute, University of California, 490 Illinois Street, Floor 2, Box 2933, San Francisco, CA 94143, USA. ¹¹Altman Clinical and Translational Research Institute, University of California, San Diego, 9452 Medical Center Drive, 1W 503, La Jolla, CA 92093, USA. ¹²UCLA Vatche and Tamar Manoukian Division of Digestive Diseases, 10945 Le Conte Ave #2114, Los Angeles, CA 90095, USA. ¹³These authors contributed equally: Bing Zhang, Anna L. Silverman, and Saroja Bangaru. ✉email: welnachef@mednet.ucla.edu; bboland@health.ucsd.edu; vivek.rudrapatna@ucsf.edu

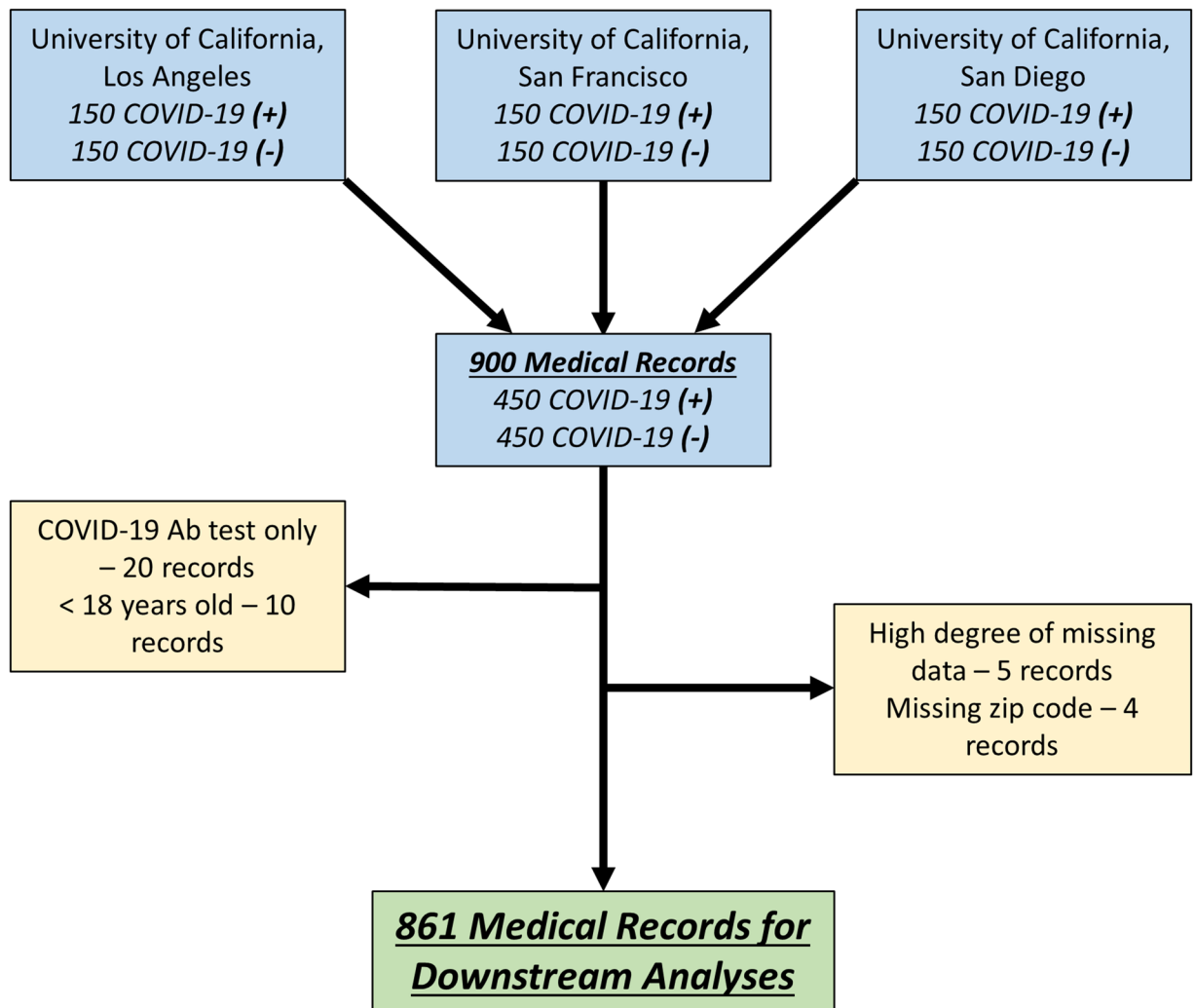


Figure 1. Selection of medical records for study. Selection of medical records was performed independently at three academic hospitals of the University of California. Each site uniformly sampled 150 COVID-19 positive (+) and 150 COVID-19 negative (–) patients. Among these, 861 medical records were retained for downstream analyses.

disease, cardiovascular disease, chronic kidney disease, malignancy and chronic HIV infection^{2–7}. The development of asthma, cardiovascular disease, hypertension, chronic kidney disease and obesity have been shown to be influenced by social determinants of health⁸. Social determinants of health comprise systemic social inequality, specifically, discrepancies in access to resources that impact health such as education, economic stability, safe living conditions and healthcare⁹.

Proton pump inhibitors (PPIs) and H2-receptor antagonists (H2RAs) are among the most commonly prescribed medications, used for gastroesophageal reflux disease (GERD) and peptic ulcer disease^{10,11}. Prior work investigating the link between PPIs and aerodigestive infections such as pneumonia has driven the initial hypothesis that PPIs could potentially compromise physiological barriers to SARS-CoV-2 infectivity^{11,12}. Although current medical literature has focused on the identification and quantification of medical-related risk factors, there is a paucity of data on how socioeconomic and behavioral influences on infection risk compare to conventional medical factors^{2–7}. Here, we sought to investigate if there is an association between chronic acid suppression and SARS-CoV-2 infection in patients seen at three urban health systems in California while accounting for concomitant social determinants of health.

Results

Of 900 medical records initially selected for manual review, twenty were excluded because they only contained the SARS-CoV-2 antibody test (not the nasopharyngeal PCR test), and ten were from patients less than 18 years old at the time of testing (Fig. 1). Five individuals with high degrees of missing data (>40%) and four with unknown residential zip codes were excluding due to inability to map geocoded social determinants of health. 861 medical records were retained for downstream analyses.

