

## Title

Association of Chronic Acid Suppression and Social Determinants of Health with COVID-19 Infection

## Authors

Bing Zhang, MD MAS<sup>1,8</sup>; Anna L. Silverman, MD<sup>2,8</sup>; Saroja Bangaru, MD<sup>3,8</sup>; Douglas Arneson, PhD<sup>4</sup>; Sonya Dasharathy, MD<sup>3</sup>; Nghia Nguyen, MD MAS<sup>5</sup>; Diane Rodden, MD<sup>2</sup>; Jonathan Shih, BS<sup>6</sup>; Atul J. Butte, MD PhD<sup>4,7</sup>; Wael Noor El-Nachef, MD PhD<sup>3\*</sup>; Brigid S. Boland, MD<sup>5\*</sup>; Vivek A. Rudrapatna, MD PhD<sup>1,4\*</sup>

<sup>1</sup>Division of Gastroenterology and Hepatology, Department of Medicine, University of California, San Francisco, San Francisco, CA 94143

<sup>2</sup>Department of Medicine, University of California, San Diego, La Jolla, CA 92093

<sup>3</sup>Vatche and Tamar Manoukian Division of Digestive Diseases, Department of Medicine, University of California, Los Angeles, Los Angeles, CA 90095

<sup>4</sup>Bakar Computational Health Sciences Institute, University of California, San Francisco, San Francisco, CA 94143

<sup>5</sup>Division of Gastroenterology and Hepatology, Department of Medicine, University of California, San Diego, La Jolla, CA 92093

<sup>6</sup>Department of Neurology, University of California, San Francisco, San Francisco, CA 94143

<sup>7</sup>Center for Data-Driven Insights and Innovation, University of California Health, Oakland, CA

<sup>8</sup>These authors contributed equally

\*Correspondence

## Abstract

Acid suppressants are a widely-used class of medications previously linked to an increased risk of aerodigestive infections. However, prior studies of these medications as potentially reversible risk factors for COVID-19 have been conflicting. We performed a case-control study involving clinician-abstracted data from 900 health records across 3 US medical centers. We incorporated sociobehavioral predictors of infectious exposure using geomapping to publicly-available data. We found no evidence for an association between chronic acid suppression and incident COVID-19 (adjusted odds ratio 1.04, 95% CI: 0.92-1.17,  $P=0.515$ ). However, we identified several medical and

social features as positive (Latinx ethnicity, BMI  $\geq$  30, dementia, public transportation use, month of the pandemic) and negative (female sex, concurrent solid tumor, alcohol use disorder) predictors of new-onset infection. These results place both medical and social factors on the same scale within the context of the COVID-19 pandemic, and underscore the importance of comprehensive models of disease.

## **Introduction**

The COVID-19 pandemic, caused by the novel severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), has affected over 70 million individuals across the globe and killed over 1.6 million as of December 2020, one year after it was first reported in Wuhan, China.<sup>1</sup> Previous studies have examined risk factors for infection and predictors of disease severity. Established medical risk factors predicting infection and severity encompass age  $\geq$  65, diabetes mellitus, obesity, smoking, chronic pulmonary disease, cardiovascular disease, chronic kidney disease, malignancy and chronic HIV infection.<sup>2-7</sup>

Proton pump inhibitors (PPIs) and H<sub>2</sub>-receptor antagonists (H<sub>2</sub>RAs) are among the most commonly prescribed medications, used for gastroesophageal reflux disease (GERD) and peptic ulcer disease.<sup>8,9</sup> 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.<sup>9,10</sup> 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 compared to conventional medical factors. 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.

## **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). The decision made by UCSF's IRB was relied upon by the IRB of the University of California, San Diego (UCSD) via a

memorandum of understanding (UC IRB Reliance #3499). The decision made by the IRB of the University of California, Los Angeles (UCLA) was that this study did not constitute human subjects research and therefore was exempt.

At the time of the study's inception in May 2020, deidentified 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.<sup>11,12</sup> 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 patients 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 (**Supplemental Methods**). Variable definitions and data extraction process was determined and fixed *a priori* and based on consensus among the co-authors.

Medical records of all patients  $\geq 18$  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 an 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.

