BMJ Open

Racial and Ethnic Disparities in SARS-CoV-2 Pandemic: Analysis of a COVID-19 Observational Registry for a Diverse U.S. Metropolitan Population

Journal:	*BMJ Open*
Manuscript ID	bmjopen-2020-039849
Article Type:	Original research
Date Submitted by the Author:	28-Apr-2020
Complete List of Authors:	Vahidy, Farhaan S; Houston Methodist Research Institute, Center for Outcomes Research; Houston Methodist Neurological Institute Nicolas, Juan Carlos; Houston Methodist Research Institute Meeks, Jennifer R; Houston Methodist Research Institute Khan, Osman; Houston Methodist Research Institute Jones, Stephen L.; Houston Methodist Research Institute; Weill Cornell Medicine Masud, Faisal; Houston Methodist Hospital; Weill Cornell Medicine Sostman, H Dirk; Houston Methodist Research Institute; Weill Cornell Medicine Phillips, Robert; Houston Methodist Hospital; Weill Cornell Medicine Andrieni, Julia D; Houston Methodist Hospital; Weill Cornell Medicine Kash, Bita A; Houston Methodist Research Institute; Texas A&M University School of Rural Public Health Nasir, Khurram; Houston Methodist Research Institute; Houston Methodist Hospital
Keywords:	EPIDEMIOLOGY, Public health < INFECTIOUS DISEASES, INFECTIOUS DISEASES

I, the Submitting Author has the right to grant and does grant on behalf of all authors of the Work (as defined in the below author licence), an exclusive licence and/or a non-exclusive licence for contributions from authors who are: i) UK Crown employees; ii) where BMJ has agreed a CC-BY licence shall apply, and/or iii) in accordance with the terms applicable for US Federal Government officers or employees acting as part of their official duties; on a worldwide, perpetual, irrevocable, royalty-free basis to BMJ Publishing Group Ltd ("BMJ") its licensees and where the relevant Journal is co-owned by BMJ to the co-owners of the Journal, to publish the Work in this journal and any other BMJ products and to exploit all rights, as set out in our licence.

The Submitting Author accepts and understands that any supply made under these terms is made by BMJ to the Submitting Author unless you are acting as an employee on behalf of your employer or a postgraduate student of an affiliated institution which is paying any applicable article publishing charge ("APC") for Open Access articles. Where the Submitting Author wishes to make the Work available on an Open Access basis (and intends to pay the relevant APC), the terms of reuse of such Open Access shall be governed by a Creative Commons licence – details of these licences and which Creative Commons licence will apply to this Work are set out in our licence referred to above.

Other than as permitted in any relevant BMJ Author's Self Archiving Policies, I confirm this Work has not been accepted for publication elsewhere, is not being considered for publication elsewhere and does not duplicate material already published. I confirm all authors consent to publication of this Work and authorise the granting of this licence.

Racial and Ethnic Disparities in SARS-CoV-2 Pandemic: Analysis of a COVID-19 Observational Registry for a Diverse U.S. Metropolitan Population

Farhaan S. Vahidy PhD MBBS MPH,[1,2] Juan Carlos Nicolas BS,[1] Jennifer R. Meeks MS,[1] Osman Khan BS,[1] Stephen L. Jones MD MSHI,[1,3,4] Faisal N. Masud MD,[4,5] H. Dirk Sostman MD,[4,6] Robert A. Phillips MD PhD,[1,4,7] Julia D. Andrieni MD,[4,8] Bita Kash PhD MBA,[1,4,9] Khurram Nasir MD MPH[1,7]

1. Center for Outcomes Research, The Houston Methodist Research Institute, Houston TX
2. Houston Methodist Neurological Institute, Houston Methodist Hospital, Houston TX
3. Department of Surgery, Houston Methodist Hospital, Houston TX
4. Weill Cornell Medical College, New York NY
5. Department of Anesthesiology and Critical Care, Houston Methodist Hospital, Houston TX
6. The Houston Methodist Research Institute and The Houston Methodist Academic Institute, Houston TX
7. Department of Cardiology, DeBakey Heart and Vascular Center, Houston Methodist Hospital, Houston TX
8. Department of Medicine, Houston Methodist Hospital, Houston TX
9. Department of Health Policy and Management, Texas A&M University School of Public Health, College Station TX

Corresponding Author

Farhaan S. Vahidy, PhD MBBS MPH
Josie Roberts Building
Houston Methodist
7550 Greenbriar Drive
Houston TX 77030
fvahidy@houstonmethodist.org

Word Count: 2,735

Key Points

What is the key question: Do race and ethnic disparities exist in susceptibility to SARS-CoV-2 infection and how can these disparities be explained?

What is the bottom line: African American race and Hispanic ethnicity are associated with higher likelihood of SARS-CoV-2 infection, potentially mediated by residence in high population density areas.

Why read on: We provide novel estimates of higher likelihood of race and ethnic disparities in susceptibility to the SARS-CoV-2 infection from a large heterogenous metropolitan in the U.S.

Abstract

Introduction: Data on race and ethnic susceptibility to SARS-CoV-2 infection are limited. We analyzed socio-demographic factors associated with higher likelihood of SARS-CoV-2 infection and explore mediating pathways for race disparities in the SARS-CoV-2 pandemic.

Methods: Cross sectional analysis of COVID-19 Surveillance and Outcomes Registry (CURATOR), which captures data for a large healthcare system comprising of one central tertiary care, seven large community hospitals, and an expansive ambulatory / emergency care network in the Greater Houston area. Nasopharyngeal samples for individuals inclusive of all ages, races, ethnicities and sex were tested for SARS-CoV-2. We analyzed, socio-demographic (age, sex, race, ethnicity, household income, residence population density) and comorbidity (hypertension, diabetes, obesity, cardiac disease) factors. Multivariable logistic regression models were fitted to provide adjusted Odds Ratios (aOR), 95% confidence intervals (CI) for likelihood of positive SARS-CoV-2 test. Structural Equation Modeling (SEM) framework was utilized to explore three mediation pathways (low income, high population density, high comorbidity burden) for association between African American race and SARS-CoV-2 infection.

Results: Among 4,513 tested individuals, 754 (16.7%) tested positive. Overall mean (SD) age was 50.6 (18.9) years, 62% females and 26% were African American. African American race was associated with lower socio-economic status, higher comorbidity burden, and population density residence. In the fully adjusted model, African American race (vs. White; aOR, CI: 1.84, 1.49–2.27) and Hispanic ethnicity (vs. non-Hispanic; aOR, CI: 1.70, 1.35–2.14) had a higher likelihood of infection. Older individuals and males were also at a higher risk of SARS-CoV-2 infection. The SEM framework demonstrated a statistically significant (p = 0.008) indirect effect

of African American race on SARS-CoV-2 infection mediated via a pathway that included residence in densely populated zip code.

Conclusions: There is strong evidence of race and ethnic disparities in the SARS-CoV-2 pandemic potentially mediated through unique social determinants of health.

Strengths and limitations of this study

- One of the first studies to systematically evaluate race and ethnic disparities in susceptibility to SARS-CoV-2 infection, while accounting for multiple socio-demographic characteristics and comorbidities
- Study population represents a large and diverse metropolitan of the U.S. with data from one of the largest healthcare providers across the greater metropolitan area
- Study evaluates potential mediation pathways for race disparities and demonstrates that residence in areas with high population density may mediate race disparities in susceptibility to SARS-CoV-2 infection
- Single center study with limited information about true burden of comorbidity and lifestyle factors

Introduction

The Coronavirus (COVID-19) disease caused by infection with the Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2) virus is a pandemic that has thus far resulted in over 2 million cases across 170 countries in under 4 months. At the time of this reporting, the U.S. has approximately 30% of total global cases, and has surpassed all countries in terms of absolute number of cases and fatalities.[1,2] Experts project these numbers to continue rising as widespread testing is instituted. The geographic distribution of cases across the U.S. suggests that the major pandemic burden has hit metropolitan areas such as New York; however, cases of COVID-19 have now been reported across all 50 states, the District of Columbia, Guam, Puerto Rico, the Northern Mariana Islands, and the U.S. Virgin Islands.[3] As of April 18, the state of Texas had 18,260 reported cases of COVID-19, with approximately one-third in the Greater Houston area.[4] The greater Houston area is home to approximately 7 million individuals, is the fourth-largest metropolitan area by population in the U.S. and is considered one of the nation's most diverse regions.

Initial reports from China and Europe indicate that specific individuals such as the elderly; males; and people with comorbidities including hypertension, diabetes, obesity, coronary artery disease and heart failure have poor COVID-19 outcomes.[5–8] As the pandemic spread over the continental U.S. during the last two months, patterns of high-risk phenotypes have started to emerge, and reports of poor outcomes (particularly high case fatality) among racial minorities have surfaced in the media.[9–11] Though it is important to understand the determinants of poor outcomes among COVID-19 patients, it is equally imperative, from a public health perspective, to systematically examine the likelihood of SARS-CoV-2 infection across large diverse communities in the U.S. More specifically, data on potential higher likelihood of SARS-CoV-2

infection among racial and ethnic minorities across diverse U.S. metropolitan areas outside of New York are limited. Furthermore, the mediators of SARS-CoV-2 infection among racial and ethnic minorities have not been described.

We explored socio-demographic characteristics such as age, sex, race, ethnicity, median household zip code income, population density of residents' zip codes, and health insurance status associated with positive SARS-CoV-2 testing in an urban and diverse population served by one of the leading healthcare systems of the greater Houston area. We further examined the association between pre-existing comorbidities and higher likelihood of SARS-CoV-2 infection in our study population. We hypothesized that older age, non-white race and ethnic minority status will be associated with significantly higher likelihood of SARS-CoV-2 infection, and factors such as low socio-economic status, residence in high population density areas (proxy for potential difficulties in social distancing) and higher comorbidity burden will mediate the effect of race on SARS-CoV-2 infection.

Methods

We analyzed data being contemporaneously collected since March 5, 2020 as a part of the COVID-19 Surveillance and Outcomes Registry (CURATOR) at the Houston Methodist Hospital system (HM). The Houston Methodist CURATOR has been approved by the HM Institutional Review Board (IRB) as an observational quality of care registry for all suspected and confirmed COVID-19 patients. CURATOR is populated from multiple data sources across the HM system such as electronic medical records, electronic databanks for laboratory and pharmacy, and electronic interactive patient interface tools. The HM system comprises a flagship tertiary care hospital in the Texas Medical Center, seven large community hospitals, a continuing care hospital, and multiple emergency centers and clinics throughout the Greater Houston area.

Data from various sources are curated into a harmonized format, assessed for quality and integrity, and stored on a secure institutional HIPAA-compliant server.

We flagged all individuals who were tested for the SARS-CoV-2 using the real time Reverse Transcriptase (RT) Polymerized Chain Reaction (PCR) diagnostic panels. These assays were verified for quantitative detection of novel SARS-CoV-2 isolated and purified from nasopharyngeal swab specimens obtained from individuals and immersed in universal transport medium. Testing was carried out for symptomatic individuals or for individuals who had a self-reported history of exposure to a COVID-19 case including recent travel to other countries with high infection rates or hotspots within the U.S. Socio-demographic characteristics including age, sex, race, ethnicity, and payer-status (insurance type) were obtained from the HM CURATOR for analyses. We utilized the U.S. Census Bureau's American Community Survey (ACS) 5-year data (2014–2018) to determine median household income by individual zip code tabulation areas (ZCTA).[12] The median ZCTA household income was inflation-adjusted to 2018 USD. We also utilized the same data source to obtain population estimates by ZCTA, and calculated ZCTA level population density (population per mile square) by standardizing it for area measurements of ZCTA. For the purpose of population density determination, land area estimates were obtained from the Census Bureau's U.S. Gazetteer Files 2010.[13] In the absence of granular and precise social distancing data, we have utilized population density as a proxy for potential difficulties in social distancing among crowded communities.

We provide descriptive summary data as means (standard deviations) and proportions. We fit univariable and multivariable logistic regression models to assess unadjusted and adjusted association between socio-demographic characteristics and likelihood of being tested positive for the SARS-CoV-2. We determined *a priori* to include all variables (age, sex, race, ethnicity, zip

code household income, insurance type, zip population density and comorbidities) in our final multivariable model. We assessed the model fit utilizing the Hosmer and Lemeshow goodness of fit test and crude and adjusted odds ratios (OR and aOR), and 95% confidence intervals (CI) are reported. Age, income and population density variables were categorized to improve model fit. Post-estimation marginal probabilities of SARS-CoV-2 infection were determined from the fully adjusted model for major covariates (race, ethnicity and age). A comorbidity burden score was calculated by assigning one point each for presence of hypertension, diabetes, obesity or a combination of Coronary Artery Disease / Myocardial Infarction / Congestive Heart Failure (CAD / MI / CHF). We explored the mediation influence of comorbidity burden, socio-economic status (median income), and lack of social distancing (population density) on the relationship between African American race and high likelihood of SARS-CoV-2 infection using the Generalized Structural Equation Modeling (GSEM) framework. The GSEM framework was set up to provide estimates of direct and indirect effect of African American race on SARS-CoV-2 infectivity. Statistically significant ($p < 0.05$) indirect effects represent full or partial mediation by a tested covariate. We included all individuals tested for SARS-CoV-2 across our healthcare system and did not perform formal sample size calculations.

Patient and public involvement: There was no direct patient involvement in the design and conduct of this study.

Results

From the HM CURATOR, during an approximate 5-week (37-day) time period, we identified a total of 4,513 presumed cases tested for SARS-CoV-2, among whom 754 (16.7%, 95% CI: 15.6 – 17.8) tested positive. Figure 1 represents temporal course of total, positive, and negative SARS-CoV-2 tests across the 37-day timeline in our hospital system.

Socio-Demographic and Comorbidity Characteristics of the Study Population

Overall, the mean (SD) age of the study population was 50.6 (18.9) years; 62% were female and 58% were Caucasian. The overall median (IQR) household income was USD $70,324 ($53,116–$97,747), and 39.8% of the study population had private or employer-based insurance. In our univariate analysis, African American race (vs. White; OR, CI: 1.52, 1.28–1.82), Hispanic (vs. non-Hispanic; OR, CI: 1.26, 1.04–1.54), and males (vs. females; OR, CI: 1.30, 1.11–1.51), were associated with significantly higher likelihood of testing positive for SARS-CoV-2. Furthermore, among the SARS-CoV-2 positive patients, 44% were in the age category of 51–75 years, and 11% were greater than 75 years. These proportions were significantly higher than the reference group (up to 35 years; OR, CI for 51-75 years vs. up to 35 years: 1.76, 1.42–2.18 and for >75 years vs. up to 35 years: 1.35, 1.01–1.79). Furthermore, individuals in higher pentiles of socio-economic status had significantly lower likelihood; whereas, those residing in higher population density ZCTAs had higher likelihood of SARS-CoV-2 infection. Among comorbidities, a significantly greater proportion of diabetic individuals had SARS-CoV-2 positive results (OR, CI: 1.40, 0.17 – 1.68). The socio-demographic characteristics and comorbidity profiles for the overall and SARS-CoV-2 positive and negative patients are summarized in Table 1.

Socio-demographic and comorbidity characteristics associated with African American Race

In order to understand the association between African American race and other socio-demographic factors, we compared age, sex, median income, population density, and comorbidity profile between African American and non-African American race. Although African Americans had higher proportion of younger individuals and greater proportion of females. A significantly higher proportion of African Americans had lower socio-economic

status, resided in ZCTAs with higher population density, and had high comorbidity burden for hypertension, diabetes, obesity and CAD / MI / CHF. Table 2 provides univariable comparison of African Americans vs. Non-African Americans across various socio-demographic and comorbidity characteristics.