Characteristic	UCLA (N=269)	UCSD (N=297)	UCSF (N=295)	Overall (N=861)
Positive SARS-CoV2 test	136 (50.6%)	147 (49.5%)	145 (49.2%)	428 (49.7%)
SARS-CoV2 test month				
Mar	67 (24.9%)	35 (11.8%)	93 (31.5%)	195 (22.6%)
Apr	103 (38.3%)	76 (25.6%)	74 (25.1%)	253 (29.4%)
May	95 (35.3%)	146 (49.2%)	119 (40.3%)	360 (41.8%)
Jun	4 (1.5%)	40 (13.5%)	9 (3.1%)	53 (6.2%)
Female sex	150 (55.8%)	153 (51.5%)	161 (54.6%)	464 (53.9%)
Age - yr				
>65 yr	66 (24.5%)	50 (16.8%)	56 (19.0%)	172 (20.0%)
≤65 yr	203 (75.5%)	247 (83.2%)	239 (81.0%)	689 (80.0%)
Race or ethnic group				
White	163 (60.6%)	212 (71.4%)	149 (50.5%)	524 (60.9%)
Asian	22 (8.2%)	33 (11.1%)	53 (18.0%)	108 (12.5%)
Black	25 (9.3%)	12 (4.0%)	27 (9.2%)	64 (7.4%)
Other	21 (7.8%)	33 (11.1%)	49 (16.6%)	103 (12.0%)
Ethnicity Latinx	56 (20.8%)	95 (32.0%)	38 (12.9%)	189 (22.0%)
BMI				
<30	170 (63.2%)	162 (54.5%)	192 (65.1%)	524 (60.9%)
≥30	64 (23.8%)	91 (30.6%)	72 (24.4%)	227 (26.4%)
PPI use	58 (21.6%)	40 (13.5%)	33 (11.2%)	131 (15.2%)
Acid Suppression use	66 (24.5%)	55 (18.5%)	38 (12.9%)	159 (18.5%)
Charlson Comorbidity Index				
0-2: Mild	175 (65.1%)	217 (73.1%)	221 (74.9%)	613 (71.2%)
3-4: Moderate	47 (17.5%)	36 (12.1%)	41 (13.9%)	124 (14.4%)
≥5: Severe	47 (17.5%)	44 (14.8%)	33 (11.2%)	124 (14.4%)
GERD	66 (24.5%)	43 (14.5%)	45 (15.3%)	154 (17.9%)
Previous coexisting disease				
Heart Failure	12 (4.5%)	16 (5.4%)	10 (3.4%)	38 (4.4%)
Stroke	14 (5.2%)	12 (4.0%)	8 (2.7%)	34 (3.9%)
Dementia	11 (4.1%)	10 (3.4%)	4 (1.4%)	25 (2.9%)
Diabetes	52 (19.3%)	32 (10.8%)	34 (11.5%)	118 (13.7%)
Severe CKD	17 (6.3%)	2 (0.7%)	7 (2.4%)	26 (3.0%)
Solid tumor	29 (10.8%)	27 (9.1%)	21 (7.1%)	77 (8.9%)
AIDS	2 (0.7%)	5 (1.7%)	2 (0.7%)	9 (1.0%)
HIV	7 (2.6%)	17 (5.7%)	14 (4.7%)	38 (4.4%)
Immunocompromised	78 (29.0%)	56 (18.9%)	55 (18.6%)	189 (22.0%)
Autoimmune disease	11 (4.1%)	16 (5.4%)	20 (6.8%)	47 (5.5%)
Chronic opiate use	13 (4.8%)	30 (10.1%)	9 (3.1%)	52 (6.0%)
EtOH use disorder	8 (3.0%)	13 (4.4%)	11 (3.7%)	32 (3.7%)
Healthcare worker	31 (11.5%)	64 (21.5%)	65 (22.0%)	160 (18.6%)
Residence				
Single Family Home	160 (59.5%)	207 (69.7%)	173 (58.6%)	540 (62.7%)
Apartment	91 (33.8%)	75 (25.3%)	104 (35.3%)	270 (31.4%)
SNF	12 (4.5%)	7 (2.4%)	6 (2.0%)	25 (2.9%)
Homeless	2 (0.7%)	5 (1.7%)	8 (2.7%)	15 (1.7%)
Cruise	0 (0%)	1 (0.3%)	0 (0%)	1 (0.1%)
Social Deprivation Index				
>65	128 (47.6%)	130 (43.8%)	94 (31.9%)	352 (40.9%)
Time spent home				
>85.0%	250 (92.9%)	266 (89.6%)	253 (85.8%)	769 (89.3%)
Public transportation usage				
>4.4%	132 (49.1%)	106 (35.7%)	256 (86.8%)	494 (57.4%)

Figure 2. Medical and sociodemographic characteristics of medical records included in the analysis. *BMI* body mass index, *GERD* gastroesophageal reflux disease, *CKD* chronic kidney disease, *SNF* skilled nursing facility.

The number of included medical records was similar across the three sites. In total there were 428 (49.7%) documented positive tests (Fig. 2). Approximately half were female, and one-fifth were ≥65 years old. The majority self-identified as white, and 22% of the total sample identified as Latinx. Most individuals were calculated to have a low Charlson Comorbidity Index (CCI; 0–2) and were not obese based on body mass index (BMI). Diabetes (13.7%) and immunocompromised status (22.0%) were the most prevalent comorbidities. Nearly one-fifth tested were healthcare workers, and the majority of people lived in single family homes. During the period