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 (**Supplemental 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 informatic 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.<sup>13-15</sup>

Geocoded covariates related to population and housing density were obtained from the 2010 United States Decennial Census.<sup>16</sup> Additional geocoded covariates were obtained from the 2018 United States American Community Survey (ACS) using the R package *tidycensus* v0.9.9.2.<sup>17</sup> 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 factors.<sup>18</sup> 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 Times.<sup>19</sup> 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.<sup>20,21</sup> 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 2020.<sup>22</sup> 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 outcomes.<sup>23</sup> The SDI used in this study was developed by the Robert Graham Center and is comprised of seven demographic characteristics from the 2015 ACS.<sup>24</sup> 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 (**Supplemental Table 1**).

### *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 convergence.<sup>25</sup>

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 10<sup>th</sup> and 90<sup>th</sup>

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.

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 intercept.<sup>26</sup> 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.

#### *Funding Source*

The UCSF Bakar Computational Health Science Institute, the National Institutes of Health, and the Barbara and Joel Marcus UCLA GI Fellowship Research Seed Grant Program had no involvement in the design, collection, analysis, interpretation and decision to publish the current study.

#### **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 (**Figure 1**). Five individuals with high degrees of missing data ( $>40\%$ ) and four with unknown residential zip codes were excluded due to inability to map geocoded social determinants of health. 861 medical records were retained for downstream analyses.

Distribution of included medical records were similar across the three sites, from which there were 428 (49.7%) documented positive tests (**Table 1**). Approximately half were female, and one-fifth were  $\geq 65$  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 CCI (0-2) and were not obese based on body mass index (BMI). Among previously

coexisting diseases captured, diabetes (13.7%) and immunocompromised (22.0%) were the most prevalent. Nearly one-fifth tested were healthcare workers, and the majority of people lived in single family homes. During the period sampled, the probability of a positive SARS-CoV2 test decreased significantly after March, likely reflecting the adoption of widespread testing (**Figure 2**).

#### *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 (**Figure 2**). This finding was confirmed with manual incorporation of 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,  $P = 0.515$ ; **Supplemental Figure 1**). 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  $\geq 30$ , and dementia (**Figure 2**). Unexpectedly, non-hematogenous malignancy and history of alcoholic use disorder 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*

Males and individuals of Latinx ethnicity were more likely to have positive test. 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. Among our sample, healthcare worker status did not strongly correlate with testing outcome, and neither did “social deprivation index” and “time spent at home”.

## **Discussion**

### *Chronic Acid Suppression and SARS-CoV-2 Risk*

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-19.<sup>27</sup> A cohort study using the United Kingdom's Biobank also reported no increased risk of COVID-19 among patients who reported taking PPI or H2RA.<sup>28</sup> Of note, medication data collection period was 2006-2010 for this study, therefore it was not confirmed 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 H2RA.<sup>29</sup> 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 proficiency.<sup>30</sup>

#### *Medical Comorbidities and SARS-CoV-2 Risk*

Consistent with previous data, our cohort found an increased risk of COVID-19 and a BMI  $\geq 30$ .<sup>31-35</sup> 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, 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 tested.<sup>36</sup>

#### *Sociodemographic Factors and SARS-CoV-2 Risk*

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.<sup>37-44</sup> 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.<sup>44,45</sup> 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.<sup>46,47</sup> 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.<sup>44,47,48</sup> 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 unknown if there are immunologic or behavioral explanations why female gender may be protective.<sup>49</sup> Future work is needed to clarify this finding.

### *Strengths and Limitations*

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 zipcode 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 include incorporating multiple predictors of the probability of testing as well as the probability of infection itself; nonetheless we cannot 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.

### *Conclusion*

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.