Multivariable Model and Marginal Probabilities for likelihood of SARS-CoV-2 infection

The significantly higher likelihood of SARS-CoV-2 infection among African Americans (compared to White) persisted after controlling for other demographics, insurance type, median household income, population density, and comorbidities. Adjusted odds ratios (CI) for African American vs. White: 1.84 (1.49–2.27). Our fully adjusted model estimated that Asians (vs. White) were also at a significantly higher risk of SARS-CoV-2 infection (aOR, CI: 1.46, 1.09–1.95). Furthermore, we also observed a statistically significant association between SARS-CoV-2 infection and Hispanic ethnicity, aOR (CI): 1.70 (1.35–2.14). Higher risk of infection among males (compared to females) and higher likelihood of SARS-CoV-2 infection among elderly also remained statistically significant. Detailed output of the fully adjusted logistic regression model is presented in Table 3. The influence of African American race (vs. White) and Hispanic (vs. Non-Hispanic) ethnicity was observed uniformly across the age spectrum of 10 – 80 years. In other words, we did not observe effect modification by age for relationship between race / ethnicity and SARS-CoV-2 infection. However older age in itself remains significantly associated with higher likelihood of SARS-CoV-2 infection. Based on the marginal probabilities obtained from our fully adjusted model, the probability of SARS-CoV-2 infection in a 40-year-old African American is 18.8% whereas it is 12.9% in a 40-year-old White individual, all other adjusted variables being constant. At the age of 75 this probability is 29.0% for an African American, and 20.8% for a Caucasian. A similar relationship differential was observed for

Hispanic vs. non-Hispanic. Probability of SARS-CoV-2 infection for African American vs. White, and for Hispanic vs. Non-Hispanic across age spectrum is presented in Figure 2 and Figure 3.

Generalized Structural Equation Modeling for Mediation by Income, Population Density and Comorbidity

Utilizing the GSEM framework, we determined the direct and indirect effects of African American race on SARS-CoV-2 infection with median income, population density and comorbidity score modeled as mediators in three separate equations. The indirect effect of African American race mediated through population density was statistically significant ($p = 0.008$); however, the indirect effects mediated via median income and comorbidity scores were not statistically significant ($p = 0.31$ and $p = 0.38$ respectively).

Discussion

There is emerging evidence of race disparities in the evolving COVID-19 pandemic across the continental U.S. Most reports indicate higher case fatality among African Americans across major U.S. metropolitan areas.[9–11] However, robust insights on the racial and ethnic differences for SARS-CoV-2 infection are limited. This is perhaps because of comparatively homogenous populations in non-U.S. regions of the world. Houston, as an exceptionally ethnically diverse population center,[14] is well suited for an investigation of racial, ethnic, and socioeconomic gradients in COVID-19 test positivity.

Our study adds to the current literature by analyzing emerging data for individuals being tested across one of the largest healthcare systems in the Greater Houston area; we report that racial minorities (non-Hispanic African American and Asian) are approximately 50–80% more likely to test positive for SARS-CoV-2 than the non-Hispanic White population. Our data also

indicate that the Hispanic population is almost 70% more likely than non-Hispanics to be susceptible to SARS-CoV-2 infection. These findings illuminate systematic racial / ethnic disparities in testing positive for SARS-CoV-2 infection. Though there are limited prior SARS-CoV-2 data, such race and ethnic disparities have previously been described for the U.S. H1N1 influenza pandemic.[15] These data indicated that Spanish-speaking Hispanics were at a greater risk of H1N1 infection primarily attributable to lack of healthcare access. Black people were also more susceptible to complications of H1N1 infection.[15]

We explored three possible mechanisms of race disparities in our data. These included lower socio-economic status, residence in higher population dense areas, and higher level of comorbidities. We demonstrate that African American race is significantly associated with all three potential disparity pathways, and in the traditional multivariable analyses, race and ethnic disparities persisted even after controlling for these pathways. However, our mediation analyses highlighted the potential influence of residence in high population density areas as a viable pathway that at least partially explains race disparity. Pathways mediating the influence of median income and comorbidity status did not demonstrate a significant effect. We utilized population density as a marker for potential inability to maintain adequate social distancing as it has been indicated that maintaining the WHO recommended safe distance between people becomes challenging with high population densities.[16] Furthermore, overall effects of population density and disease spread has been previously described in literature.[17,18] In addition to lack of social distancing, higher population density may also be associated with several other behavioral and socio-demographic attributes that may predispose to both viral spread and increased susceptibility. For example, there are reports linking obesity, lack of physical activity, and higher mortality with residence in densely populated neighborhoods.[19,20]

As reported, our data also corroborate that older populations may be more susceptible to SARS-CoV-2 infection;[8] however, younger populations still have cause for concern, as nearly 1 in 4 of the infected cases in our sample were between 36–50 years. Finally, our data demonstrate that males may be approximately 30% more likely to test positive for the SARS-CoV-2 infection. Potential sex differences in infectivity to SARS-CoV-2 and intersectionality with racial and ethnic socioeconomic factors need to be explored further in future analyses. Additional policy-oriented research should prioritize study on the intersectionality of these vulnerable economic statuses and racial disparities in COVID infection indicated by the present study.

Findings of our study need to be interpreted in the light of certain limitations. First, our data are from a single center and may not be generalizable to the wider U.S. population. These findings need to be replicated in larger data sets across other large heterogenous U.S. metropolitans. However, the Houston metropolitan area is one of the most diverse and representative in the U.S.,[14] and our healthcare system is one of the largest systems providing care to COVID-19 patients in the Greater Houston area. Our sample was composed of 26% Black, 19% Hispanic, and 62% female population. Second, we did not have information on certain demographic covariates such as education. Educational status has been linked to healthcare awareness and may be important to adjust for in analyses of potential disparities. However, we obtained and adjusted for zip code income data from the U.S. Census, as income has previously been shown to have strong correlation with educational attainment.[21] Third, since testing was based on suspicion of infection and may have been influenced by factors such as access to care, the potential for selection bias cannot be ruled out. Finally, we did not have detailed information on comorbidities and their management in the study population. However,

we did control for major comorbidities which are being reported as associated with COVID-19 outcomes.[22]

Conclusions

The strong association between racial and ethnic minorities and SARS-CoV-2 infection demonstrated in our data, even after adjustment for other important socio-demographic and comorbidity factors, highlight a potential catastrophe of inequality within the existential crisis of a global pandemic. Our data, representing a large heterogeneous U.S. metropolitan area, also provide preliminary evidence into the potential pathways for this disparity. It is highly likely that higher comorbidity burden and detrimental effects of adverse social determinants, including those that may not adequately permit safe practices of social distancing, mediate higher SARS-CoV-2 infectivity among racial and ethnic minorities.

As the pandemic continues to spread and evolve across the continental U.S., emerging data on association between SARS-CoV-2 infection and various socio-demographic factors will continue to enhance our understanding of targeted risks related to SARS-CoV-2 infection, and such data would enable us to comprehend healthcare services and access factors related to development and outcomes of COVID-19 among minority populations.

Acknowledgment: The authors thank Jacob M. Kolman, Senior Scientific Writer of the Houston Methodist Center for Outcomes Research, for reviewing the language and format of the manuscript.

IRB Approval: This work was carried out under an approved protocol for the Houston Methodist COVID-19 Surveillance and Outcomes Registry (HM CURATOR) by the Houston Methodist Research Institute Institutional Review Board (HMRI IRB).

Contribution Statement:

FV: design, data analysis and interpretation, drafting the manuscript, critical revision for important intellectual content, final approval

JCN: data acquisition, data analysis, drafting the manuscript, final approval

OK: data acquisition, data analysis, drafting the manuscript, final approval

SLJ: data acquisition, data interpretation, critical revision for important intellectual content, final approval

FNM: critical revision for important intellectual content, final approval

HDS: critical revision for important intellectual content, final approval

RAP: critical revision for important intellectual content, final approval

JDA: critical revision for important intellectual content, final approval

BK: critical revision for important intellectual content, final approval

KN: design, interpretation of data, critical revision for important intellectual content, final approval

Competing Interests: No authors declare a competing interest to this work

Funding: This work was supported internally by the Houston Methodist Academic Institute.

Date sharing statement: All requests for de-identified data should be made to the corresponding author. All reasonable requests will be evaluated by the CURATOR Data Governance and Sharing Committee comprising of FV, SLJ, BK and KN in the light of institutional policies and guidelines.

References

1. World Health Organization. WHO COVID-19 dashboard. https://covid19.who.int/. Published April 6, 2020. Accessed April 6, 2020.

2. Center for Systems Science and Engineering (CSSE) at Johns Hopkins University. COVID-19 map. Johns Hopkins Coronavirus Resource Center. https://coronavirus.jhu.edu/map.html. Published April 6, 2020. Accessed April 6, 2020.

3. Centers for Disease Control and Prevention. Coronavirus Disease 2019 (COVID-19) in the U.S. cdc.gov. https://www.cdc.gov/coronavirus/2019-ncov/cases-updates/cases-in-us.html. Published April 14, 2020. Accessed April 14, 2020.

4. Texas Health and Human Services. COVID-19 in Texas (dashboard). https://txdshs.maps.arcgis.com/apps/opsdashboard/index.html#/ed483ecd702b4298ab01e8b9cafc8b83. Published April 18, 2020. Accessed April 18, 2020.

5. Shi S, Qin M, Shen B, et al. Association of cardiac injury with mortality in hospitalized patients with COVID-19 in Wuhan, China. *JAMA Cardiol*. March 2020. doi:10.1001/jamacardio.2020.0950

6. Bonow RO, Fonarow GC, O'Gara PT, Yancy CW. Association of coronavirus disease 2019 (COVID-19) with myocardial injury and mortality. *JAMA Cardiol*. March 2020. doi:10.1001/jamacardio.2020.1105

7. 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*. April 2020. doi:10.1001/jama.2020.5394

8. Zhou F, Yu T, Du R, et al. Clinical course and risk factors for mortality of adult inpatients with COVID-19 in Wuhan, China: a retrospective cohort study. *Lancet*. 2020;395(10229):1054-1062. doi:10.1016/S0140-6736(20)30566-3

9. Deslatte M. Louisiana data: virus hits blacks, people with hypertension. *AP News*. https://apnews.com/cde897293f88f788d3f3bfdb3d46c466. Published April 7, 2020. Accessed April 20, 2020.

10. Reyes C, Husain N, Gutowski C, St. Clair S, Pratt G. Chicago's coronavirus disparity: Black Chicagoans are dying at nearly six times the rate of white residents, data show. *Chicago Tribune*. https://www.chicagotribune.com/coronavirus/ct-coronavirus-chicago-coronavirus-deaths-demographics-lightfoot-20200406-77nlylhiavgjzb2wa4ckivh7mu-story.html. Published April 7, 2020. Accessed April 20, 2020.

11. Thebault R, Tran AB, Williams V. The coronavirus is infecting and killing black Americans at an alarmingly high rate. *Washington Post*. https://www.washingtonpost.com/nation/2020/04/07/coronavirus-is-infecting-killing-black-americans-an-alarmingly-high-rate-post-analysis-shows/. Published April 7, 2020. Accessed April 20, 2020.

12. United States Census Bureau. American Community Survey (ACS). https://www.census.gov/programs-surveys/acs. Published 2018. Accessed April 14, 2020.

13. United States Census Bureau. Gazetteer Files. https://www.census.gov/geographies/reference-files/time-series/geo/gazetteer-files.html. Accessed April 20, 2020.

14. Harris County Public Health. *Harris Cares: A 2020 Vision of Health in Harris County*. Houston, TX: Harris County Public Health; 2019. http://publichealth.harriscountytx.gov/Resources/Harris-Cares. Accessed November 21, 2019.

15. Quinn SC, Kumar S, Freimuth VS, Musa D, Casteneda-Angarita N, Kidwell K. Racial disparities in exposure, susceptibility, and access to health care in the US H1N1 influenza pandemic. *Am J Public Health*. 2011;101(2):285-293. doi:10.2105/AJPH.2009.188029

16. Rocklov J, Sjodin H. High population densities catalyze the spread of COVID-19. J Travel Med. 2020. doi: 10.1093/jtm/taaa038

17. Tarwater PM, Martin CF. Effects of population density on the spread of disease. *Complexity*. 2001;6(6):29-36. doi:10.1002/cplx.10003

18. Yang K, LeJeune J, Alsdorf D, Lu B, Shum CK, Liang S. Global distribution of outbreaks of water-associated infectious diseases. *PLoS Negl Trop Dis*. 2012;6(2):e1483. doi:10.1371/journal.pntd.0001483

19. Wang R, Feng Z, Xue D, Liu Y, Wu R. Exploring the links between population density, lifestyle, and being overweight: secondary data analyses of middle-aged and older Chinese adults. *Health Qual Life Outcomes*. 2019;17(1):100. doi:10.1186/s12955-019-1172-3

20. Beenackers MA, Oude Groeniger J, Kamphuis CBM, Van Lenthe FJ. Urban population density and mortality in a compact Dutch city: 23-year follow-up of the Dutch GLOBE study. *Health & Place*. 2018;53:79-85. doi:10.1016/j.healthplace.2018.06.010

21. U.S. Bureau of Labor Statistics. Unemployment rates and earnings by educational attainment. https://www.bls.gov/emp/chart-unemployment-earnings-education.htm. Published September 4, 2019. Accessed April 15, 2020.

22. Centers for Disease Control and Prevention. Groups at higher risk for severe illness. cdc.gov. https://www.cdc.gov/coronavirus/2019-ncov/need-extra-precautions/groups-at-higher-risk.html. Last updated April 17, 2020. Accessed April 20, 2020.