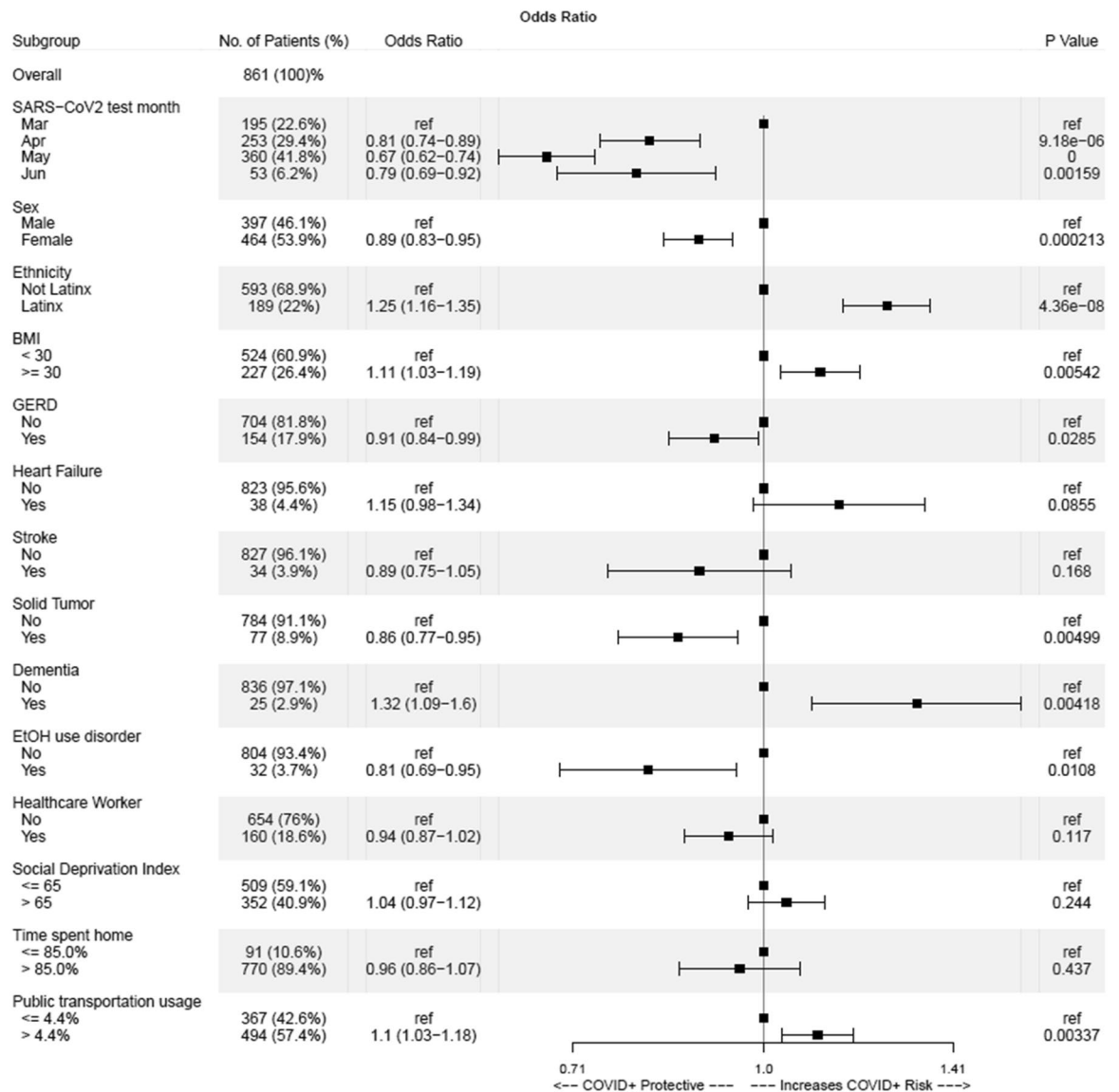


Figure 3. Features identified as correlating with the SARS-CoV-2 testing result. Adjusted odds ratios and P-values were calculated when compared to baseline (ref).

sampled, the probability of a positive SARS-CoV2 test decreased significantly after March (adjusted odds ratios of testing positive for SARS-CoV2 of 0.81, 0.67, and 0.79 in April, May, and June respectively), likely reflecting the adoption of widespread testing (Fig. 3).

Chronic acid suppression and SARS-CoV-2 risk. To evaluate the effect of chronic acid suppression on SARS-CoV-2 infection risk, logistic model selection using 62 candidate features including chronic PPI use and chronic H2RA use was performed. This process identified neither medication class as significant influences on infection risk (Fig. 3). This finding was unchanged even after manually incorporating chronic acid suppression use (combination of both PPI and H2RA) into the model (adjusted odds ratio 1.04, 95% CI 0.92 to 1.17, $P=0.515$; Supplementary Fig. S1). Interestingly, GERD was associated with a slightly lower odds of COVID-19 in both models.

Medical conditions associated with SARS-CoV-2 risk. Several medical comorbidities were identified as significant predictors of incident infection. These included a BMI ≥ 30 (adjusted odds ratio 1.11, 95% CI 1.03 to 1.19, $P=0.0054$; Fig. 3), and dementia (adjusted odds ratio 1.32, 95% CI 1.09 to 1.6, $P=0.0042$; Fig. 3). Unexpectedly, non-hematogenous malignancy (adjusted odds ratio 0.86, 95% CI 0.77 to 0.95, $P=0.0050$; Fig. 3) and history of alcoholic use disorder (adjusted odds ratio 0.81, 95% CI 0.69 to 0.95, $P=0.0108$; Fig. 3) were associated with a negative test. Other comorbidities including heart failure and stroke were not found to be statistically significant predictors of COVID-19.

Sociodemographic factors and SARS-CoV-2 risk. Females (adjusted odds ratio 0.89, 95% CI 0.83–0.95, $P = 0.000213$; Fig. 3) were less likely to test positive, but individuals of Latinx ethnicity (adjusted odds ratio 1.25, 95% CI 1.16 to 1.35, $P = 4.36 \times 10^{-8}$; Fig. 3) were more likely to have positive result. Finally, those who resided in localities with more than 4.4% of the adult population utilizing public transportation as their main means of commute also were more likely to test positive (adjusted odds ratio 1.1, 95% CI 1.03 to 1.18, $P = 0.0034$; Fig. 3). Among our sample, healthcare worker status did not strongly correlate with testing outcome, and neither did “social deprivation index” and “time spent at home” (Fig. 3).