## References

1. Coronavirus Resource Center. COVID-19 Dashboard by the Center for Systems Science and Engineering (CCSE) at Johns Hopkins University (JHU). Johns Hopkins University. <https://coronavirus.jhu.edu/map.html>. Accessed November 10, 2020.
2. Blanco JL, Ambrosioni J, Garcia F, et al. COVID-19 in patients with HIV: clinical case series. *Lancet HIV*. 2020;7(5):e314-e316.
3. Dai M, Liu D, Liu M, et al. Patients with Cancer Appear More Vulnerable to SARS-CoV-2: A Multicenter Study during the COVID-19 Outbreak. *Cancer Discov*. 2020;10(6):783-791.
4. Grasselli G, Zangrillo A, Zanella A, et al. Baseline Characteristics and Outcomes of 1591 Patients Infected With SARS-CoV-2 Admitted to ICUs of the Lombardy Region, Italy. *JAMA*. 2020;323(16):1574-1581.
5. Guan WJ, Liang WH, Zhao Y, et al. Comorbidity and its impact on 1590 patients with COVID-19 in China: a nationwide analysis. *Eur Respir J*. 2020;55(5).
6. Lighter J, Phillips M, Hochman S, et al. Obesity in Patients Younger Than 60 Years Is a Risk Factor for COVID-19 Hospital Admission. *Clin Infect Dis*. 2020;71(15):896-897.
7. Park SE. Epidemiology, virology, and clinical features of severe acute respiratory syndrome -coronavirus-2 (SARS-CoV-2; Coronavirus Disease-19). *Clin Exp Pediatr*. 2020;63(4):119-124.
8. Kantor ED, Rehm CD, Haas JS, Chan AT, Giovannucci EL. Trends in Prescription Drug Use Among Adults in the United States From 1999-2012. *JAMA*. 2015;314(17):1818-1831.
9. Vaezi MF, Yang YX, Howden CW. Complications of Proton Pump Inhibitor Therapy. *Gastroenterology*. 2017;153(1):35-48.
10. Laheij RJ, Sturkenboom MC, Hassing RJ, Dieleman J, Stricker BH, Jansen JB. Risk of community-acquired pneumonia and use of gastric acid-suppressive drugs. *JAMA*. 2004;292(16):1955-1960.
11. Ding S, Liang TJ. Is SARS-CoV-2 Also an Enteric Pathogen With Potential Fecal-Oral Transmission? A COVID-19 Virological and Clinical Review. *Gastroenterology*. 2020;159(1):53-61.
12. Vilcu AM, Sabatte L, Blanchon T, 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*. 2019;2(11):e1916205.

13. 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%20a real%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%20a real%20representations%20of,station%20associated%20with%20mailing%20addresses). Accessed October 15, 2020.
14. UDS Mapper. ZIP Code to ZCTA Crosswalk. <https://udsmapper.org/zip-code-to-zcta-crosswalk/>. Accessed October 15, 2020.
15. Williams-Holt J. Spatial Crosswalks. <https://github.com/jwilliamsholt/safegraph-in-r/blob/master/spatial-crosswalks.Rmd>. Accessed October 10, 2020.
16. 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](https://www.census.gov/geographies/reference-files/time-series/geo/relationship-files.html#par_textimage_674173622). Accessed October 15, 2020.
17. 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.
18. Community Resilience Estimates. United States Census Bureau. <https://www.census.gov/data/experimental-data-products/community-resilience-estimates.html>. Accessed October 15, 2020.
19. The New York Times. Mask-Wearing Survey Data. <https://github.com/nytimes/covid-19-data/tree/master/mask-use>. Accessed October 20, 2020.
20. SafeGraph. COVID-19 Data Consortium. <https://www.safegraph.com/covid-19-data-consortium>. Accessed October 20, 2020.
21. SafeGraph. Social Distancing Metrics. <https://docs.safegraph.com/docs/social-distancing-metrics>. Accessed, October 20, 2020.
22. Google LLC. COVID-19 Community Mobility Reports. <https://www.google.com/covid19/mobility/>. Accessed October 25, 2020.
23. Butler DC, Petterson S, Phillips RL, Bazemore AW. Measures of social deprivation that predict health care access and need within a rational area of primary care service delivery. *Health Serv Res.* 2013;48(2 Pt 1):539-559.