Tables and Figures

Figure 1: Schematic representation of the temporal sequence for total, positive and negative numbers of SARS-CoV-2 tests in the Houston Methodist CURATOR

Figure 2: Adjusted Probability and 95% Confidence Interval of Positive SARS-CoV-2 PCR in African American vs. Caucasian by increasing age

Figure 3: Adjusted Probability and 95% Confidence Interval of Positive SARS-CoV-2 PCR in Hispanic vs. Non-Hispanic by increasing age

Table 1: Summary measures and univariable association of socio-demographic characteristics with SARS-CoV-2 infection from HM CURATOR

Characteristics	Overall (n = 4,513)	SARS-CoV-2 Negative (n = 3,759)	SARS-CoV-2 Positive (n = 754)	OR (95% CI)
Age, mean (SD)	50.6 (18.9)	50.1 (19.1)	53.2 (17.8)	1.01 (1.00–1.01)
Age Categories (%)				
Up to 35 years	24.8	26.0	19.0	Reference Category
36–50 years	27.4	27.8	25.3	1.25 (0.99–1.58)
51–75 years	35.9	34.3	44.0	1.76 (1.42–2.18)
>75 years	11.9	11.9	11.7	1.35 (1.01–1.79)
Females (%)	62.1	63.2	56.9	0.77 (0.66–0.90)
Race (%)				
White	57.7	58.8	51.7	Reference Category
African American	25.7	24.3	32.6	1.52 (1.28–1.82)
Asian	9.4	9.2	10.3	1.27 (0.97–1.67)
More Than One Race / Other / Unknown	7.3	7.7	5.2	0.76 (0.54–1.08)
Hispanic (%)*	18.7	18.1	21.8	1.26 (1.04–1.54)
Median Zip Household Income (IQR)†	70,324 (53,116–97,747)	70,658 (53,313–97,747)	66,983 (50,665–95,835)	-3675‡ (-6667.01, -682.95)
Median Zip Household Income Pentiles (%)				
I: 24,993–50,462	19.8	19.1	23.6	Reference Category
II: 50,465–65,339	18.5	18.4	19.4	0.85 (0.67–1.09)
III: 65,742–78,487	21.6	21.9	20.3	0.75 (0.59–0.95)
IV: 78,06–102,583	19.7	20.3	16.8	0.67 (0.52–0.86)
V: 103,48–230,750	20.4	20.4	19.9	0.79 (0.62–1.00)
Median (IQR) Population Density	2991.6 (1538. 9 – 4299.7)	2883.9 (1504 – 4261)	3320.3 (2123.6 – 4665.6)	436.4‡ (265.7 – 607.1)
Median Population Density Pentiles (%)				
I: 1.5 – 1370.7	19.9	20.8	15.8	Reference Category
II: 1393.7 – 2524.0	19.5	19.8	18.0	1.19 (0.91 – 1.56)
III: 2603.0 – 3486.5	20.4	20.1	22.0	1.44 (1.11 – 1.86)
IV: 3513.1 – 4789.7	19.8	19.7	20.4	1.36 (1.05 – 1.76)
V: 4801.1 – 61610.2	20.4	19.8	23.7	1.57 (1.22 – 2.03)
Insurance Status (%)				
Medicare	28.8	28.5	30.1	Reference Category
Medicaid	3.6	3.8	2.5	0.63 (0.38–1.04)
Pvt / Employer based	39.8	39.7	40.2	0.96 (0.79–1.16)
HC Exchange	2.4	2.2	3.6	1.57 (0.99–2.49)
Self-Pay	24.3	24.5	23.2	0.90 (0.72–1.11)
VA	1.2	1.3	0.4	0.28 (0.09–0.92)
Hypertension	42.3	41.9	44.4	1.11 (0.95 – 1.30)
Diabetes	21.7	20.7	26.8	1.40 (1.17 – 1.68)
Obesity	8.0	8.1	7.7	0.92 (0.68 – 1.23)

Characteristics	Overall (n = 4,513)	SARS-CoV-2 Negative (n = 3,759)	SARS-CoV-2 Positive (n = 754)	OR (95% CI)
CAD / MI / CHF	16.5	16.8	15.0	0.88 (0.70 – 1.09)

*Missing, Unknown, Declined n = 154 (3.4%).
†2018 inflation adjusted USD. Missing n = 76 (1.7%)
‡ Difference in median and 95% CI of difference obtained via quantile regression

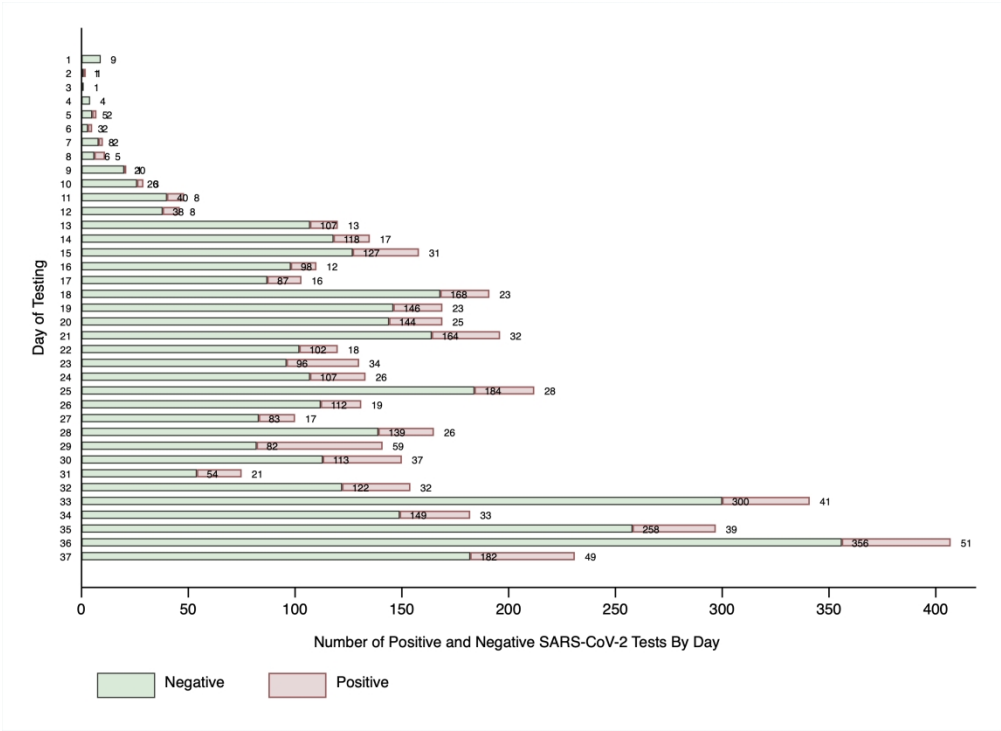
Table 2: Univariable comparison of socio-demographic and comorbidity factors between African American and non-African American

	African American n = 1,159	Non-African American n = 3,354	OR / Median Difference (95% CI)	P value
Age: Mean (SD)	48.8 (17.7)	51.2 (19.2)	0.99 (0.99 – 0.99)	< 0.001
Age Category (%)				
Up to 35	23.6	23.3	Reference	
36 – 50	30.6	26.3	1.25 (1.04 – 1.50)	0.02
51 – 75	37.6	35.3	1.14 (0.96 – 1.36)	0.14
> 75	8.2	13.1	0.67 (0.52 – 0.87)	0.003
Females	66.4	60.6	1.28 (1.12 – 1.48)	<0.001
Median (IQR) Zip Income	60,765 (46,801 – 76,163)	75,793 (57,252 – 102,008)	-15,028 (-1,667, -12,388)	< 0.001
Median Zip Income Pentiles (%) – Pentiles of increasing Income				
Category I	31.8	15.7	Reference	
Category II	22.4	17.2	0.64 (0.53 – 0.79)	< 0.001
Category III	22.0	21.5	0.50 (0.41 – 0.61)	< 0.001
Category IV	11.7	22.4	0.26 (0.21 – 0.32)	< 0.001
Category V	12.0	23.2	0.25 (0.20 – 0.32)	< 0.001
Population Density for Zip: Median (IQR)	3256.8 (2123.6 – 4439.7)	2814.2 (1439.1 – 4260.9)	442.6 (306.7 – 578.5)	< 0.001
Population Density for Zip Pentiles (%) – Pentiles of increasing population density				
Category I	12.3	22.6	Reference	
Category II	23.0	18.2	2.32 (1.85 – 2.93)	< 0.001
Category III	21.3	19.0	1.95 (1.54 – 2.46)	< 0.001
Category IV	22.0	19.0	3.12 (1.69 – 2.68)	< 0.001
Category V	21.5	20.1	1.97 (1.59 – 2.48)	< 0.001
Hypertension	51.7	39.1	1.69 (1.46 – 1.91)	< 0.001
Diabetes	25.0	20.5	1.29 (1.10 – 1.51)	0.001
Obesity	9.8	7.4	1.37 (1.09 – 1.73)	0.008
CAD / MI / CHF	19.1	15.5	1.28 (1.07 – 1.52)	0.006

Table 3: Adjusted Odds Ratios and 95% Confidence Intervals for socio-demographic and comorbidity factors associated with SARS-CoV-2 infection

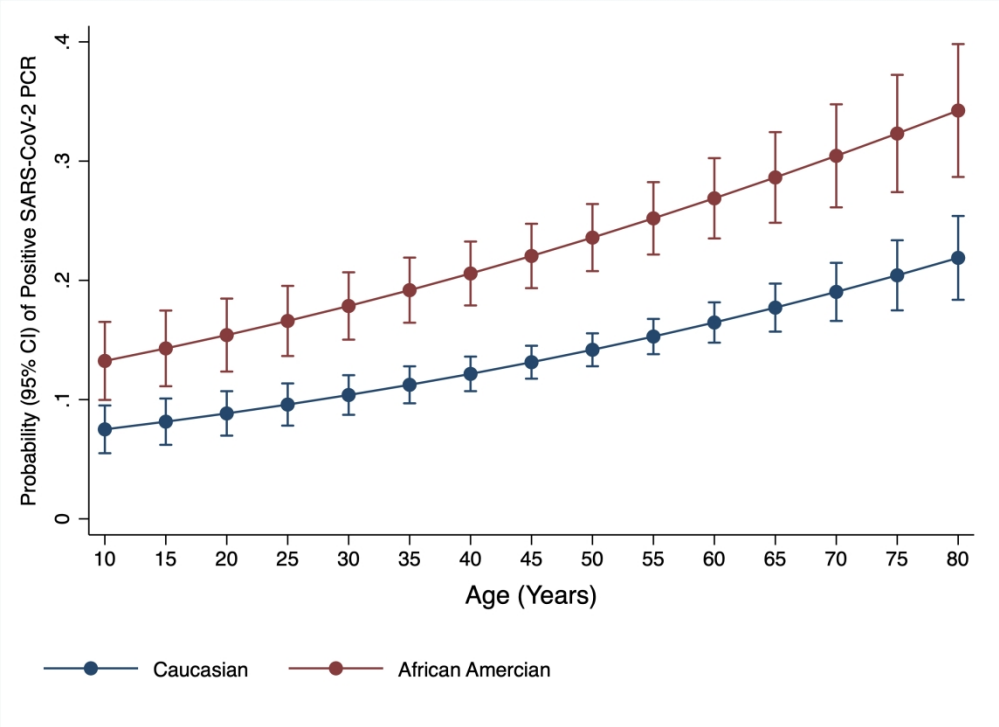
Covariate	Adjusted Odds Ratio	95% Confidence Interval	P value
Age Categories			
Up to 35 years	*Reference Category*		
36–50 years	1.24	0.96 – 1.58	0.09
51–75 years	2.02	1.57 – 2.61	<0.001
>75 years	1.87	1.28 – 2.73	0.001
Male (vs. Female)	1.32	1.12 – 1.57	0.001
Race Categories (Non-Hispanic)			
White	*Reference Category*		
African American	1.84	1.49 – 2.27	<0.001
Asian	1.46	1.09 – 1.95	0.01
More Than One Race / Other / Unknown	0.56	0.34 – 0.90	0.02
Hispanic (vs. Non-Hispanic)	1.70	1.35 – 2.14	<0.001
Median Zip Household Income Categories (Pentiles of Increasing Income)			
Category I	*Reference Category*		
Category II	0.96	0.74 – 1.25	0.77
Category III	0.88	0.68 – 1.15	0.36
Category IV	0.88	0.67 – 1.17	0.39
Category V	0.98	0.74 – 1.30	0.91
Primary Insurance Type			
Medicare	*Reference Category*		
Medicaid	0.69	0.40 – 1.17	0.17
Private / Employer Based	1.18	0.92 – 1.51	0.21
Healthcare Exchange	1.58	0.96 – 2.60	0.30
Self-Pay	1.16	0.87 – 1.54	0.30
Veterans Affairs	0.22	0.05 – 0.93	0.04
Zip Population Density (Pentiles of Increasing Density)			
Category I	*Reference Category*		
Category II	1.08	0.82 – 1.44	0.58
Category III	1.30	0.99 – 1.71	0.06
Category IV	1.32	1.00 – 1.74	0.05
Category V	1.34	1.01 – 1.77	0.04
Hypertension	0.86	0.70 – 1.06	0.15
Diabetes	1.27	1.03 – 1.57	0.03
Obesity	0.95	0.70 – 1.30	0.75
CAD / MI / CHF	0.70	0.55 – 0.91	0.007

**Hosmer and Lemeshow goodness of fit p-value: 0.63 (H_0: Model fit is correct)*



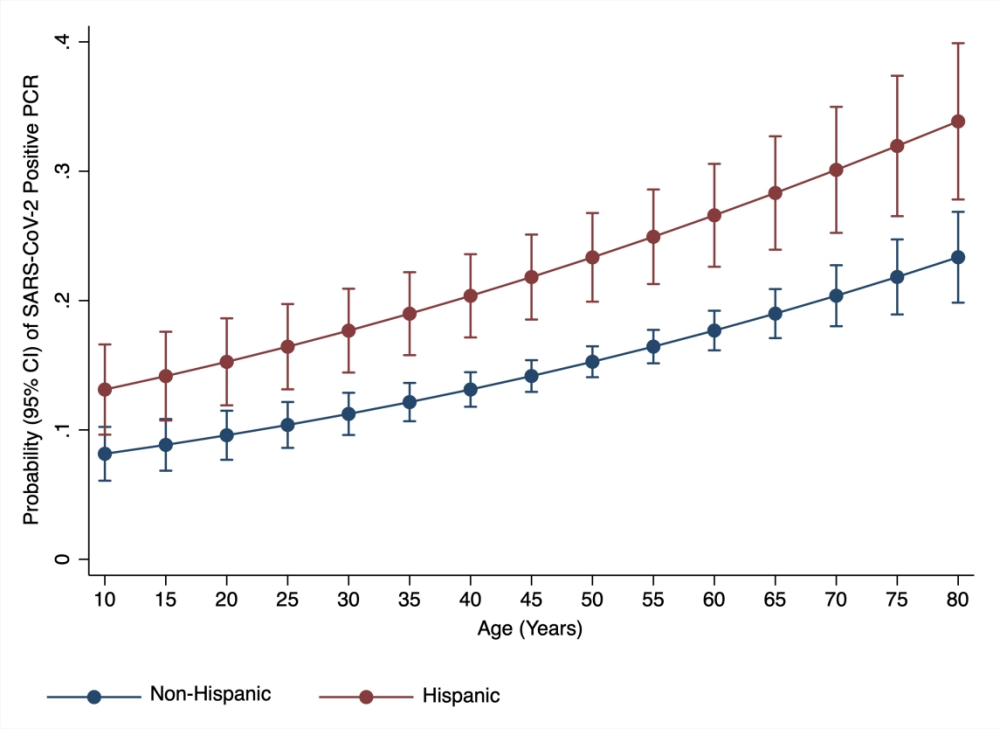
Schematic representation of the temporal sequence for total, positive and negative numbers of SARS-CoV-2 tests in the Houston Methodist CURATOR

489x355mm (144 x 144 DPI)



Adjusted Probability and 95% Confidence Interval of Positive SARS-CoV-2 PCR in African American vs. Caucasian by increasing age

489x355mm (144 x 144 DPI)



Adjusted Probability and 95% Confidence Interval of Positive SARS-CoV-2 PCR in Hispanic vs. Non-Hispanic by increasing age

489x355mm (144 x 144 DPI)

STROBE Statement—Checklist of items that should be included in reports of ***cross-sectional studies***