Discussion

We identified no association between long-term PPI or H2RA use and COVID-19 after controlling for medical and sociodemographic predictors of infection risk. Our results are consistent with, and complement, two previously published studies utilizing a clinical informatics approach. A case–control study using South Korean health insurance claims reported no association between current and past use of PPIs and COVID-19¹³. A cohort study using the United Kingdom’s Biobank also reported no increased risk of COVID-19 among patients who reported taking PPI or H2RA¹⁴. Of note, medication data collection period was 2006–2010 for this study, therefore the study investigators could not confirm that participants were still taking these medications following COVID-19 outbreak. By contrast, a single large United States (US) survey-based study found a positive association between adults self-reporting COVID-19 and daily PPI use, but did not find an association with H2RA¹⁵. A notable limitation of this study is the reliance of self-reported COVID-19 when testing was available. The study reported a COVID-19 prevalence of 6.4% which was higher than the national estimates at the time, raising concerns regarding the accuracy of self-report. Furthermore, the survey was administered in English, functionally excluding the near quarter of US Census respondents who report limited to no English proficiency¹⁶.

Consistent with previous publications, our data found an increased risk of COVID-19 and a BMI ≥ 30 ^{17–21}. Furthermore, patients with dementia were more likely to test positive, possibly reflecting frequent interactions with the healthcare system or caregivers and a decreased ability to socially distance. Interestingly, those with underlying solid tumors or alcohol use disorder had a lower risk of infection with SARS-CoV-2. We surmised that this may be due to reduced exposure to infected individuals by spending significant time at home and adhering more strictly to precautions such as mask wearing, handwashing and social distancing, but still having a high need for medical services and therefore receiving more testing. Finally, the negative association between COVID-19 positivity and GERD may reflect functional heartburn, which is comorbid with anxiety and possible proclivity to be tested²².

An important finding of our study is that members of the Latinx community are at a disproportionate risk of contracting SARS-CoV-2, even after controlling for a variety of medical, socioeconomic and behavioral factors. Current literature in the US and globally has commented mainly on the increased risk of infection, severe disease and death in black patients^{23–30}. Our California-based cohort showed that being of Latinx ethnicity is associated with higher rates of infection with SARS-CoV-2. In addition to data from New York City reporting 34% of COVID-19 deaths in Latinx people despite only representing 29% of the population, our California-based cohort enriched in Latinx patients highlights increased risk^{30, 31}. In the 45 states reporting data by ethnic group, 20 report the proportion of cases among Latinx people is twice as high and 11 report it is three times as high as would be expected based on the general population^{32, 33}. These collective data underscore the substantial vulnerability of this ethnic group.

Patient zip codes and geolocation data provided indirect insight into the physical environments and behaviors of the communities in which the patients reside and evaluated effects of social determinants of health on COVID-19 risk. We found patients living in zip codes with increased public transportation reliance were associated with SARS-CoV-2 positivity. Since the ability to work from home, avoid public transportation, financially accommodate for furloughs from work are not equally distributed, the ability to social distance effectively may be one of privilege^{30, 33, 34}. Although these differences in social determinants of health are receiving specific attention in the current pandemic, they speak to important and systematic disparities across healthcare that need to be better understood and addressed. In our cohort, females had a significantly lower probability of infection with SARS-CoV-2. It is currently unclear if there are immunologic or behavioral explanations why female sex may be protective from infection³⁵. It has been proposed that sex hormone regulation of angiotensin-converting enzyme 2 and transmembrane serine protease-2 may be implicated; however, future work is needed to clarify this finding³⁶.

Our study has several strengths. The multi-center design incorporates patients from three large urban academic health systems spanning Northern and Southern California providing care for a diverse patient cohort. Data included confirmed SARS-CoV-2 by PCR test results in addition to medications and comorbidities data collected and curated by physicians to improve accuracy. Application of geolocation using zip code enabled identification of sociodemographic determinants of infection risk. An important limitation of our study is possible selection bias. Our study cohort consisted only of tested patients. Attempts to correct this included incorporating multiple predictors of the probability of testing as well as the probability of infection itself; nonetheless we could not exclude residual bias. Other potential limitations include possible measurement bias (e.g. discordance between EHR documentation and actual patient use of medications), ecological fallacy (i.e. attribution of locality-level data to individuals), and cohort unrepresentativeness (i.e. differences between urban, tertiary-care cohorts and the general US population). It should be noted that testing was performed both for symptom-triggered testing and pre-procedural or hospitalization screening. Thus, the pretest probability was not homogeneous. Associations identified in this observational study should not be assumed causal and require prospective validation.

This multicenter case–control study demonstrated no association between chronic acid suppression use and risk of COVID-19. In addition, we found that Latinx ethnicity and residing in a zip code with high public transportation usage were associated with increased risk of COVID-19 which is in keeping with growing evidence that

this pandemic has disproportionately affected vulnerable communities in our society. Further work is needed to address reducing the burden of the SARS-CoV-2 pandemic on these communities.

Methods

Study design. This is a case–control study involving the review of medical records from three campuses of the University of California (UC) Health system: San Francisco, San Diego, and Los Angeles. The study was approved by the Institutional Review Board (IRB) of the University of California, San Francisco (UCSF; IRB #20-30549), which was accepted and approved by the IRB of the University of California, San Diego (UCSD) via a memorandum of understanding (UC IRB Reliance #3499). Informed consent was exempted by the IRB of UCSF and UCSD. The IRB of the University of California, Los Angeles (UCLA) determined that the study did not constitute human subjects research and therefore waived approval and informed consent. The study was performed in accordance with relevant guidelines and regulations of the three respective academic medical institutions and their IRB committees. The STROBE guideline for case–control studies was referenced for study design and reporting the findings (Supplementary Table S1)³⁷.

At the time of the study's inception in May 2020, de-identified databases of electronic health records (EHR) data at UC Health were queried to support sample size calculations. At that time, the estimated prevalence of positivity for SARS-CoV-2 virus was 5.6% among those tested across all campuses of UC Health. At UCSF, the prevalence of PPI use was estimated to be 7.5%. We obtained approximate effect sizes from a prior study that assessed the relative risk of acute gastroenteritis among chronic PPI users, motivated in part by work suggesting the possible enteric transmission of the SARS-CoV-2 virus^{38, 39}. We calculated that at least 416 cases and controls would be needed to detect two-fold relative risk among chronic PPI users compared to non-users with 80% power. These numbers were rounded up to obtain 450 cases and controls each, or 900 medical records in total. These were divided evenly across all three participating sites (300 records each).