24. 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.
25. Stef van Buuren KG-O. mice: Multivariate Imputation by Chained Equations in R. *J Stat Softw*. 2011;45(3).
26. Bates D, Machler M, Bolker BM, Walker SC. Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software*. 2015;67(1):1-48.
27. Lee SW, Ha EK, Yeniova AO, et al. Severe clinical outcomes of COVID-19 associated with proton pump inhibitors: a nationwide cohort study with propensity score matching. *Gut*. 2021;70(1):76-84.
28. Fan X, Liu Z, Miyata T, et al. Effect of Acid Suppressants on the Risk of COVID-19: A Propensity Score-Matched Study Using UK Biobank. *Gastroenterology*. 2020;169(1):455-458.e5.
29. Almario CV, Chey WD, Spiegel BMR. Increased Risk of COVID-19 Among Users of Proton Pump Inhibitors. *Am J Gastroenterol*. 2020;115(10):1707-1715.
30. Ryan C. Language Use in the United States: 2011. <https://www.census.gov/library/publications/2013/acs/acs-22.html>. 2010. Updated November 9, 2018. Accessed November 5, 2020.
31. Colaneri M, Novelli V, Cutti S, 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)*. 2020;doi:10.1093/pubmed/fdaa195.
32. Omrani AS, Almaslamani MA, Daghfal J, et al. The first consecutive 5000 patients with Coronavirus Disease 2019 from Qatar; a nation-wide cohort study. *BMC Infect Dis*. 2020;20(1):777.
33. Ranjan P, Kumar A, Chowdhury S, et al. Is excess weight a risk factor for the development of COVID 19 infection? A preliminary report from India. *Diabetes Metab Syndr*. 2020;14(6):1805-1807.
34. Razieh C, Zaccardi F, Davies MJ, Khunti K, Yates T. Body mass index and the risk of COVID-19 across ethnic groups: Analysis of UK Biobank. *Diabetes Obes Metab*. 2020;22(10):1953-1954.
35. Sattar N, Ho FK, Gill JM, 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*. 2020;14(5):1149-1151.
36. Fass R, Tougas G. Functional heartburn: the stimulus, the pain, and the brain. *Gut*. 2002;51(6):885-892.

37. Bailey ZD, Moon JR. Racism and the Political Economy of COVID-19: Will We Continue to Resurrect the Past? *J Health Polit Policy Law*. 2020;45(6):937-950.
38. Jon Zelner RT, Ramya Naraharisetti, Alex Cao, et al. Racial disparities in COVID-19 mortality are driven by unequal infection risks. *Clin Infect Dis*. 2020;doi:10.1093/cid/ciaa1723.
39. Mahajan UV, Larkins-Pettigrew M. Racial demographics and COVID-19 confirmed cases and deaths: a correlational analysis of 2886 US counties. *J Public Health (Oxf)*. 2020;42(3):445-447.
40. 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*. 2020;101:138-148.
41. Nepomuceno MR, Acosta E, Alburez-Gutierrez D, Aburto JM, Gagnon A, Turra CM. Besides population age structure, health and other demographic factors can contribute to understanding the COVID-19 burden. *Proc Natl Acad Sci U S A*. 2020;117(25):13881-13883.
42. Parcha V, Malla G, Suri SS, et al. Geographic Variation of Racial Disparities in Health and COVID-19 Mortality. *Mayo Clin Proc Innov Qual Outcomes*. 2020;4(6):703-716.
43. Price-Haywood EG, Burton J, Fort D, Seoane L. Hospitalization and Mortality among Black Patients and White Patients with Covid-19. *N Engl J Med*. 2020;382(26):2534-2543.
44. Yancy CW. COVID-19 and African Americans. *JAMA*. 2020;323(19):1891-1892.
45. Workbook: NYS-COVID19-Tracker. <https://covid19tracker.health.ny.gov>. 2020. Accessed November 1, 2020.
46. The COVID Racial Data Tracker. <https://covidtracking.com/race>. Accessed November 8, 2020.
47. McClure ES, Vasudevan P, Bailey Z, Patel S, Robinson WR. Racial Capitalism Within Public Health-How Occupational Settings Drive COVID-19 Disparities. *Am J Epidemiol*. 2020;189(11):1244-1253.
48. Page KR, Flores-Miller A. Lessons We've Learned - Covid-19 and the Undocumented Latinx Community. *N Engl J Med*. 2020;doi:10.1056/NEJMp2024897.
49. Kelada M, Anto A, Dave K, Saleh SN. The Role of Sex in the Risk of Mortality From COVID-19 Amongst Adult Patients: A Systematic Review. *Cureus*. 2020;12(8):e10114.

## **Acknowledgments**

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.

## **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/darneson/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](https://figshare.com/articles/dataset/Datasets_supporting_analytical_workflow_of_Chronic_Acid_Suppression_and_Social_Determinants_of_COVID-19_Infection/13380356).

## **Competing Interest Statement**

*The authors declare no 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, 10x 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, 10x Genomics, Illumina, CVS, Nuna Health, Assay Depot, Vet24seven, 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.