	Item No	Recommendation	Page No
Title and abstract	1	(*a*) Indicate the study's design with a commonly used term in the title or the abstract	1
		(*b*) Provide in the abstract an informative and balanced summary of what was done and what was found	3
Introduction			
Background/rationale	2	Explain the scientific background and rationale for the investigation being reported	5
Objectives	3	State specific objectives, including any prespecified hypotheses	6
Methods			
Study design	4	Present key elements of study design early in the paper	7
Setting	5	Describe the setting, locations, and relevant dates, including periods of recruitment, exposure, follow-up, and data collection	7
Participants	6	(*a*) Give the eligibility criteria, and the sources and methods of selection of participants	7
Variables	7	Clearly define all outcomes, exposures, predictors, potential confounders, and effect modifiers. Give diagnostic criteria, if applicable	7,8
Data sources/ measurement	8*	For each variable of interest, give sources of data and details of methods of assessment (measurement). Describe comparability of assessment methods if there is more than one group	7
Bias	9	Describe any efforts to address potential sources of bias	N/A
Study size	10	Explain how the study size was arrived at	8
Quantitative variables	11	Explain how quantitative variables were handled in the analyses. If applicable, describe which groupings were chosen and why	8
Statistical methods	12	(*a*) Describe all statistical methods, including those used to control for confounding	8
		(*b*) Describe any methods used to examine subgroups and interactions	8
		(*c*) Explain how missing data were addressed	N/A
		(*d*) If applicable, describe analytical methods taking account of sampling strategy	N/A
		(*e*) Describe any sensitivity analyses	N/A
Results			
Participants	13*	(a) Report numbers of individuals at each stage of study—eg numbers potentially eligible, examined for eligibility, confirmed eligible, included in the study, completing follow-up, and analysed	8
		(b) Give reasons for non-participation at each stage	N/A
		(c) Consider use of a flow diagram	N/A
Descriptive data	14*	(a) Give characteristics of study participants (eg demographic, clinical, social) and information on exposures and potential confounders	9, 19
		(b) Indicate number of participants with missing data for each variable of interest	19,20
Outcome data	15*	Report numbers of outcome events or summary measures	8,9
Main results	16	(*a*) Give unadjusted estimates and, if applicable, confounder-adjusted estimates and their precision (eg, 95% confidence interval). Make clear which confounders were adjusted for and why they were included	9,19,20

		(*b*) Report category boundaries when continuous variables were categorized	19
		(*c*) If relevant, consider translating estimates of relative risk into absolute risk for a meaningful time period	N/A
Other analyses	17	Report other analyses done—eg analyses of subgroups and interactions, and sensitivity analyses	9,10
Discussion			
Key results	18	Summarise key results with reference to study objectives	11
Limitations	19	Discuss limitations of the study, taking into account sources of potential bias or imprecision. Discuss both direction and magnitude of any potential bias	13
Interpretation	20	Give a cautious overall interpretation of results considering objectives, limitations, multiplicity of analyses, results from similar studies, and other relevant evidence	12,14
Generalisability	21	Discuss the generalisability (external validity) of the study results	13
Other information			
Funding	22	Give the source of funding and the role of the funders for the present study and, if applicable, for the original study on which the present article is based	15

*Give information separately for exposed and unexposed groups.

Note: An Explanation and Elaboration article discusses each checklist item and gives methodological background and published examples of transparent reporting. The STROBE checklist is best used in conjunction with this article (freely available on the Web sites of PLoS Medicine at http://www.plosmedicine.org/, Annals of Internal Medicine at http://www.annals.org/, and Epidemiology at http://www.epidem.com/). Information on the STROBE Initiative is available at www.strobe-statement.org.

BMJ Open

Racial and Ethnic Disparities in SARS-CoV-2 Pandemic: Analysis of a COVID-19 Observational Registry for a Diverse U.S. Metropolitan Population

Journal:	*BMJ Open*
Manuscript ID	bmjopen-2020-039849.R1
Article Type:	Original research
Date Submitted by the Author:	06-Jul-2020
Complete List of Authors:	Vahidy, Farhaan S; Houston Methodist Research Institute, Center for Outcomes Research; Houston Methodist Neurological Institute Nicolas, Juan Carlos; Houston Methodist Research Institute Meeks, Jennifer R; Houston Methodist Research Institute Khan, Osman; Houston Methodist Research Institute Pan, Alan; Houston Methodist Research Institute, Center for Outcomes Research Jones, Stephen L.; Houston Methodist Research Institute; Weill Cornell Medicine Masud, Faisal; Houston Methodist Hospital; Weill Cornell Medicine Sostman, H Dirk; Houston Methodist Research Institute; Weill Cornell Medicine Phillips, Robert; Houston Methodist Hospital; Weill Cornell Medicine Andrieni, Julia D; Houston Methodist Hospital; Weill Cornell Medicine Kash, Bita A; Houston Methodist Research Institute; Texas A&M University School of Rural Public Health Nasir, Khurram; Houston Methodist Research Institute; Houston Methodist Hospital
Primary Subject Heading:	Epidemiology
Secondary Subject Heading:	Epidemiology, Infectious diseases, Public health
Keywords:	EPIDEMIOLOGY, Public health < INFECTIOUS DISEASES, INFECTIOUS DISEASES

I, the Submitting Author has the right to grant and does grant on behalf of all authors of the Work (as defined in the below author licence), an exclusive licence and/or a non-exclusive licence for contributions from authors who are: i) UK Crown employees; ii) where BMJ has agreed a CC-BY licence shall apply, and/or iii) in accordance with the terms applicable for US Federal Government officers or employees acting as part of their official duties; on a worldwide, perpetual, irrevocable, royalty-free basis to BMJ Publishing Group Ltd ("BMJ") its licensees and where the relevant Journal is co-owned by BMJ to the co-owners of the Journal, to publish the Work in this journal and any other BMJ products and to exploit all rights, as set out in our licence.

The Submitting Author accepts and understands that any supply made under these terms is made by BMJ to the Submitting Author unless you are acting as an employee on behalf of your employer or a postgraduate student of an affiliated institution which is paying any applicable article publishing charge ("APC") for Open Access articles. Where the Submitting Author wishes to make the Work available on an Open Access basis (and intends to pay the relevant APC), the terms of reuse of such Open Access shall be governed by a Creative Commons licence – details of these licences and which Creative Commons licence will apply to this Work are set out in our licence referred to above.

Other than as permitted in any relevant BMJ Author's Self Archiving Policies, I confirm this Work has not been accepted for publication elsewhere, is not being considered for publication elsewhere and does not duplicate material already published. I confirm all authors consent to publication of this Work and authorise the granting of this licence.

Racial and Ethnic Disparities in SARS-CoV-2 Pandemic: Analysis of a COVID-19 Observational Registry for a Diverse U.S. Metropolitan Population

Farhaan S. Vahidy PhD MBBS MPH,[1,2] Juan Carlos Nicolas BS,[1] Jennifer R. Meeks MS,[1] Osman Khan BS,[1] Alan Pan MS,[1] Stephen L. Jones MD MSHI,[1,3,4] Faisal N. Masud MD,[4,5] H. Dirk Sostman MD,[4,6] Robert A. Phillips MD PhD,[1,4,7] Julia D. Andrieni MD,[4,8] Bita Kash PhD MBA,[1,4,9] Khurram Nasir MD MPH[1,7]

1. Center for Outcomes Research, The Houston Methodist Research Institute, Houston TX
2. Houston Methodist Neurological Institute, Houston Methodist Hospital, Houston TX
3. Department of Surgery, Houston Methodist Hospital, Houston TX
4. Weill Cornell Medical College, New York NY
5. Department of Anesthesiology and Critical Care, Houston Methodist Hospital, Houston TX
6. The Houston Methodist Research Institute and The Houston Methodist Academic Institute, Houston TX
7. Department of Cardiology, DeBakey Heart and Vascular Center, Houston Methodist Hospital, Houston TX
8. Department of Medicine, Houston Methodist Hospital, Houston TX
9. Department of Health Policy and Management, Texas A&M University School of Public Health, College Station TX

Corresponding Author

Farhaan S. Vahidy, PhD MBBS MPH
Josie Roberts Building, Suite 4.123
Houston Methodist
7550 Greenbriar Drive
Houston TX 77030
fvahidy@houstonmethodist.org

Word Count: 3,062

Abstract

Introduction: Data on race and ethnic disparities for SARS-CoV-2 infection are limited. We analyzed socio-demographic factors associated with higher likelihood of SARS-CoV-2 infection and explore mediating pathways for race and ethnic disparities in the SARS-CoV-2 pandemic.

Methods: Cross-sectional analysis of COVID-19 Surveillance and Outcomes Registry (CURATOR), which captures data for a large healthcare system, comprising of one central tertiary care hospital, seven large community hospitals, and an expansive ambulatory / emergency care network in the Greater Houston area. Nasopharyngeal samples for individuals inclusive of all ages, races, ethnicities and sex were tested for SARS-CoV-2. We analyzed socio-demographic (age, sex, race, ethnicity, household income, residence population density) and comorbidity (Charlson Comorbidity Index, hypertension, diabetes, obesity) factors. Multivariable logistic regression models were fitted to provide adjusted Odds Ratios (aOR) and 95% confidence intervals (CI) for likelihood of a positive SARS-CoV-2 test. Structural Equation Modeling (SEM) framework was utilized to explore three mediation pathways (low income, high population density, high comorbidity burden) for association between Non-Hispanic Black race (NHB), Hispanic ethnicity, and SARS-CoV-2 infection.

Results: Among 20,228 tested individuals, 1,551 (7.7%) tested positive. Overall mean (SD) age was 51.1 (19.0) years, 62% females, 22% Black and 18% were Hispanic. NHB and Hispanic ethnicity was associated with lower socio-economic status and higher population density residence. In the fully adjusted model, NHB (vs. NHW; aOR, CI: 2.23, 1.90-2.60) and Hispanic ethnicity (vs. non-Hispanic; aOR, CI: 1.95, 1.72-2.20) had a higher likelihood of infection. Older individuals and males were also at higher risk of infection. The SEM framework demonstrated a

statistically significant indirect effect of NHB and Hispanic ethnicity on SARS-CoV-2 infection mediated via a pathway including residence in densely populated zip code.

Conclusions: There is strong evidence of race and ethnic disparities in the SARS-CoV-2 pandemic, potentially mediated through unique social determinants of health.

Strengths and limitations of this study

- One of the first studies to systematically evaluate race and ethnic disparities in susceptibility to SARS-CoV-2 infection, while accounting for multiple socio-demographic characteristics and comorbidities
- Study population represents a large and diverse metropolitan of the U.S. with data from one of the largest healthcare providers across the greater metropolitan area
- Study evaluates potential mediation pathways for race disparities and demonstrates that residence in areas with high population density may mediate race and ethnic disparities in susceptibility to SARS-CoV-2 infection
- Single center study with limited information about burden of comorbidity and lifestyle factors

INTRODUCTION

The Coronavirus disease (COVID-19), caused by infection with the Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2), is a pandemic that has thus far resulted in over 9.5 million cases globally in under 6 months. At the time of this reporting, the United States (U.S.) has approximately 25% of total global cases and has surpassed all countries in terms of absolute number of cases, cases per 1 million population, and fatalities.[1,2] Experts project these numbers to continue rising as widespread testing is instituted and newer patterns of infectivity emerge. The geographic distribution of cases across the U.S. demonstrates that the predominant pandemic burden hit major metropolitan areas. However, cases of COVID-19 have been reported across all 50 states, the District of Columbia, Guam, Puerto Rico, the Northern Mariana Islands, and the U.S. Virgin Islands.[3] As of May 31, 2020, the state of Texas had 64,287 reported cases of COVID-19, with about one-third in the Greater Houston area.[4] The Greater Houston area is home to approximately 7 million individuals, is the fourth-largest metropolitan area by population in the U.S., and is considered one of the nation's most diverse regions.[5-6]

Initial reports indicate that specific individuals such as the elderly; males; and people with comorbidities including hypertension, diabetes, obesity, coronary artery disease and heart failure have poor COVID-19 outcomes.[7–10] As the pandemic spread over the continental U.S. during the last four months, patterns of high-risk phenotypes started to emerge and reports of poor outcomes (particularly high case fatality) among racial minorities surfaced.[11–13] Though it is important to understand the determinants of poor outcomes among COVID-19 patients, it is equally imperative, from a public health perspective, to systematically examine the likelihood of SARS-CoV-2 infection across large diverse communities in the U.S. Data on higher likelihood of SARS-CoV-2 infection among racial and ethnic minorities across diverse U.S. metropolitan

areas are limited. Furthermore, the mediators of SARS-CoV-2 infection among racial and ethnic minorities have not been described.

We explored socio-demographic characteristics such as age, sex, race, ethnicity, median household income by zip codes, population density of residents' zip codes, and health insurance status associated with positive SARS-CoV-2 testing in an urban and diverse population served by one of the leading healthcare systems of the Greater Houston area. We further examined the association between pre-existing comorbidities and higher likelihood of SARS-CoV-2 infection in our study population. We hypothesized that older age, and racial and ethnic minorities will be associated with significantly higher likelihood of SARS-CoV-2 infection, and factors such as low socio-economic status, residence in high population density areas (proxy for potential difficulties in social distancing) and higher comorbidity burden will mediate the effect of race and ethnicity on SARS-CoV-2 infection.

METHODS

We analyzed data between March 5 and May 31, 2020 collected as a part of the COVID-19 Surveillance and Outcomes Registry (CURATOR) at Houston Methodist (HM). The HM CURATOR has been approved by the HM Institutional Review Board (IRB) as an observational quality of care registry for all suspected and confirmed COVID-19 patients. HM IRB granted CURATOR a waiver of informed consent and HIPAA (Health Insurance Portability and Accountability Act) authorization in accordance with current federal regulations. The CURATOR, designed and managed by the big data team at the Center for Outcomes Research (COR) at HM, is populated from multiple data sources across the HM system such as electronic medical records, electronic databanks for laboratory and pharmacy, and electronic interactive

patient interface tools. The HM system comprises a flagship tertiary care hospital in the Texas Medical Center, seven large community hospitals, a continuing care hospital, and multiple emergency centers and clinics throughout the Greater Houston area. Data from various sources are curated into a harmonized format, assessed for quality and integrity, and stored on a secure institutional HIPAA-compliant server.

We flagged all individuals who were tested for the SARS-CoV-2 using the real time Reverse Transcriptase (RT) Polymerized Chain Reaction (PCR) diagnostic panels. The three cross-validated PCR tests utilized were the World Health Organization (WHO) nucleic acid amplification test, Panther Fusion® SARS-CoV-2 Assay, and Cepheid Xpert® Xpress SARS-CoV-2 Assay. These assays were verified for quantitative detection of novel SARS-CoV-2 isolated and purified from nasopharyngeal swab specimens obtained from individuals and immersed in universal transport medium. Testing was carried out for symptomatic individuals or for individuals who had a self-reported history of exposure to a COVID-19 case including recent travel to other countries with high infection rates or hotspots within the U.S.

Socio-demographic characteristics including age, sex, race, ethnicity, and payer-status (insurance type) were obtained from the HM CURATOR for analyses. We also extracted information on presence of comorbidities comprising the Charlson Comorbidity Index (CCI) which include past history of myocardial infarction, congestive heart failure, peripheral vascular disease, cerebrovascular disease, dementia, chronic pulmonary disease, rheumatic disease, peptic ulcer disease, liver disease, diabetes with or without complications, hemiplegia, renal disease, any malignancy (excluding skin neoplasms), metastatic solid tumors, and AIDS/HIV. Data on hypertension and obesity were additionally obtained. We utilized the U.S. Census Bureau's American Community Survey (ACS) 5-year data (2014–2018) to determine median household

income by individual zip code tabulation areas (ZCTA).[14] The median ZCTA household income was inflation-adjusted to 2018 USD. We also utilized the same data source to obtain population estimates by ZCTA, and calculated ZCTA level population density (population per mile square) by standardizing it for area measurements of ZCTA. For the purpose of population density determination, land area estimates were obtained from the Census Bureau's U.S. Gazetteer Files 2010.[15] In the absence of granular and precise social distancing data, we have utilized population density as a proxy for potential difficulties in social distancing among crowded communities.