Patient selection and data collection. Records of all individuals who underwent SARS-CoV-2 nasopharyngeal PCR testing at any of the three health systems (UCSF, UCSD, UCLA) from March 1, 2020 to June 10, 2020 were separately retrieved via EHR database query. Stratified uniform sampling was performed to yield 150 cases and 150 controls for manual review at each site. A standardized case report form and accompanying data dictionary were developed. Data elements included PPI and H2RA use, Charlson Comorbidity Index (CCI), additional medical comorbidities and immunocompromised states, and sociodemographic factors (Supplementary Methods). Variable definitions and data extraction process was determined and fixed a priori and based on consensus among the co-authors.

For eligibility criteria of both cases and controls (collectively referred to as patients), medical records of all patients ≥ 18 years old at the time of their initial SARS-CoV-2 test were included in the study database. Result of the SARS-CoV-2 PCR testing was determined by documentation in the EHR of each site. For patients who were tested more than once during the data collection period and were consistently negative, the date of the first result was recorded. Patients who tested positive after initial negative test(s) were recorded as positive, and the date of the first positive test was recorded. Chronic acid suppression was defined as the use of PPI or H2RA daily or as needed for at least four weeks, documented by medical providers in charts or prescription refill records. The decision for this is based on studies demonstrating the incubation period of SARS-CoV-2 to possibly be two to three weeks or more, thus a minimal of four weeks ensures adequate acid suppression prior to SARS-CoV-2 exposure^{40, 41}.

All medical history was reviewed and annotated by physicians (BZ, AS, SB, SD, DR). Disagreements during extraction pertaining to the coding of specific variables were adjudicated through discussion and resolved through consensus voting (Supplementary Methods). The extraction sheet was modified once to better capture race and ethnicity. Following the completion of data abstraction, datasheets were compiled and subject to quality control measures via both manual (JS) and informatics methods in order to ensure consistent and accurate data capture across all three sites.

Geomapping social determinants of health. Patient records were mapped to geographic areas using postal zip codes corresponding to their home address. However, most geocoded data including reports from the United States Census utilize zip code tabulation areas (ZCTAs). Postal zip codes were mapped to zip code tabulation areas (ZCTAs), with the method implemented by Williams-Holt^{42–44}.

Geocoded covariates related to population and housing density were obtained from the 2010 United States Decennial Census⁴⁵. Additional geocoded covariates were obtained from the 2018 United States American Community Survey (ACS) using the R package *tidycensus* v0.9.9.2⁴⁶. For each ZCTA, median household income, public transportation as means of transportation to work for workers 16 and over, and walking as means of transportation to work for workers 16 and over were extracted.

The United States Census Bureau provides Community Resilience Estimates (CREs) at the census tract level defining what percentage of residents in a census tract have 3 + risk factors, 1–2 risk factors, or 0 risk factors⁴⁷. These risk factors are derived from the 2018 American Community Survey (ACS) and the 2018 National Health Interview Survey (NHIS).

Mask-wearing by county was obtained from a survey conducted by The New York Times⁴⁸. Briefly, survey responses were collected from 250,000 individuals between 7/2/2020 and 7/14/2020 who were asked: “How often do you wear a mask in public when you expect to be within six feet of another person?” The possible responses were: never, rarely, sometimes, frequently, and always.

Daily social distancing mobility data was obtained from the data company SafeGraph^{49, 50}. Briefly, daily anonymized and aggregated mobile device data were obtained from SafeGraph from 2-12-2020 through

7/12/2020. Mobile devices were aggregated by Census Block Groups (CBGs) based on home location and the median percentage of time a device was observed at home versus observed at all during a given time period was calculated for all observed devices in a CBG each day.

SARS-CoV-2 Community Mobility Reports were obtained from Google from February 2020 through July 2020⁵¹. Briefly, these reports quantify how visits and length of stay at various locations changed compared to a baseline using aggregated and anonymized data from users who opted into Google Location History. Google reported daily values for time spent at the following locations aggregated at the county level: grocery & pharmacy, parks, transit stations, retail & recreation, residential, and workplaces. The time spent at a location was represented as a percent change of time spent at the location compared to the baseline.

The Social Deprivation Index (SDI) has previously been used to quantify levels of disadvantage and evaluate their associations with health outcomes⁵². The SDI used in this study was developed by the Robert Graham Center and is comprised of seven demographic characteristics from the 2015 ACS⁵³. These characteristics include: income, education, employment, housing, household characteristics, transportation, and demographics. Reported SDI measures for ZCTAs were used in this analysis.

Outcomes. The primary outcome was association between chronic acid suppression (PPIs, H2RAs) and risk of SARS-CoV-2 infection. Secondary outcomes included the association between SARS-CoV-2 infection and the month of testing, demographics, comorbid medical conditions, immunocompromised state, CCI, and social and mobility variables (Supplementary Table S2).

Statistical analysis. Medical records and covariates with a high degree of missing data (>40%) were excluded from subsequent analysis. Multiple imputation was then performed using the R package *mice* v3.11.0 to impute 10 datasets with 20 iterations per dataset to ensure convergence⁵⁴.

Continuous covariates were converted to binary variables for interpretability in the logistic regression model and corresponding forest plot. To identify binary cut points which may be associated with an increased chance of a positive SARS-CoV-2 test, we conducted a grid search of all possible values between the 10th and 90th percentiles for each continuous covariate. The binary cut point associated with the lowest p-value resulting from a chi-square test was used to binarize each continuous covariate.

Prior to regression analysis, chi-squared and ANOVA tests were performed where appropriate to test for associations between all variables and positivity for SARS-CoV2 (Supplementary Table S3). Subsequently, to identify and select the subset of input covariates that were most relevant to a positive SARS-CoV-2 test, a multistep process was employed. First, a stepwise linear model to predict a positive SARS-CoV-2 test was fit separately for each of the 10 imputed datasets. Both forward and backward selection using the Akaike Information Criterion were used to identify the subset of covariates to be included in the final model for each imputed dataset. Next, covariates which were included in all 10 reduced models were chosen to be included in the final pooled model. Covariates which appeared in $\geq 50\%$ of the reduced models were considered individually for inclusion in the final pooled model with a Wald test. The selected covariates were used in the final reduced logistic regression model. We initially fit a mixed-effects model on the combined dataset using study site as a random intercept⁵⁵. These results suggested a low degree of heterogeneity attributable to study site (i^2 statistic = 0%). Consequently, all subsequent results were obtained by fixed-effect (pooled) models.