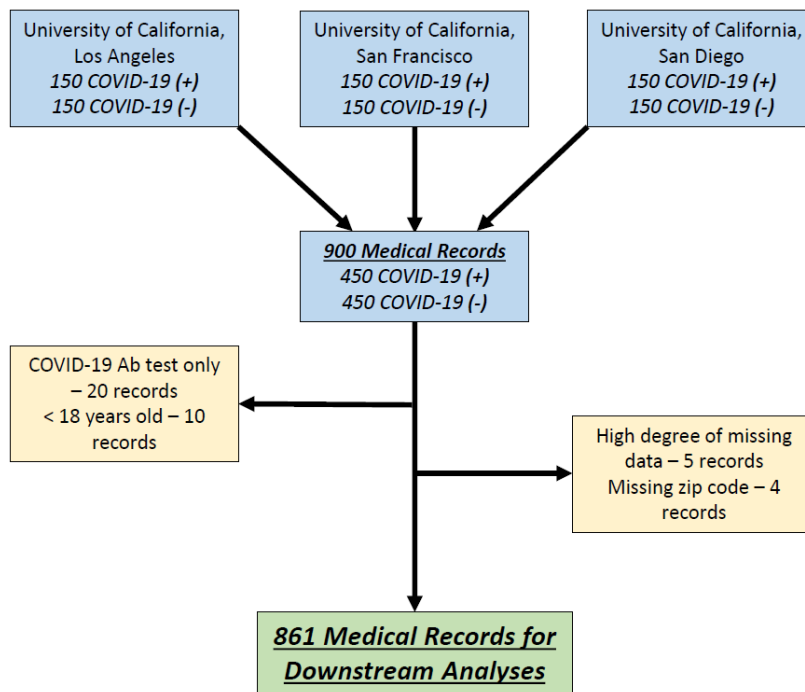
AJB's research has been funded by NIH, Northrup Grumman (as the prime on an NIH contract), Genentech, Johnson and Johnson, FDA, Robert Wood Johnson Foundation, Leon Lowenstein Foundation, Intervalien Foundation, Priscilla Chan and Mark Zuckerberg, the Barbara and Gerson Bakar Foundation, and in the recent 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 the design, collection, analysis, interpretation and decision to publish.

### **Funding Source**

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.



**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.



**Table 1** – Medical and sociodemographic characteristics of medical records included in analysis

Characteristic	UCLA (N=269)	UCSD (N=297)	UCSF (N=295)	Overall (N=861)
<b>Positive SARS-CoV2 test</b>	136 (50.6%)	147 (49.5%)	145 (49.2%)	428 (49.7%)
<b>SARS-CoV2 test month</b>				
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%)
<b>Female sex</b>	150 (55.8%)	153 (51.5%)	161 (54.6%)	464 (53.9%)
<b>Age - yr</b>				
>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%)
<b>Race or ethnic group</b>				
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%)
<b>Ethnicity Latinx</b>	56 (20.8%)	95 (32.0%)	38 (12.9%)	189 (22.0%)
<b>BMI</b>				
<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%)
<b>PPI use</b>	58 (21.6%)	40 (13.5%)	33 (11.2%)	131 (15.2%)
<b>Acid Suppression use</b>	66 (24.5%)	55 (18.5%)	38 (12.9%)	159 (18.5%)
<b>Charlson Comorbidity Index</b>				
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%)
<b>GERD</b>	66 (24.5%)	43 (14.5%)	45 (15.3%)	154 (17.9%)
<b>Previous coexisting disease</b>				
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%)
<b>Chronic opiate use</b>	13 (4.8%)	30 (10.1%)	9 (3.1%)	52 (6.0%)
<b>EtOH use disorder</b>	8 (3.0%)	13 (4.4%)	11 (3.7%)	32 (3.7%)
<b>Healthcare worker</b>	31 (11.5%)	64 (21.5%)	65 (22.0%)	160 (18.6%)
<b>Residence</b>				
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%)
<b>Social Deprivation Index</b>				
>65	128 (47.6%)	130 (43.8%)	94 (31.9%)	352 (40.9%)
<b>Time spent home</b>				
>85.0%	250 (92.9%)	266 (89.6%)	253 (85.8%)	769 (89.3%)
<b>Public transportation usage</b>				
>4.4%	132 (49.1%)	106 (35.7%)	256 (86.8%)	494 (57.4%)