We provide descriptive summary data as means (standard deviations) and proportions. We fit univariable and multivariable logistic regression models to assess unadjusted and adjusted association between socio-demographic characteristics and likelihood of being tested positive for SARS-CoV-2. We additionally provide univariable comparison of various socio-demographic and comorbidity variables between non-Hispanic Black (NHB) and non-Hispanic White (NHW) race categories, as well as between Hispanic and Non-Hispanic ethnic groups. Age, income, population density and CCI were categorized for certain analyses. We determined *a priori* to include all variables (age, sex, race, ethnicity, zip code household income, insurance type, zip population density and CCI) in our initial multivariable model. Factors demonstrating mediation were excluded from the final model. We assessed the model fit utilizing the Hosmer-Lemeshow goodness of fit test, and crude and adjusted odds ratios (OR and aOR) and 95% confidence intervals (CI) are reported. Post-estimation marginal probabilities of SARS-CoV-2 infection were determined from the fully adjusted model for major covariates (race, ethnicity and age). We explored the mediation influence of comorbidity burden (CCI), socio-economic status (median income), and lack of social distancing (population density) on the relationship of Black race and Hispanic ethnicity with high likelihood of SARS-CoV-2 infection using the Generalized

Structural Equation Modeling (GSEM) framework. The GSEM framework was set up to provide estimates of direct and indirect effect of Black race and Hispanic ethnicity on SARS-CoV-2 infectivity. Statistically significant ($p < 0.05$) indirect effects represent full or partial mediation by a tested covariate. We included all individuals tested for SARS-CoV-2 across our healthcare system and did not perform formal sample size calculations.

Patient and Public Involvement

There was no direct patient or public involvement in the design and conduct of this study.

RESULTS

Socio-demographic and comorbidity characteristics of the study population

Across the time period of analysis, we identified a total of 20,228 presumed cases tested for SARS-CoV-2, among whom 1,551 (7.7%, CI: 7.3-8.0) tested positive. Overall, the mean (SD) age of the study population was 51.1 (19.0) years; 61.9% were female and 62.3% were White (including Hispanic ethnicity). The study sample was comparable to the overall population of patients treated across HM, who have a mean (SD) age of 49.0 (22) years, are 56% female, and 53% White. The HM system metrics was derived from a sample of 3,216,290 patients managed across the system since May 22, 2016.

The overall median (IQR) household income was USD $70,658 ($53,313–$99,276), and 42.6% of the study population had private or employer-based insurance. In our univariate analysis, Black race (vs. White; OR, CI: 1.55, 1.37–1.75), Hispanic ethnicity (vs. non-Hispanic; OR, CI: 2.02, 1.79–2.27), and males (vs. females; OR, CI: 1.17, 1.06–1.31) were associated with significantly higher likelihood of testing positive for SARS-CoV-2. Among the SARS-CoV-2 positive patients, 40.8% were in the age category of 51–75 years, and 11.4% were greater than

75 years. These proportions were significantly higher than the reference group (up to 35 years; OR, CI for 51-75 years vs. up to 35 years: 1.29, 1.12–1.48 and for >75 years vs. up to 35 years: 1.23, 1.02–1.49). Furthermore, individuals in higher pentiles of socio-economic status had significantly lower likelihood, whereas those residing in higher population density ZCTAs had higher likelihood of SARS-CoV-2 infection. We observed a significantly higher proportion of SARS-CoV-2 positive individuals in the CCI 1-2 category compared to CCI of 0 (OR, CI: 1.35, 1.18–1.54). However, similar differences for higher CCI categories were not observed. For specific comorbidities, a significantly greater proportion of diabetic individuals had SARS-CoV-2 positive results (OR, CI: 1.40, 0.17–1.68). The socio-demographic characteristics and comorbidity profiles for the overall and SARS-CoV-2 positive and negative patients are summarized in Table 1.

Socio-demographic and comorbidity characteristics associated with minority race and ethnicity

In our study sample comprising of 13,754 Non-Hispanic Black and White individuals, we compared the association between race and various socio-demographic and comorbidity characteristics (Table 2). Similarly, we also evaluated univariable differences for socio-demographic variables and co-morbidities between Hispanic and non-Hispanic individuals (Table 3). Minority race (NHB) and ethnicity (Hispanic) were both associated with younger age, higher proportion of females, and residence in low income and higher population density ZCTAs. However, NHB and Hispanic groups were both associated with an overall lower burden of comorbidities (as demonstrated by significantly lower median CCI) compared respectively to NHW and non-Hispanic categories. A higher proportion of individuals among minority race and

ethnicity were diabetic, and a higher proportion of NHB were also hypertensive compared to NHW.

Multivariable model and marginal probabilities for likelihood of SARS-CoV-2 infection and racial and ethnic minorities

The significantly higher likelihood of SARS-CoV-2 infection among minority race and ethnic groups persisted after controlling for other demographics, insurance type, median household income, population density, and comorbidities. Adjusted odds ratios (CI) for NHB vs. NHW was 2.23 (1.90 – 2.60) and for Hispanic vs. Non-Hispanic was 1.95 (1.72 – 2.20). Higher risk of infection among males (compared to females) and higher likelihood of SARS-CoV-2 infection among elderly also remained statistically significant. Detailed outputs of the fully adjusted logistic regression models for minority race and ethnic groups are presented in Table 4. Based on the marginal probabilities obtained from our fully adjusted model, the probability of SARS-CoV-2 infection in a 45-year-old NHB is 9.6% whereas it is 4.5% in a 45-year-old NHW individual, all other adjusted variables being constant. At the age of 75, this probability is 14.0% for an NHB and 6.9% for a NHW. A similar relationship differential was observed for Hispanic vs. non-Hispanic individuals. Multivariable model derived probabilities of SARS-CoV-2 infection for NHB vs. NHW and for Hispanic vs. Non-Hispanic across age spectrum are presented in Figure 1 and Figure 2.

Generalized Structural Equation Modeling for mediation by income, population density and Comorbidity Index

Utilizing the GSEM framework, we determined the direct and indirect effects of NHB and Hispanic ethnicity on SARS-CoV-2 infection with median income, population density and CCI modeled as mediators in six separate equations adjusted for age and sex. The indirect effect

of NHB mediated through population density was statistically significant (OR, CI: 1.03, 1.01 – 1.05, p = 0.001); however, the indirect effects mediated via median income and comorbidity scores were not statistically significant (p = 0.14 and p = 0.64 respectively). Among individuals identifying as Hispanic or Latino, both population density and income partially mediated the effect of ethnicity on SARS-CoV-2 positivity (OR, CI for population density: 1.02, 1.01 – 1.02, p < 0.001 and OR, CI for income: 1.04, 1.02 – 1.06, p < 0.001). Evaluation of comorbidities did not suggest a mediation influence for either NHB or Hispanic categories.

DISCUSSION

The underlying race and ethnic healthcare disparities have been painfully highlighted in the wake of the COVID-19 pandemic. Most reports indicate higher mortality or case fatality among minority racial groups (Black / African American) across major U.S. metropolitan areas.[11–13] However, robust insights on the racial differences for SARS-CoV-2 infection are limited. Furthermore, comprehensive data evaluating higher susceptibility to SARS-CoV-2 infection among Hispanic communities are also scarce. This is perhaps because of comparatively homogenous populations in non-U.S. regions of the world. Houston, as an exceptionally ethnically diverse population center,[16] is well suited for an investigation of racial, ethnic, and socioeconomic gradients in COVID-19 test positivity. We focus on highlighting the mechanisms of racial and ethnic disparities in susceptibility to SARS-CoV-2 infection and provide evidence of mediation of such disparities by novel social determinants of health (SDoH).

Our study adds to the current literature by analyzing emerging data for individuals being tested across one of the largest healthcare systems in the Greater Houston area. We report that racial and ethnic minorities (non-Hispanic Black and Hispanic individuals) are almost twice as likely to test positive for SARS-CoV-2 than the non-Hispanic White and non-Hispanic

population. These findings illuminate systematic racial / ethnic disparities in testing positive for SARS-CoV-2 infection. Though there are limited prior SARS-CoV-2 data, such racial and ethnic disparities have previously been described for the U.S. H1N1 influenza pandemic.[17] These data indicated that Spanish-speaking Hispanic and Black individuals were at a greater risk of H1N1 infection, primarily attributable to lack of healthcare access.

We explored three possible mechanisms of race disparities in our data. These included lower socio-economic status, residence in higher population dense areas, and higher level of comorbidities. We demonstrate that NHB race is significantly associated with all three potential disparity pathways, and in the traditional multivariable analyses, racial and ethnic disparities persisted even after controlling for these pathways. However, our mediation analyses highlighted the potential influence of residence in high population density areas as a viable pathway that at least partially explains the observed racial and ethnic disparity. Furthermore, residence in low income areas emerged as a significant mediation pathway for ethnic differences in SARS-CoV-2 positivity. Pathways mediating the influence of comorbidity status did not demonstrate a significant effect. We utilized population density as a marker for potential inability to maintain adequate social distancing as it has been indicated that maintaining the WHO recommended safe distance between people becomes challenging with high population densities.[18] Furthermore, overall effects of population density and disease spread has been previously described in literature.[19,20] In addition to lack of social distancing, higher population density may also be associated with several other behavioral and socio-demographic attributes that may predispose populations to both viral spread and increased susceptibility. For example, there are reports linking obesity, lack of physical activity, and higher mortality with residence in densely populated neighborhoods.[21,22]

As reported, our data also corroborate that older populations may be more susceptible to SARS-CoV-2 infection.[10] However, younger populations still have cause for concern as nearly 1 in 4 of the infected cases in our sample were between 36–50 years of age. Finally, our data demonstrate that males may be approximately 20% more likely to test positive for the SARS-CoV-2 infection. Potential sex differences in infectivity to SARS-CoV-2 and intersectionality with racial and ethnic socioeconomic factors need to be explored further in future analyses. Additional policy-oriented research should prioritize studying the intersectionality of these vulnerable economic statuses and racial disparities in COVID infection indicated by the present study.

Findings of our study need to be interpreted in the light of certain limitations. Our data are from a single center and may not be generalizable to the wider U.S. population. These findings need to be replicated in larger data sets across other large heterogenous U.S. metropolitans. However, the Houston metropolitan area is one of the most diverse and representative in the U.S.,[16] and our healthcare system is one of the largest systems providing care to COVID-19 patients in the Greater Houston area. Our sample was composed of 22% Black, 18% Hispanic, and 62% female population. We did not have information on certain demographic covariates such as education or household size. Educational status has been linked to healthcare awareness and may be important to adjust for in analyses of potential disparities, and household size may be used to provide more precise estimates of socio-economic status. However, we obtained and adjusted for zip code income data from the U.S. Census, as income has previously been shown to have strong correlation with educational attainment and socio-economic status.[23] Since testing was based on suspicion of infection and may have been influenced by factors such as access to care, the potential for selection bias cannot be ruled out.

Furthermore, lack of sensitivity of SARS-CoV-2 diagnostics tests have been reported; however, the three assays utilized for testing were cross validated for internal consistency. Finally, we did not have detailed information on comorbidities and their management in the study population. However, we did control for major comorbidities which are being reported as associated with COVID-19 outcomes.[24]

Conclusions

The strong association between racial and ethnic minorities and SARS-CoV-2 infection demonstrated in our data, even after adjustment for other important socio-demographic and comorbidity factors, highlight a potential catastrophe of inequality within the existential crisis of a global pandemic. Our data, representing a large heterogeneous U.S. metropolitan area, also provide preliminary evidence into the potential pathways for this disparity. It is highly likely that higher comorbidity burden and detrimental effects of adverse social determinants, including those that may not adequately permit safe practices of social distancing, mediate higher SARS-CoV-2 infectivity among racial and ethnic minorities.

As the pandemic continues to spread and evolve across the continental U.S., emerging data on association between SARS-CoV-2 infection and various socio-demographic factors will continue to enhance our understanding of targeted risks related to SARS-CoV-2 infection, and such data would enable us to comprehend healthcare services and access factors related to development and outcomes of COVID-19 among minority populations. Our findings substantiate prior calls for collection of robust data on race and ethnicity as a part of international collaborations,[25] and further drive home the critical importance of quantifying novel SDoH.

Acknowledgment: The authors thank Jacob M. Kolman, Senior Scientific Writer of the Houston Methodist Center for Outcomes Research, for reviewing the language and format of the manuscript.

IRB Approval: This work was carried out under an approved protocol for the Houston Methodist COVID-19 Surveillance and Outcomes Registry (HM CURATOR) by the Houston Methodist Research Institute Institutional Review Board (HMRI IRB).

Contribution Statement:

FV: design, data analysis and interpretation, drafting the manuscript, critical revision for important intellectual content, final approval

JCN: data acquisition, data analysis, drafting the manuscript, final approval

JRM: data acquisition, drafting the manuscript, final approval

OK: data acquisition, data analysis, drafting the manuscript, final approval

AP: data acquisition, data analysis, drafting the manuscript, final approval

SLJ: data acquisition, data interpretation, critical revision for important intellectual content, final approval

FNM: critical revision for important intellectual content, final approval

HDS: critical revision for important intellectual content, final approval

RAP: critical revision for important intellectual content, final approval

JDA: critical revision for important intellectual content, final approval

BK: critical revision for important intellectual content, final approval

KN: design, interpretation of data, critical revision for important intellectual content, final approval

Competing Interests: No authors declare a competing interest to this work

Funding: There are no funders to report for this submission.

Date sharing statement: All requests for de-identified data should be made to the corresponding author. All reasonable requests will be evaluated by the CURATOR Data Governance and

Sharing Committee comprising of FV, SLJ, BK and KN in the light of institutional policies and guidelines.

References

1. World Health Organization. WHO COVID-19 dashboard. https://covid19.who.int/. Published April 6, 2020. Accessed April 6, 2020.

2. Center for Systems Science and Engineering (CSSE) at Johns Hopkins University. COVID-19 map. Johns Hopkins Coronavirus Resource Center. https://coronavirus.jhu.edu/map.html. Published April 6, 2020. Accessed April 6, 2020.

3. Centers for Disease Control and Prevention. Coronavirus Disease 2019 (COVID-19) in the U.S. cdc.gov. https://www.cdc.gov/coronavirus/2019-ncov/cases-updates/cases-in-us.html. Published April 14, 2020. Accessed April 14, 2020.

4. Texas Health and Human Services. COVID-19 in Texas (dashboard). https://txdshs.maps.arcgis.com/apps/opsdashboard/index.html#/ed483ecd702b4298ab01e8b9cafc8b83. Published April 18, 2020. Accessed April 18, 2020.

5. McCann A. Most diverse cities in the U.S. Wallethub. https://wallethub.com/edu/most-diverse-cities/12690/. Published April 10, 2019. Accessed June 29, 2020.

6. Mejia B. How Houston has become the most diverse place in America. *Los Angeles Times*. https://www.latimes.com/nation/la-na-houston-diversity-2017-htmlstory.html. Published May 9, 2017. Accessed June 29, 2020.