Data availability

The data used for this study contains protected health information, including dates and zip codes. As such, the underlying data have not been made available for reuse. However, the analytic codes along with randomly generated demo data and detailed tutorial have been made publicly available on github at: <https://github.com/darne/son/CovidPPI>. Datasets supporting analytical workflow of this manuscript is available at: https://figshare.com/articles/dataset/Datasets_supporting_analytical_workflow_of_Chronic_Acid_Suppression_and_Social_Determinants_of_COVID-19_Infection/13380356.

Received: 27 March 2021; Accepted: 28 September 2021

Published online: 25 October 2021

References

1. Coronavirus Resource Center. *COVID-19 dashboard by the Center for Systems Science and Engineering (CCSE) at Johns Hopkins University (JHU)*. <https://coronavirus.jhu.edu/map.html>. Accessed August 25, 2021 (2021).
2. Blanco, J. L. *et al.* COVID-19 in patients with HIV: Clinical case series. *Lancet HIV* **7**, e314–e316. [https://doi.org/10.1016/S2352-3018\(20\)30111-9](https://doi.org/10.1016/S2352-3018(20)30111-9) (2020).
3. Dai, M. *et al.* Patients with cancer appear more vulnerable to SARS-CoV-2: A multicenter study during the COVID-19 outbreak. *Cancer Discov.* **10**, 783–791. <https://doi.org/10.1158/2159-8290.CD-20-0422> (2020).
4. Grasselli, G. *et al.* Baseline characteristics and outcomes of 1591 patients infected with SARS-CoV-2 admitted to ICUs of the Lombardy region, Italy. *JAMA* **323**, 1574–1581. <https://doi.org/10.1001/jama.2020.5394> (2020).
5. Guan, W. J. *et al.* Comorbidity and its impact on 1590 patients with COVID-19 in China: A nationwide analysis. *Eur. Respir. J.* **55**, 2000547. <https://doi.org/10.1183/13993003.00547-2020> (2020).
6. Lighter, J. *et al.* Obesity in patients younger than 60 years is a risk factor for COVID-19 hospital admission. *Clin. Infect. Dis.* **71**, 896–897. <https://doi.org/10.1093/cid/ciaa415> (2020).
7. Park, S. E. Epidemiology, virology, and clinical features of severe acute respiratory syndrome-coronavirus-2 (SARS-CoV-2; Coronavirus Disease-19). *Clin. Exp. Pediatr.* **63**, 119–124. <https://doi.org/10.3345/cep.2020.00493> (2020).
8. Singu, S., Acharya, A., Challagundla, K. & Byrreddy, S. N. Impact of Social Determinants of Health on the Emerging COVID-19 Pandemic in the United States. *Front. Public Health* **8**, 406. <https://doi.org/10.3389/fpubh.2020.00406> (2020).
9. Abrams, E. M. & Szefler, S. J. COVID-19 and the impact of social determinants of health. *Lancet Respir. Med.* **8**, 659–661. [https://doi.org/10.1016/S2213-2600\(20\)30234-4](https://doi.org/10.1016/S2213-2600(20)30234-4) (2020).

10. Kantor, E. D., Rehm, C. D., Haas, J. S., Chan, A. T. & Giovannucci, E. L. Trends in prescription drug use among adults in the United States from 1999–2012. *JAMA* **314**, 1818–1831. <https://doi.org/10.1001/jama.2015.13766> (2015).
11. Vaezi, M. F., Yang, Y. X. & Howden, C. W. Complications of proton pump inhibitor therapy. *Gastroenterology* **153**, 35–48. <https://doi.org/10.1053/j.gastro.2017.04.047> (2017).
12. Laheij, R. J. *et al.* Risk of community-acquired pneumonia and use of gastric acid-suppressive drugs. *JAMA* **292**, 1955–1960. <https://doi.org/10.1001/jama.292.16.1955> (2004).
13. Lee, S. W. *et al.* Severe clinical outcomes of COVID-19 associated with proton pump inhibitors: A nationwide cohort study with propensity score matching. *Gut* **70**, 76–84. <https://doi.org/10.1136/gutjnl-2020-322248> (2020).
14. Fan, X. *et al.* Effect of acid suppressants on the risk of COVID-19: A propensity score-matched study using UK Biobank. *Gastroenterology* **169**, 455–458.e5. <https://doi.org/10.1053/j.gastro.2020.09.028> (2020).
15. Almario, C. V., Chey, W. D. & Spiegel, B. M. R. Increased risk of COVID-19 among users of proton pump inhibitors. *Am. J. Gastroenterol.* **115**, 1707–1715. <https://doi.org/10.14309/ajg.0000000000000798> (2020).
16. Ryan, C. *Language use in the United States: 2011*. <https://www.census.gov/library/publications/2013/acs/acs-22.html> (2010). Accessed November 5, 2020.
17. Colaneri, M. *et al.* The experience of the health care workers of a severely hit SARS-CoV-2 referral Hospital in Italy: Incidence, clinical course and modifiable risk factors for COVID-19 infection. *J. Public Health (Oxf)*. <https://doi.org/10.1093/pubmed/fdaa195> (2020).
18. Omrani, A. S. *et al.* The first consecutive 5000 patients with Coronavirus Disease 2019 from Qatar; a nation-wide cohort study. *BMC Infect. Dis.* **20**, 777. <https://doi.org/10.1186/s12879-020-05511-8> (2020).
19. Ranjan, P. *et al.* Is excess weight a risk factor for the development of COVID 19 infection? A preliminary report from India. *Diabetes Metab. Syndr.* **14**, 1805–1807. <https://doi.org/10.1016/j.dsx.2020.09.012> (2020).
20. Razieh, C., Zaccardi, F., Davies, M. J., Khunti, K. & Yates, T. Body mass index and the risk of COVID-19 across ethnic groups: Analysis of UK Biobank. *Diabetes Obes. Metab.* **22**, 1953–1954. <https://doi.org/10.1111/dom.14125> (2020).
21. Sattar, N. *et al.* BMI and future risk for COVID-19 infection and death across sex, age and ethnicity: Preliminary findings from UK biobank. *Diabetes Metab. Syndr.* **14**, 1149–1151. <https://doi.org/10.1016/j.dsx.2020.06.060> (2020).
22. Fass, R. & Tougas, G. Functional heartburn: The stimulus, the pain, and the brain. *Gut* **51**, 885–892. <https://doi.org/10.1136/gut.51.6.885> (2002).
23. Bailey, Z. D. & Moon, J. R. Racism and the political economy of COVID-19: Will we continue to resurrect the past?. *J. Health Polit. Policy Law* **45**, 937–950. <https://doi.org/10.1215/03616878-8641481> (2020).
24. Zelner, J. *et al.* Racial disparities in COVID-19 mortality are driven by unequal infection risks. *Clin. Infect. Dis.* **72**, e88–e95. <https://doi.org/10.1093/cid/ciaa1723> (2020).
25. Mahajan, U. V. & Larkins-Pettigrew, M. Racial demographics and COVID-19 confirmed cases and deaths: A correlational analysis of 2886 US counties. *J. Public Health (Oxf)* **42**, 445–447. <https://doi.org/10.1093/pubmed/fdaa070> (2020).
26. Meyerowitz-Katz, G. & Merone, L. A systematic review and meta-analysis of published research data on COVID-19 infection fatality rates. *Int. J. Infect. Dis.* **101**, 138–148. <https://doi.org/10.1016/j.ijid.2020.09.1464> (2020).
27. Nepomuceno, M. R. *et al.* Besides population age structure, health and other demographic factors can contribute to understanding the COVID-19 burden. *Proc. Natl. Acad. Sci. U S A* **117**, 13881–13883. <https://doi.org/10.1073/pnas.2008760117> (2020).
28. Parcha, V. *et al.* Geographic variation of racial disparities in health and COVID-19 mortality. *Mayo Clin. Proc. Innov. Qual. Outcomes* **4**, 703–716. <https://doi.org/10.1016/j.mayocpiqo.2020.09.005> (2020).
29. Price-Haywood, E. G., Burton, J., Fort, D. & Seoane, L. Hospitalization and mortality among black patients and white patients with Covid-19. *N. Engl. J. Med.* **382**, 2534–2543. <https://doi.org/10.1056/NEJMsa2011686> (2020).
30. Yancy, C. W. COVID-19 and African Americans. *JAMA* **323**, 1891–1892. <https://doi.org/10.1001/jama.2020.6548> (2020).
31. Workbook. *NYS-COVID19-Tracker*. <https://covid19tracker.health.ny.gov> (2020). Accessed November 1, 2020.
32. The Atlantic. *The COVID Racial Data Tracker*. <https://covidtracking.com/race> (2021). Accessed October 15, 2020.
33. McClure, E. S., Vasudevan, P., Bailey, Z., Patel, S. & Robinson, W. R. Racial capitalism within public health-how occupational settings drive COVID-19 disparities. *Am. J. Epidemiol.* **189**, 1244–1253. <https://doi.org/10.1093/aje/kwaa126> (2020).
34. Page, K. R. & Flores-Miller, A. Lessons we've learned - Covid-19 and the undocumented Latinx community. *N. Engl. J. Med.* **384**, 5–7. <https://doi.org/10.1056/NEJMp2024897> (2020).
35. Kelada, M., Anto, A., Dave, K. & Saleh, S. N. The role of sex in the risk of mortality from COVID-19 amongst adult patients: A systematic review. *Cureus* **12**, e10114. <https://doi.org/10.7759/cureus.10114> (2020).
36. Foresta, C., Rocca, M. S. & Di Nisio, A. Gender susceptibility to COVID-19: A review of the putative role of sex hormones and X chromosome. *J. Endocrinol. Invest.* **44**, 951–956. <https://doi.org/10.1007/s40618-020-01383-6> (2021).
37. von Elm, E. *et al.* The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement: Guidelines for reporting observational studies. *J. Clin. Epidemiol.* **61**, 344–349. <https://doi.org/10.1016/j.jclinepi.2007.11.008> (2008).
38. Ding, S. & Liang, T. J. Is SARS-CoV-2 also an enteric pathogen with potential fecal-oral transmission? A COVID-19 virological and clinical review. *Gastroenterology* **159**, 53–61. <https://doi.org/10.1053/j.gastro.2020.04.052> (2020).
39. Vilcu, A. M. *et al.* Association between acute gastroenteritis and continuous use of proton pump inhibitors during winter periods of highest circulation of enteric viruses. *JAMA Netw. Open* **2**, e1916205. <https://doi.org/10.1001/jamanetworkopen.2019.16205> (2019).
40. Hu, B., Guo, H., Zhou, P. & Shi, Z. L. Characteristics of SARS-CoV-2 and COVID-19. *Nat. Rev. Microbiol.* **19**, 141–154. <https://doi.org/10.1038/s41579-020-00459-7> (2021).
41. Lauer, S. A. *et al.* The incubation period of coronavirus disease 2019 (COVID-19) from publicly reported confirmed cases: Estimation and application. *Ann. Intern. Med.* **172**, 577–582. <https://doi.org/10.7326/M20-0504> (2020).
42. *ZIP Code Tabulation Areas (ZCTAs)*. United States Census Bureau. [https://www.census.gov/programs-surveys/geography/guidance/geo-areas/zctas.html#:~:text=ZIP%20Code%20Tabulation%20Areas%20\(ZCTAs\)%20are%20generalized%20areal%20representations%20of,station%20associated%20with%20mailing%20addresses](https://www.census.gov/programs-surveys/geography/guidance/geo-areas/zctas.html#:~:text=ZIP%20Code%20Tabulation%20Areas%20(ZCTAs)%20are%20generalized%20areal%20representations%20of,station%20associated%20with%20mailing%20addresses). Accessed October 15, 2020.
43. UDS Mapper. *ZIP Code to ZCTA Crosswalk*. <https://udsmapper.org/zip-code-to-zcta-crosswalk/>. Accessed October 15, 2020.
44. Williams-Holt, J. *Spatial Crosswalks*. <https://github.com/jwilliamsholt/safegraph-in-r/blob/master/spatial-crosswalks.Rmd>. Accessed October 10, 2020.
45. *ZCTA to County Relationship File, 2010 United States Decennial Census*. United States Census Bureau. https://www.census.gov/geographies/reference-files/time-series/geo/relationship-files.html#par_textimage_674173622. Accessed October 15, 2020.
46. Walker K. *Tidycensus: Load US Census Boundary and Attribute Data as 'tidyverse' and 'sf': Ready Data Frames*. <https://CRAN.R-project.org/package=tidycensus>. Accessed October 15, 2020.
47. Community Resilience Estimates. United States Census Bureau. <https://www.census.gov/data/experimental-data-products/community-resilience-estimates.html>. Accessed October 15, 2020.
48. The New York Times. *Mask-Wearing Survey Data*. <https://github.com/nytimes/covid-19-data/tree/master/mask-use>. Accessed October 20, 2020.
49. SafeGraph. *COVID-19 Data Consortium*. <https://www.safegraph.com/covid-19-data-consortium>. Accessed October 20, 2020.
50. SafeGraph. *Social Distancing Metrics*. <https://docs.safegraph.com/docs/social-distancing-metrics>. Accessed, October 20, 2020.
51. Google LLC. *COVID-19 Community Mobility Reports*. <https://www.google.com/covid19/mobility/>. Accessed October 25, 2020.