7. Shi S, Qin M, Shen B, et al. Association of cardiac injury with mortality in hospitalized patients with COVID-19 in Wuhan, China. *JAMA Cardiol*. March 2020. doi:10.1001/jamacardio.2020.0950

8. Bonow RO, Fonarow GC, O'Gara PT, Yancy CW. Association of coronavirus disease 2019 (COVID-19) with myocardial injury and mortality. *JAMA Cardiol*. March 2020. doi:10.1001/jamacardio.2020.1105

9. 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*. April 2020. doi:10.1001/jama.2020.5394

10. Zhou F, Yu T, Du R, et al. Clinical course and risk factors for mortality of adult inpatients with COVID-19 in Wuhan, China: a retrospective cohort study. *Lancet*. 2020;395(10229):1054-1062. doi:10.1016/S0140-6736(20)30566-3

11. Deslatte M. Louisiana data: virus hits blacks, people with hypertension. *AP News*. https://apnews.com/cde897293f88f788d3f3bfdb3d46c466. Published April 7, 2020. Accessed April 20, 2020.

12. Reyes C, Husain N, Gutowski C, St. Clair S, Pratt G. Chicago's coronavirus disparity: Black Chicagoans are dying at nearly six times the rate of white residents, data show. *Chicago Tribune*. https://www.chicagotribune.com/coronavirus/ct-coronavirus-chicago-coronavirus-

deaths-demographics-lightfoot-20200406-77nlylhiavgjzb2wa4ckivh7mu-story.html. Published April 7, 2020. Accessed April 20, 2020.

13. Thebault R, Tran AB, Williams V. The coronavirus is infecting and killing black Americans at an alarmingly high rate. *Washington Post*. https://www.washingtonpost.com/nation/2020/04/07/coronavirus-is-infecting-killing-black-americans-an-alarmingly-high-rate-post-analysis-shows/. Published April 7, 2020. Accessed April 20, 2020.

14. United States Census Bureau. American Community Survey (ACS). https://www.census.gov/programs-surveys/acs. Published 2018. Accessed April 14, 2020.

15. United States Census Bureau. Gazetteer Files. https://www.census.gov/geographies/reference-files/time-series/geo/gazetteer-files.html. Accessed April 20, 2020.

16. Harris County Public Health. *Harris Cares: A 2020 Vision of Health in Harris County*. Houston, TX: Harris County Public Health; 2019. http://publichealth.harriscountytx.gov/Resources/Harris-Cares. Accessed November 21, 2019.

17. Quinn SC, Kumar S, Freimuth VS, Musa D, Casteneda-Angarita N, Kidwell K. Racial disparities in exposure, susceptibility, and access to health care in the US H1N1 influenza pandemic. *Am J Public Health*. 2011;101(2):285-293. doi:10.2105/AJPH.2009.188029

18. Rocklov J, Sjodin H. High population densities catalyze the spread of COVID-19. J Travel Med. 2020. doi: 10.1093/jtm/taaa038

19. Tarwater PM, Martin CF. Effects of population density on the spread of disease. *Complexity*. 2001;6(6):29-36. doi:10.1002/cplx.10003

20. Yang K, LeJeune J, Alsdorf D, Lu B, Shum CK, Liang S. Global distribution of outbreaks of water-associated infectious diseases. *PLoS Negl Trop Dis*. 2012;6(2):e1483. doi:10.1371/journal.pntd.0001483

21. Wang R, Feng Z, Xue D, Liu Y, Wu R. Exploring the links between population density, lifestyle, and being overweight: secondary data analyses of middle-aged and older Chinese adults. *Health Qual Life Outcomes*. 2019;17(1):100. doi:10.1186/s12955-019-1172-3

22. Beenackers MA, Oude Groeniger J, Kamphuis CBM, Van Lenthe FJ. Urban population density and mortality in a compact Dutch city: 23-year follow-up of the Dutch GLOBE study. *Health & Place*. 2018;53:79-85. doi:10.1016/j.healthplace.2018.06.010

23. U.S. Bureau of Labor Statistics. Unemployment rates and earnings by educational attainment. https://www.bls.gov/emp/chart-unemployment-earnings-education.htm. Published September 4, 2019. Accessed April 15, 2020.

24. Centers for Disease Control and Prevention. Groups at higher risk for severe illness. cdc.gov. https://www.cdc.gov/coronavirus/2019-ncov/need-extra-precautions/groups-at-higher-risk.html. Last updated April 17, 2020. Accessed April 20, 2020.

25. Abuelgasim E, Saw LJ, Shirke M, Zeinah M, Harky A. COVID-19: Unique public health issues facing Black, Asian and minority ethnic communities. *Curr Probl Cardiol*. 2020;100621. doi:10.1016/j.cpcardiol.2020.100621

Tables and Figures

Figure 1: Adjusted Probability and 95% Confidence Interval of Positive SARS-CoV-2 PCR in Non-Hispanic Black vs. Non-Hispanic White by increasing age

Figure 2: Adjusted Probability and 95% Confidence Interval of Positive SARS-CoV-2 PCR in Hispanic vs. Non-Hispanic by increasing age

Table 1: Summary measures and univariable association of socio-demographic characteristics with SARS-CoV-2 infection from HM CURATOR

Characteristics	Overall (n = 20,228)	SARS-CoV-2 Negative (n = 18,677)	SARS-CoV-2 Positive (n = 1,551)	OR (95% CI)[a]
Age, mean (SD)	51.1 (19.0)	51.0 (19.1)	52.1 (18.1)	1.00 (1.00–1.01)
Age Categories (%)				
Up to 35 years	25.1	25.5	20.9	Reference Category
36–50 years	25.0	24.9	27.0	1.32 (1.14 – 1.54)
51–75 years	38.7	38.5	40.8	1.29 (1.12 – 1.48)
>75 years	11.2	11.2	11.4	1.23 (1.02 – 1.49)
Females (%)	61.9	62.3	58.3	0.85 (0.76 – 0.94)
Race (%)[b]				
White	62.3	62.9	55.7	Reference Category
Black	21.6	21.0	28.8	1.55 (1.37 – 1.75)
Asian	9.2	9.1	9.7	1.20 (1.00 – 1.44)
Mixed / Other	1.3	1.2	1.5	1.36 (0.88 – 2.11)
Hispanic (%)[c]	17.8	16.8	29.2	2.02 (1.79 – 2.27)
Median Zip Household Income (IQR)[d]	70,658 (53,313–99,276)	70,758 (53,633–100,107)	66,523 (50,485–94,226)	-4,168[e] (-6463.2, -1872.8)
Median Zip Household Income Pentiles (%)				
I: 13,893 – 50,485	18.2	17.8	23.6	Reference Category
II: 50,642 – 65,805	18.4	18.1	21.2	0.89 (0.75 – 1.04)
III: 65,897 – 79,869	18.3	18.2	19.6	0.82 (0.70 – 0.96)
IV: 80,039 – 106,067	18.6	19.0	13.7	0.55 (0.46 – 0.65)
V: 106,415 – 240,417	17.2	17.3	16.1	0.71 (0.60 – 0.83)
Median (IQR) Population Density[f]	2797.2 (1439.1 – 4260.9)	2797.2 (1439.1 – 4211.4)	3320.3 (1904.4 – 4439.7)	523.1[e] (454.9 – 591.3)
Median Population Density Pentiles (%)				
I: 1.5 – 1026.6	18.2	18.6	13.5	Reference Category
II: 1034.6 – 2306.3	19.0	19.1	18.1	1.31 (1.09 – 1.58)
III: 2330.8 – 3328.9	17.4	17.4	18.3	1.45 (1.21 – 1.75)
IV: 3360.1 – 4665.6	18.1	17.8	22.1	1.71 (1.43 – 2.05)
V: 4742.6 – 98025.9	18.1	17.7	22.6	1.76 (1.47 – 2.10)
Insurance Status (%)				
Medicare	29.0	29.1	27.2	Reference Category
Medicaid	4.7	4.8	4.3	0.96 (0.73 – 1.26)
Pvt / Employer based	42.6	43.2	36.4	0.90 (0.79 – 1.03)
HC Exchange	1.7	1.6	3.0	2.02 (1.46 – 2.80)
Self-Pay	20.6	20.0	28.4	1.52 (1.33 – 1.75)
VA	1.3	1.4	0.7	0.54 (0.29 – 1.00)
Charlson Co-morbidity Index, Median (IQR)	2 (0 – 6)	2 (0 – 6)	2 (0 – 5)	0 (-0.36, 0.36)[f]
Charlson Co-morbidity Index (CCI) Categories (%)				
CCI: 0	33.1	33.4	30.4	Reference Category

Characteristics	Overall (n = 20,228)	SARS-CoV-2 Negative (n = 18,677)	SARS-CoV-2 Positive (n = 1,551)	OR (95% CI)[a]
CCI: 1 – 2	23.7	23.1	28.6	1.35 (1.18 – 1.54)
CCI: 3 – 6	20.3	20.2	20.4	0.10 (0.65 – 1.28)
CCI: > 6	22.9	23.1	20.6	0.98 (0.84 – 1.13)
Hypertension	47.2	47.1	48.4	1.06 (0.95 – 1.17)
Diabetes (without complications)	24.2	23.7	30.3	1.40 (1.24– 1.57)
Obesity	28.0	28.2	25.2	0.86 (0.76 – 0.96)

[a] Unadjusted Odds Ratios and 95% Confidence Intervals for association between individual co-variates in SRAS-CoV-2 positivity.
[b] Race: Missing, Unknown, Declined, n = 1157 (5.7%).
[c] Ethnicity: Missing, Unknown, Declined, n = 613 (3.0%).
[d] 2018 inflation adjusted USD. Missing n = 1,883 (9.3%). Pentiles were defined by categorizing the ordered distribution of median income into five categories.
[e] Difference in median and 95% CI of difference obtained via quantile regression.
[f] Population density Missing n = 1,854 (9.2%). Pentiles were defined by categorizing the ordered distribution of population density into five categories.
HM CURATOR: Houston Methodist, COVID-19 Surveillance and Outcomes Registry
SARS-CoV-2: Severe Acute Respiratory Syndrome Coronavirus 2

Table 2: Univariable comparison of socio-demographic and comorbidity factors between non-Hispanic Black (NHB) and non-Hispanic White (NHW) race categories

	Non-Hispanic Black n = 4,285	Non-Hispanic White n = 9,469	OR[a] / Median Difference (95% CI)	P value
Age: Mean (SD)	49.4 (17.8)	56.0 (19.2)	0.98 (0.98 – 0.98)	< 0.001
Age Category (%)				
Up to 35	24.3	18.8	Reference	
36 – 50	28.8	19.8	1.12 (1.01 – 1.25)	0.03
51 – 75	39.1	44.9	0.68 (0.61 – 0.74)	< 0.001
> 75	7.8	16.5	0.37 (0.32 – 0.42)	< 0.001
Females	68.1	58.1	1.54 (1.42 – 1.66)	< 0.001
Median (IQR) Zip Income	64,022 (47,303 – 79,658)	76,163 (60,130 – 102,019)	-12,141 (-14,018, -10,263)	< 0.001
Median Zip Income Pentiles (%) – Pentiles of increasing Income[b]				
Category I	33.5	13.1	Reference	
Category II	21.3	19.7	0.42 (0.38 – 0.47)	< 0.001
Category III	22.7	19.1	0.47 (0.42 – 0.52)	< 0.001
Category IV	10.5	26.2	0.16 (0.14 – 0.18)	< 0.001
Category V	11.9	22.0	0.21 (0.19 – 0.24)	< 0.001
Population Density for Zip: Median (IQR)	3217.6 (2040.7 – 4439.7)	2488.5 (812.2 – 4084.3)	729.1 (603.0 – 855.2)	< 0.001
Population Density for Zip Pentiles (%) – Pentiles of increasing population density[c]				
Category I	11.5	28.3	Reference	
Category II	21.7	20.9	2.52 (2.21 – 2.86)	< 0.001
Category III	21.8	16.9	3.20 (2.81 – 3.65)	< 0.001
Category IV	23.0	15.5	3.66 (3.21 – 4.17)	< 0.001
Category V	22.2	18.3	2.99 (2.63 – 3.41)	< 0.001
Charlson Comorbidity Index, median (IQR)	2 (0 – 6)	3 (1 – 7)	-1 (-1.22, -0.78)	< 0.001
Charlson Comorbidity Index (CCI) Categories (%)				
CCI: 0	30.8	24.9	Reference	
CCI: 1 – 2	25.1	22.3	0.91 (0.82 – 1.01)	0.07
CCI: 3 – 6	19.7	24.0	0.66 (0.60 – 0.74)	< 0.001
CCI: > 6	24.4	28.8	0.68 (0.62 – 0.75)	< 0.001
Hypertension	56.0	52.6	1.14 (1.07 – 1.23)	< 0.001
Diabetes (without complications)	29.6	21.9	1.50 (1.38 – 1.63)	< 0.001
Obesity	33.1	30.8	1.11 (1.03 – 1.20)	0.008

[a] *Unadjusted Odds Ratios and 95% Confidence Intervals for association between individual co-variates in SRAS-CoV-2 positivity or Difference in median and 95% CI of difference obtained via quantile regression.*

[b,c] *Pentiles were defined by categorizing the ordered distribution of median income and population density into five categories.*

Table 3: Univariable comparison of socio-demographic and comorbidity factors between Hispanic and non-Hispanic ethnicities

	Hispanic n = 3,590	Non-Hispanic n = 16,025	OR[a] / Median Difference[a] (95% CI)	P value
Age: Mean (SD)	45.1 (18.3)	52.8 (18.9)	0.98 (0.98 – 0.98)	< 0.001
Age Category (%)				
Up to 35	36.0	21.8	Reference	
36 – 50	26.1	24.6	0.64 (0.58 – 0.71)	< 0.001
51 – 75	32.0	40.1	0.47 (0.43 – 0.52)	< 0.001
> 75	5.9	12.7	0.28 (0.24 – 0.36)	< 0.001
Females	64.5	61.4	1.14 (1.06 – 1.23)	0.001
Median (IQR) Zip Income	65,742 (48,345 – 82,708)	73,742 (56,288 – 102,008)	-8,000 (-9,450.5, -6,549.5)	< 0.001
Median Zip Income Pentiles (%) – Pentiles of increasing Income[b]				
Category I	29.2	18.2	Reference	
Category II	23.3	19.6	0.74 (0.66 – 0.83)	< 0.001
Category III	21.8	19.8	0.68 (0.61 – 0.76)	< 0.001
Category IV	15.0	21.6	0.43 (0.38 – 0.49)	< 0.001
Category V	10.6	20.8	0.32 (0.28 – 0.36)	< 0.001
Population Density for Zip: Median (IQR)	3256.8 (1504.0 – 4299.7)	2741.6 (1408.1 – 4110.7)	515.2 (469.2 – 561.2)	< 0.001
Population Density for Zip Pentiles (%) – Pentiles of increasing population density[c]				
Category I	15.5	21.2	Reference	
Category II	21.1	20.9	1.39 (1.22 – 1.57)	< 0.001
Category III	15.3	20.0	1.05 (0.92 – 1.20)	0.48
Category IV	27.0	18.3	2.02 (1.79 – 2.28)	< 0.001
Category V	21.1	19.5	1.48 (1.31 – 1.68)	< 0.001
Charlson Comorbidity Index, median (IQR)	1 (0 – 4)	2 (0 – 6)	-1 (-1.21, -0.79)	< 0.001
Charlson Comorbidity Index (CCI) Categories (%)				
CCI: 0	41.8	29.7	Reference	
CCI: 1 – 2	25.4	23.6	0.76 (0.70 – 0.84)	< 0.001
CCI: 3 – 6	15.9	21.7	0.52 (0.47 – 0.58)	< 0.001
CCI: > 6	16.9	24.9	0.48 (0.43 – 0.54)	< 0.001
Hypertension	38.0	50.5	0.60 (0.56 – 0.65)	< 0.001
Diabetes (without complications)	27.4	23.6	1.22 (1.12 – 1.33)	< 0.001
Obesity	27.3	28.9	0.92 (0.85 – 1.00)	0.06

[a] *Unadjusted Odds Ratios and 95% Confidence Intervals for association between individual co-variates in SRAS-CoV-2 positivity or Difference in median and 95% CI of difference obtained via quantile regression.*
[b,c] *Pentiles were defined by categorizing the ordered distribution of median income and population density into five categories.*

Table 4: Adjusted Odds Ratios and 95% Confidence Intervals for likelihood of SARS-CoV-2 positivity among minority race and ethnic groups

Covariate	NHB vs. NHW, aOR (95% CI)[a]	Hispanic vs. Non-Hispanic, aOR (95% CI)[b]
Non-Hispanic Black (vs. Non-Hispanic White)	**2.23 (1.90 – 2.60)[c]**	
Hispanic (vs. Non-Hispanic)		**1.95 (1.72 – 2.20)[c]**
Age Categories		
Up to 35 years	*Reference Category*	
36–50 years	1.36 (1.07 – 1.72)	1.42 (1.20 – 1.68)
51–75 years	1.60 (1.21 – 2.11)	1.71 (1.39 – 2.11)
>75 years	2.20 (1.52 – 3.19)	2.08 (1.56 – 2.77)
Male (vs. Female)	1.20 (1.04 – 1.39)	1.17 (1.05 – 1.32)
Median Zip Household Income Categories (Pentiles of Increasing Income)		
Category I	*Reference Category*	
Category II	0.95 (0.76 – 1.18)	
Category III	1.02 (0.84 – 1.26)	
Category IV	0.74 (0.58 – 0.94)	
Category V	0.97 (0.77 – 1.22)	
Primary Insurance Type		
Medicare	*Reference Category*	
Medicaid	0.86 (0.55 – 1.33)	1.01 (0.74 – 1.37)
Private / Employer Based	1.11 (0.88 – 1.40)	0.93 (0.80 – 1.12)
Healthcare Exchange	2.06 (1.25 – 3.40)	1.80 (1.27 – 2.54)
Self-Pay	1.75 (1.33 – 2.30)	1.80 (1.27 – 2.54)
Veterans Affairs	0.72 (0.36 – 1.43)	0.58 (0.31 – 1.07)
Charlson Comorbidity Index Categories		
CCI: 0	*Reference Category*	
CCI: 1 – 2	1.26 (0.99 – 1.60)	1.03 (0.87 – 1.23)
CCI: 3 – 6	0.94 (0.68 – 1.28)	0.71 (0.56 – 0.91)
CCI: > 6	0.91 (0.65 – 1.30)	0.58 (0.44 – 0.76)
Hypertension	0.87 (0.73 – 1.05)	1.00 (0.87 – 1.15)
Diabetes (Without Complications)	1.42 (1.17 – 1.71)	1.62 (1.40 – 1.87)
Obesity	0.82 (0.70 – 0.97)	0.83 (0.73 – 0.94)

[a,b] Hosmer and Lemeshow goodness of fit p-value: 0.58 and 0.73 (H_0: Model fit is correct).

[c] ORs (95% CIs) represent the direct association adjusted for all other covariates in the model.

NHB: Non-Hispanic Black, NHW: Non-Hispanic White. Grayed out cells were not included in respective models either due to collinearity or demonstration of statistically significant mediation towards likelihood of SARS-CoV-2 positivity.

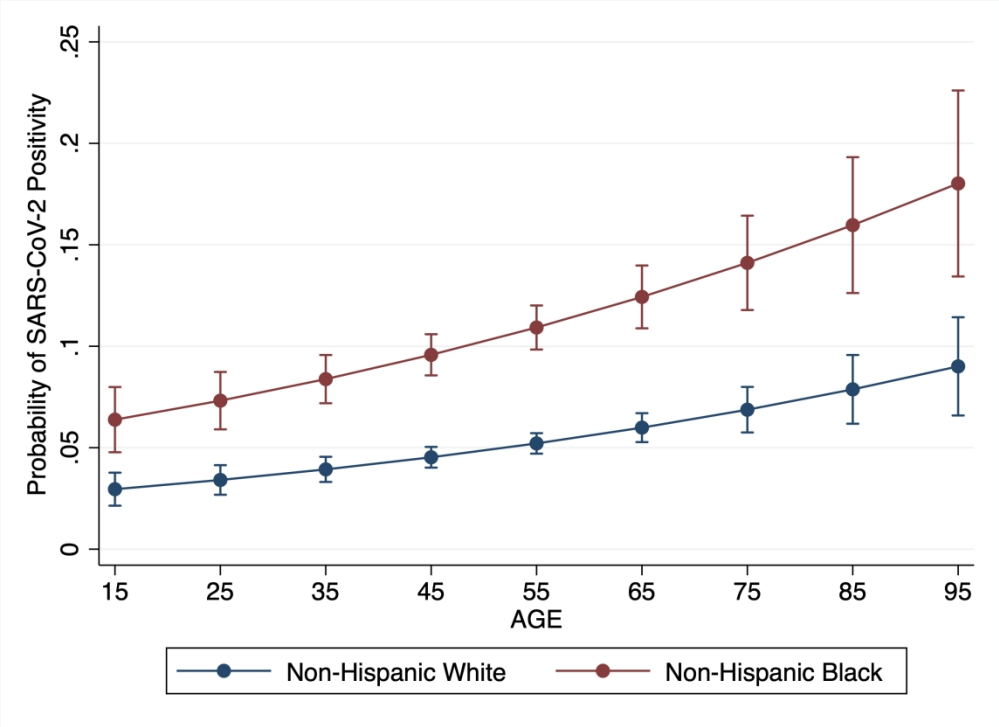


Figure 1: Adjusted Probability and 95% Confidence Interval of Positive SARS-CoV-2 PCR in Non-Hispanic Black vs. Non-Hispanic White by increasing age

467x339mm (144 x 144 DPI)

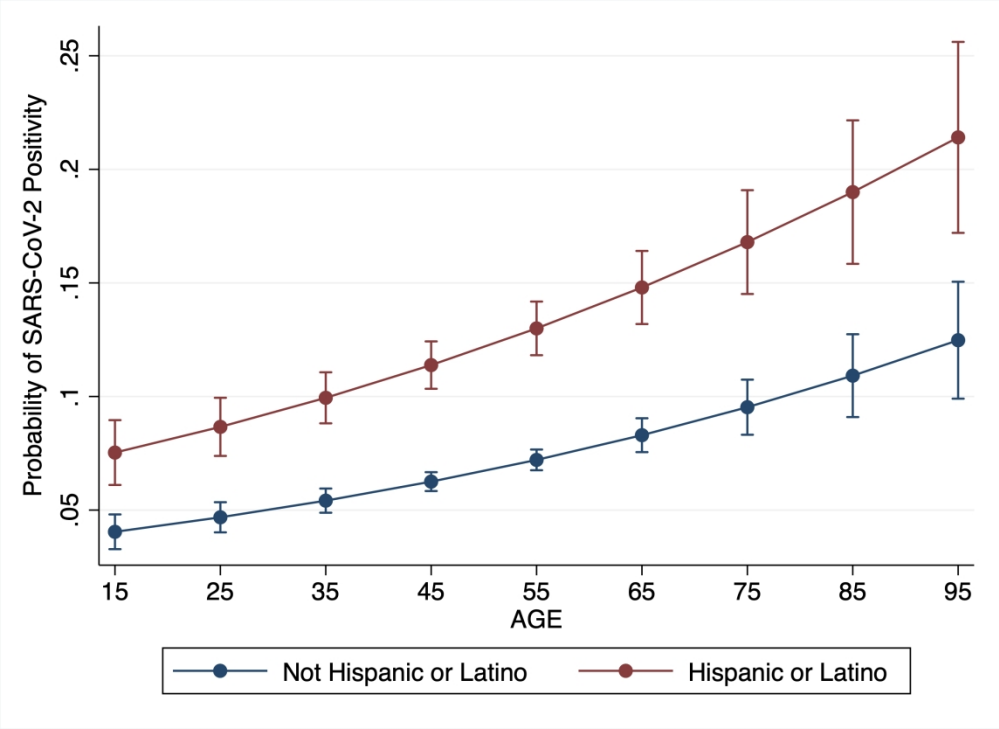


Figure 2: Adjusted Probability and 95% Confidence Interval of Positive SARS-CoV-2 PCR in Hispanic vs. Non-Hispanic by increasing age

467x339mm (144 x 144 DPI)

STROBE Statement—Checklist of items that should be included in reports of ***cross-sectional studies***

	Item No	Recommendation	Page No
Title and abstract	1	(*a*) Indicate the study's design with a commonly used term in the title or the abstract	2
		(*b*) Provide in the abstract an informative and balanced summary of what was done and what was found	2-3
Introduction			
Background/rationale	2	Explain the scientific background and rationale for the investigation being reported	4
Objectives	3	State specific objectives, including any prespecified hypotheses	5
Methods			
Study design	4	Present key elements of study design early in the paper	5
Setting	5	Describe the setting, locations, and relevant dates, including periods of recruitment, exposure, follow-up, and data collection	5-6
Participants	6	(*a*) Give the eligibility criteria, and the sources and methods of selection of participants	6
Variables	7	Clearly define all outcomes, exposures, predictors, potential confounders, and effect modifiers. Give diagnostic criteria, if applicable	6-7
Data sources/ measurement	8*	For each variable of interest, give sources of data and details of methods of assessment (measurement). Describe comparability of assessment methods if there is more than one group	6-7
Bias	9	Describe any efforts to address potential sources of bias	N/A
Study size	10	Explain how the study size was arrived at	8
Quantitative variables	11	Explain how quantitative variables were handled in the analyses. If applicable, describe which groupings were chosen and why	6-7
Statistical methods	12	(*a*) Describe all statistical methods, including those used to control for confounding	7
		(*b*) Describe any methods used to examine subgroups and interactions	7-8
		(*c*) Explain how missing data were addressed	N/A
		(*d*) If applicable, describe analytical methods taking account of sampling strategy	N/A
		(*e*) Describe any sensitivity analyses	N/A
Results			
Participants	13*	(a) Report numbers of individuals at each stage of study—eg numbers potentially eligible, examined for eligibility, confirmed eligible, included in the study, completing follow-up, and analysed	8
		(b) Give reasons for non-participation at each stage	N/A
		(c) Consider use of a flow diagram	N/A
Descriptive data	14*	(a) Give characteristics of study participants (eg demographic, clinical, social) and information on exposures and potential confounders	8, 21-22
		(b) Indicate number of participants with missing data for each variable of interest	21-22
Outcome data	15*	Report numbers of outcome events or summary measures	8
Main results	16	(*a*) Give unadjusted estimates and, if applicable, confounder-adjusted estimates and their precision (eg, 95% confidence interval). Make clear which confounders were adjusted for and why they were included	8-9, 10, 21-22

		(*b*) Report category boundaries when continuous variables were categorized	21-22
		(*c*) If relevant, consider translating estimates of relative risk into absolute risk for a meaningful time period	N/A
Other analyses	17	Report other analyses done—eg analyses of subgroups and interactions, and sensitivity analyses	9-10. 23-25
Discussion			
Key results	18	Summarise key results with reference to study objectives	11
Limitations	19	Discuss limitations of the study, taking into account sources of potential bias or imprecision. Discuss both direction and magnitude of any potential bias	13-14
Interpretation	20	Give a cautious overall interpretation of results considering objectives, limitations, multiplicity of analyses, results from similar studies, and other relevant evidence	12,14
Generalisability	21	Discuss the generalisability (external validity) of the study results	13
Other information			
Funding	22	Give the source of funding and the role of the funders for the present study and, if applicable, for the original study on which the present article is based	15

*Give information separately for exposed and unexposed groups.

Note: An Explanation and Elaboration article discusses each checklist item and gives methodological background and published examples of transparent reporting. The STROBE checklist is best used in conjunction with this article (freely available on the Web sites of PLoS Medicine at http://www.plosmedicine.org/, Annals of Internal Medicine at http://www.annals.org/, and Epidemiology at http://www.epidem.com/). Information on the STROBE Initiative is available at www.strobe-statement.org.

BMJ Open

Racial and Ethnic Disparities in SARS-CoV-2 Pandemic: Analysis of a COVID-19 Observational Registry for a Diverse U.S. Metropolitan Population

Journal:	*BMJ Open*
Manuscript ID	bmjopen-2020-039849.R2
Article Type:	Original research
Date Submitted by the Author:	14-Jul-2020
Complete List of Authors:	Vahidy, Farhaan S; Houston Methodist Research Institute, Center for Outcomes Research; Houston Methodist Neurological Institute Nicolas, Juan Carlos; Houston Methodist Research Institute Meeks, Jennifer R; Houston Methodist Research Institute Khan, Osman; Houston Methodist Research Institute Pan, Alan; Houston Methodist Research Institute, Center for Outcomes Research Jones, Stephen L.; Houston Methodist Research Institute; Weill Cornell Medicine Masud, Faisal; Houston Methodist Hospital; Weill Cornell Medicine Sostman, H Dirk; Houston Methodist Research Institute; Weill Cornell Medicine Phillips, Robert; Houston Methodist Hospital; Weill Cornell Medicine Andrieni, Julia D; Houston Methodist Hospital; Weill Cornell Medicine Kash, Bita A; Houston Methodist Research Institute; Texas A&M University School of Rural Public Health Nasir, Khurram; Houston Methodist Research Institute; Houston Methodist Hospital
Primary Subject Heading:	Epidemiology
Secondary Subject Heading:	Epidemiology, Infectious diseases, Public health
Keywords:	EPIDEMIOLOGY, Public health < INFECTIOUS DISEASES, INFECTIOUS DISEASES

I, the Submitting Author has the right to grant and does grant on behalf of all authors of the Work (as defined in the below author licence), an exclusive licence and/or a non-exclusive licence for contributions from authors who are: i) UK Crown employees; ii) where BMJ has agreed a CC-BY licence shall apply, and/or iii) in accordance with the terms applicable for US Federal Government officers or employees acting as part of their official duties; on a worldwide, perpetual, irrevocable, royalty-free basis to BMJ Publishing Group Ltd ("BMJ") its licensees and where the relevant Journal is co-owned by BMJ to the co-owners of the Journal, to publish the Work in this journal and any other BMJ products and to exploit all rights, as set out in our licence.

The Submitting Author accepts and understands that any supply made under these terms is made by BMJ to the Submitting Author unless you are acting as an employee on behalf of your employer or a postgraduate student of an affiliated institution which is paying any applicable article publishing charge ("APC") for Open Access articles. Where the Submitting Author wishes to make the Work available on an Open Access basis (and intends to pay the relevant APC), the terms of reuse of such Open Access shall be governed by a Creative Commons licence – details of these licences and which Creative Commons licence will apply to this Work are set out in our licence referred to above.

Other than as permitted in any relevant BMJ Author's Self Archiving Policies, I confirm this Work has not been accepted for publication elsewhere, is not being considered for publication elsewhere and does not duplicate material already published. I confirm all authors consent to publication of this Work and authorise the granting of this licence.