52. Butler, D. C., Petterson, S., Phillips, R. L. & Bazemore, A. W. Measures of social deprivation that predict health care access and need within a rational area of primary care service delivery. *Health Serv. Res.* **48**, 539–559. <https://doi.org/10.1111/j.1475-6773.2012.01449.x> (2013).
53. Robert Graham Center. *Social Deprivation Index (SDI)*. <https://www.graham-center.org/rgc/maps-data-tools/sdi/social-deprivation-index.html>. Accessed October 25, 2020.
54. Stef van Buuren, K.G.-O. mice: Multivariate imputation by chained equations in R. *J. Stat. Softw.* **45**, 1–67 (2011).
55. Bates, D., Machler, M., Bolker, B. M. & Walker, S. C. Fitting linear mixed-effects models using lme4. *J. Stat. Softw.* **67**, 1–48. <https://doi.org/10.18637/jss.v067.i01> (2015).

Acknowledgements

This study was supported by funding from the UCSF Bakar Computational Health Science Institute and the National Center for Advancing Translational Sciences of the NIH, grant number UL1TR001872 (UCSF), UL1TR001881 (UCLA), and UL1TR001442 (UCSD). UCLA's effort was supported by the Barbara and Joel Marcus UCLA GI Fellowship Research Seed Grant Program. BZ is supported by NIH/NIDDK, grant number T32DK007007-45. NN is supported by NIH/NIDDK, grant number T32DK007202. BSB is supported by NIH/NIDDK, grant number K23DK123406. VAR was supported by NIH/National Center for Advancing Translational Sciences, grant number TL1TR001871. The authors acknowledge the US Census, The New York Times, SafeGraph, and Google LLC for providing data relevant to this study. They thank the leadership and governance of the UC Office of the President and UC Health, as well as the Clinical and Translational Science Institutes of all member campuses for their support of the clinical informatics infrastructure necessary for this work (CORDS, UCSF COVID-19 Research Data Mart). The acknowledged organizations and funding sources had no involvement in any aspects of the study and the decision to publish.

Author contributions

B.Z.—conceptualization, data curation, analysis, investigation, methodology, supervision, validation, writing, review and editing. A.L.S.—conceptualization, data curation, analysis, investigation, methodology, supervision, validation, writing, review and editing. S.B.—conceptualization, data curation, formal analysis, investigation, methodology, supervision, validation, writing, review and editing. D.A.—analysis, methodology, validation, writing, review and editing. S.D.—conceptualization, data curation, investigation, methodology, review and editing. N.N.—conceptualization, data curation, investigation, supervision, review and editing. D.R.—data curation, investigation, reviewing and editing. J.S.—data curation, investigation, review and editing. A.J.B.—administration, resources, reviewing and editing. W.N.E.—conceptualization, methodology, administration, resources, review and editing. B.S.B.—conceptualization, investigation, methodology, administration, review and editing. V.A.R.—conceptualization, analysis, investigation, methodology, administration, resources, review and editing.

Competing interests

BSB has received consulting fees from Bristol Myers Squibb, Pfizer, and Takeda, unrelated to current research. VAR has received support from Janssen Inc. for research unrelated this work. AJB is a co-founder and consultant to Personalis and NuMedii; consultant to Samsung, Mango Tree Corporation, and in the recent past, 10× Genomics, Helix, Pathway Genomics, and Verinata (Illumina); has served on paid advisory panels or boards for Geisinger Health, Regenstrief Institute, Gerson Lehman Group, AlphaSights, Covance, Novartis, Genentech, and Merck, and Roche; is a shareholder in Personalis and NuMedii; is a minor shareholder in Apple, Facebook, Alphabet (Google), Microsoft, Amazon, Snap, 10× Genomics, Illumina, CVS, Nuna Health, Assay Depot, Vet-24seven, Regeneron, Sanofi, Royalty Pharma, AstraZeneca, Moderna, Biogen, Paraxel, and Sutro, and several other non-health related companies and mutual funds; and has received honoraria and travel reimbursement for invited talks from Johnson and Johnson, Roche, Genentech, Pfizer, Merck, Lilly, Takeda, Varian, Mars, Siemens, Optum, Abbott, Celgene, AstraZeneca, AbbVie, Westat, and many academic institutions, medical or disease specific foundations and associations, and health systems. AJB receives royalty payments through Stanford University, for several patents and other disclosures licensed to NuMedii and Personalis. AJB's research has been funded by NIH, Northrup Grumman (as the prime on an NIH contract), Genentech, Johnson and Johnson, FDA, Robert Wood Johnson Foundation, Leon Lowenstein Foundation, Intervallen Foundation, Priscilla Chan and Mark Zuckerberg, the Barbara and Gerson Bakar Foundation, and in the recent past, the March of Dimes, Juvenile Diabetes Research Foundation, California Governor's Office of Planning and Research, California Institute for Regenerative Medicine, L'Oreal, and Progenity. All other authors have nothing to disclose relevant to the manuscript. The disclosures had no involvement in any aspects of the study and the decision to publish.

Additional information

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1038/s41598-021-00367-7>.

Correspondence and requests for materials should be addressed to W.N.E.-N., B.S.B. or V.A.R.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2021