Racial and Ethnic Disparities in SARS-CoV-2 Pandemic: Analysis of a COVID-19 Observational Registry for a Diverse U.S. Metropolitan Population

Farhaan S. Vahidy PhD MBBS MPH,[1,2] Juan Carlos Nicolas BS,[1] Jennifer R. Meeks MS,[1] Osman Khan BS,[1] Alan Pan MS,[1] Stephen L. Jones MD MSHI,[1,3,4] Faisal N. Masud MD,[4,5] H. Dirk Sostman MD,[4,6] Robert A. Phillips MD PhD,[1,4,7] Julia D. Andrieni MD,[4,8] Bita Kash PhD MBA,[1,4,9] Khurram Nasir MD MPH[1,7]

1. Center for Outcomes Research, The Houston Methodist Research Institute, Houston TX
2. Houston Methodist Neurological Institute, Houston Methodist Hospital, Houston TX
3. Department of Surgery, Houston Methodist Hospital, Houston TX
4. Weill Cornell Medical College, New York NY
5. Department of Anesthesiology and Critical Care, Houston Methodist Hospital, Houston TX
6. The Houston Methodist Research Institute and The Houston Methodist Academic Institute, Houston TX
7. Department of Cardiology, DeBakey Heart and Vascular Center, Houston Methodist Hospital, Houston TX
8. Department of Medicine, Houston Methodist Hospital, Houston TX
9. Department of Health Policy and Management, Texas A&M University School of Public Health, College Station TX

Corresponding Author

Farhaan S. Vahidy, PhD MBBS MPH

Josie Roberts Building, Suite 4.123

Houston Methodist

7550 Greenbriar Drive

Houston TX 77030

fvahidy@houstonmethodist.org

Word Count: 3,063

Abstract

Introduction: Data on race and ethnic disparities for SARS-CoV-2 infection are limited. We analyzed socio-demographic factors associated with higher likelihood of SARS-CoV-2 infection and explore mediating pathways for race and ethnic disparities in the SARS-CoV-2 pandemic.

Methods: Cross-sectional analysis of COVID-19 Surveillance and Outcomes Registry (CURATOR), which captures data for a large healthcare system, comprising of one central tertiary-care hospital, seven large community hospitals, and an expansive ambulatory / emergency care network in the Greater Houston area. Nasopharyngeal samples for individuals inclusive of all ages, races, ethnicities and sex were tested for SARS-CoV-2. We analyzed socio-demographic (age, sex, race, ethnicity, household income, residence population density) and comorbidity (Charlson Comorbidity Index, hypertension, diabetes, obesity) factors. Multivariable logistic regression models were fitted to provide adjusted Odds Ratios (aOR) and 95% confidence intervals (CI) for likelihood of a positive SARS-CoV-2 test. Structural Equation Modeling (SEM) framework was utilized to explore three mediation pathways (low income, high population density, high comorbidity burden) for association between Non-Hispanic Black race (NHB), Hispanic ethnicity, and SARS-CoV-2 infection.

Results: Among 20,228 tested individuals, 1,551 (7.7%) tested positive. Overall mean (SD) age was 51.1 (19.0) years, 62% females, 22% Black and 18% were Hispanic. NHB and Hispanic ethnicity was associated with lower socio-economic status and higher population density residence. In the fully adjusted model, NHB (vs. Non-Hispanic White; aOR, CI: 2.23, 1.90-2.60) and Hispanic ethnicity (vs. non-Hispanic; aOR, CI: 1.95, 1.72-2.20) had a higher likelihood of infection. Older individuals and males were also at higher risk of infection. The SEM framework

demonstrated a significant indirect effect of NHB and Hispanic ethnicity on SARS-CoV-2 infection mediated via a pathway including residence in densely populated zip code.

Conclusions: There is strong evidence of race and ethnic disparities in the SARS-CoV-2 pandemic, that is potentially mediated through unique social determinants of health.

Strengths and limitations of this study

- One of the first studies to systematically evaluate race and ethnic disparities in susceptibility to SARS-CoV-2 infection, while accounting for multiple socio-demographic characteristics and comorbidities
- Study population represents a large and diverse metropolitan of the U.S. with data from one of the largest healthcare providers across the greater metropolitan area
- Study evaluates potential mediation pathways for race disparities and demonstrates that residence in areas with high population density may mediate race and ethnic disparities in susceptibility to SARS-CoV-2 infection
- Single center study with limited information about burden of comorbidity and lifestyle factors

INTRODUCTION

The Coronavirus disease (COVID-19), caused by infection with the Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2), is a pandemic that has thus far resulted in over 9.5 million cases globally in under 6 months. At the time of this reporting, the United States (U.S.) has approximately 25% of total global cases and has surpassed all countries in terms of absolute number of cases, cases per 1 million population, and fatalities.[1,2] Experts project these numbers to continue rising as widespread testing is instituted and newer patterns of infectivity emerge. The geographic distribution of cases across the U.S. demonstrates that the predominant pandemic burden hit major metropolitan areas. However, cases of COVID-19 have been reported across all 50 states, the District of Columbia, Guam, Puerto Rico, the Northern Mariana Islands, and the U.S. Virgin Islands.[3] As of May 31, 2020, the state of Texas had 64,287 reported cases of COVID-19, with about one-third in the Greater Houston area.[4] The Greater Houston area is home to approximately 7 million individuals, is the fourth-largest metropolitan area by population in the U.S., and is considered one of the nation's most diverse regions.[5-6]

Initial reports indicate that specific individuals such as the elderly; males; and people with comorbidities including hypertension, diabetes, obesity, coronary artery disease and heart failure have poor COVID-19 outcomes.[7–10] As the pandemic spread over the continental U.S. during the last four months, patterns of high-risk phenotypes started to emerge and reports of poor outcomes (particularly high case fatality) among racial minorities surfaced.[11–13] Though it is important to understand the determinants of poor outcomes among COVID-19 patients, it is equally imperative, from a public health perspective, to systematically examine the likelihood of SARS-CoV-2 infection across large diverse communities in the U.S. Data on higher likelihood of SARS-CoV-2 infection among racial and ethnic minorities across diverse U.S. metropolitan

areas are limited. Furthermore, the mediators of SARS-CoV-2 infection among racial and ethnic minorities have not been described.

We explored socio-demographic characteristics such as age, sex, race, ethnicity, median household income by zip codes, population density of residents' zip codes, and health insurance status associated with positive SARS-CoV-2 testing in an urban and diverse population served by one of the leading healthcare systems of the Greater Houston area. We further examined the association between pre-existing comorbidities and higher likelihood of SARS-CoV-2 infection in our study population. We hypothesized that older age, and racial and ethnic minorities will be associated with significantly higher likelihood of SARS-CoV-2 infection, and factors such as low socio-economic status, residence in high population density areas (proxy for potential difficulties in social distancing) and higher comorbidity burden will mediate the effect of race and ethnicity on SARS-CoV-2 infection.

METHODS

We analyzed data between March 5 and May 31, 2020 collected as a part of the COVID-19 Surveillance and Outcomes Registry (CURATOR) at Houston Methodist (HM). The HM CURATOR has been approved by the HM Institutional Review Board (IRB) as an observational quality of care registry for all suspected and confirmed COVID-19 patients. HM IRB granted CURATOR a waiver of informed consent and HIPAA (Health Insurance Portability and Accountability Act) authorization in accordance with current federal regulations. The CURATOR, designed and managed by the big data team at the Center for Outcomes Research (COR) at HM, is populated from multiple data sources across the HM system such as electronic medical records, electronic databanks for laboratory and pharmacy, and electronic interactive

patient interface tools. The HM system comprises a flagship tertiary care hospital in the Texas Medical Center, seven large community hospitals, a continuing care hospital, and multiple emergency centers and clinics throughout the Greater Houston area. Data from various sources are curated into a harmonized format, assessed for quality and integrity, and stored on a secure institutional HIPAA-compliant server.

We flagged all individuals who were tested for the SARS-CoV-2 using the real time Reverse Transcriptase (RT) Polymerized Chain Reaction (PCR) diagnostic panels. The three cross-validated PCR tests utilized were the World Health Organization (WHO) nucleic acid amplification test, Panther Fusion® SARS-CoV-2 Assay, and Cepheid Xpert® Xpress SARS-CoV-2 Assay. These assays were verified for quantitative detection of novel SARS-CoV-2 isolated and purified from nasopharyngeal swab specimens obtained from individuals and immersed in universal transport medium. Testing was carried out for symptomatic individuals or for individuals who had a self-reported history of exposure to a COVID-19 case including recent travel to other countries with high infection rates or hotspots within the U.S.

Socio-demographic characteristics including age, sex, race, ethnicity, and payer-status (insurance type) were obtained from the HM CURATOR for analyses. We also extracted information on presence of comorbidities comprising the Charlson Comorbidity Index (CCI) which include past history of myocardial infarction, congestive heart failure, peripheral vascular disease, cerebrovascular disease, dementia, chronic pulmonary disease, rheumatic disease, peptic ulcer disease, liver disease, diabetes with or without complications, hemiplegia, renal disease, any malignancy (excluding skin neoplasms), metastatic solid tumors, and AIDS/HIV. Data on hypertension and obesity were additionally obtained. We utilized the U.S. Census Bureau's American Community Survey (ACS) 5-year data (2014–2018) to determine median household

income by individual zip code tabulation areas (ZCTA).[14] The median ZCTA household income was inflation-adjusted to 2018 USD. We also utilized the same data source to obtain population estimates by ZCTA, and calculated ZCTA level population density (population per mile square) by standardizing it for area measurements of ZCTA. For the purpose of population density determination, land area estimates were obtained from the Census Bureau's U.S. Gazetteer Files 2010.[15] In the absence of granular and precise social distancing data, we have utilized population density as a proxy for potential difficulties in social distancing among crowded communities.

We provide descriptive summary data as means (standard deviations) and proportions. We fit univariable and multivariable logistic regression models to assess unadjusted and adjusted association between socio-demographic characteristics and likelihood of being tested positive for SARS-CoV-2. We additionally provide univariable comparison of various socio-demographic and comorbidity variables between non-Hispanic Black (NHB) and non-Hispanic White (NHW) race categories, as well as between Hispanic and Non-Hispanic ethnic groups. Age, income, population density and CCI were categorized for certain analyses. We determined *a priori* to include all variables (age, sex, race, ethnicity, zip code household income, insurance type, zip population density and CCI) in our initial multivariable model. Subsequently, factors demonstrating mediation were excluded from the final model. We assessed the model fit utilizing the Hosmer-Lemeshow goodness of fit test, and crude and adjusted odds ratios (OR and aOR) and 95% confidence intervals (CI) are reported. Post-estimation marginal probabilities of SARS-CoV-2 infection were determined from the final adjusted model for major covariates (race, ethnicity and age). We explored the mediation influence of comorbidity burden (CCI), socio-economic status (median income), and lack of social distancing (population density) on the relationship of Black race and Hispanic ethnicity with high likelihood of SARS-CoV-2 infection

using the Generalized Structural Equation Modeling (GSEM) framework. The GSEM framework was set up to provide estimates of direct and indirect effect of Black race and Hispanic ethnicity on SARS-CoV-2 infectivity. Statistically significant ($p < 0.05$) indirect effects represent full or partial mediation by a tested covariate. We included all individuals tested for SARS-CoV-2 across our healthcare system and did not perform formal sample size calculations.

Patient and Public Involvement

There was no direct patient or public involvement in the design and conduct of this study.

RESULTS

Socio-demographic and comorbidity characteristics of the study population

Across the time period of analysis, we identified a total of 20,228 presumed cases tested for SARS-CoV-2, among whom 1,551 (7.7%, CI: 7.3-8.0) tested positive. Overall, the mean (SD) age of the study population was 51.1 (19.0) years; 61.9% were female and 62.3% were White (including Hispanic ethnicity). The study sample was comparable to the overall population of patients treated across HM, who have a mean (SD) age of 49.0 (22) years, are 56% female, and 53% White. The HM system metrics was derived from a sample of 3,216,290 patients managed across the system since May 22, 2016.

The overall median (IQR) household income was USD $70,658 ($53,313–$99,276), and 42.6% of the study population had private or employer-based insurance. In our univariate analysis, Black race (vs. White; OR, CI: 1.55, 1.37–1.75), Hispanic ethnicity (vs. non-Hispanic; OR, CI: 2.02, 1.79–2.27), and males (vs. females; OR, CI: 1.17, 1.06–1.31) were associated with significantly higher likelihood of testing positive for SARS-CoV-2. Among the SARS-CoV-2 positive patients, 40.8% were in the age category of 51–75 years, and 11.4% were greater than

75 years. These proportions were significantly higher than the reference group (up to 35 years; OR, CI for 51-75 years vs. up to 35 years: 1.29, 1.12–1.48 and for >75 years vs. up to 35 years: 1.23, 1.02–1.49). Furthermore, individuals in higher pentiles of socio-economic status had significantly lower likelihood, whereas those residing in higher population density ZCTAs had higher likelihood of SARS-CoV-2 infection. We observed a significantly higher proportion of SARS-CoV-2 positive individuals in the CCI 1-2 category compared to CCI of 0 (OR, CI: 1.35, 1.18–1.54). However, similar differences for higher CCI categories were not observed. For specific comorbidities, a significantly greater proportion of diabetic individuals had SARS-CoV-2 positive results (OR, CI: 1.40, 0.17–1.68). The socio-demographic characteristics and comorbidity profiles for the overall and SARS-CoV-2 positive and negative patients are summarized in Table 1.

Socio-demographic and comorbidity characteristics associated with minority race and ethnicity

In our study sample comprising of 13,754 Non-Hispanic Black and White individuals, we compared the association between race and various socio-demographic and comorbidity characteristics (Table 2). Similarly, we also evaluated univariable differences for socio-demographic variables and co-morbidities between Hispanic and non-Hispanic individuals (Table 3). Minority race (NHB) and ethnicity (Hispanic) were both associated with younger age, higher proportion of females, and residence in low income and higher population density ZCTAs. However, NHB and Hispanic groups were both associated with an overall lower burden of comorbidities (as demonstrated by significantly lower median CCI) compared respectively to NHW and non-Hispanic categories. A higher proportion of individuals among minority race and

ethnicity were diabetic, and a higher proportion of NHB were also hypertensive compared to NHW.

Multivariable model and marginal probabilities for likelihood of SARS-CoV-2 infection and racial and ethnic minorities

The significantly higher likelihood of SARS-CoV-2 infection among minority race and ethnic groups persisted after controlling for other demographics, insurance type, median household income, population density, and comorbidities. Adjusted odds ratios (CI) for NHB vs. NHW was 2.23 (1.90 – 2.60) and for Hispanic vs. Non-Hispanic was 1.95 (1.72 – 2.20). Higher risk of infection among males (compared to females) and higher likelihood of SARS-CoV-2 infection among elderly also remained statistically significant. Detailed outputs of the fully adjusted logistic regression models for minority race and ethnic groups are presented in Table 4. Based on the marginal probabilities obtained from our fully adjusted model, the probability of SARS-CoV-2 infection in a 45-year-old NHB is 9.6% whereas it is 4.5% in a 45-year-old NHW individual, all other adjusted variables being constant. At the age of 75, this probability is 14.0% for an NHB and 6.9% for a NHW. A similar relationship differential was observed for Hispanic vs. non-Hispanic individuals. Multivariable model derived probabilities of SARS-CoV-2 infection for NHB vs. NHW and for Hispanic vs. Non-Hispanic across age spectrum are presented in Figure 1 and Figure 2.

Generalized Structural Equation Modeling for mediation by income, population density and Comorbidity Index

Utilizing the GSEM framework, we determined the direct and indirect effects of NHB and Hispanic ethnicity on SARS-CoV-2 infection with median income, population density and CCI modeled as mediators in six separate equations adjusted for age and sex. The indirect effect

of NHB mediated through population density was statistically significant (OR, CI: 1.03, 1.01 – 1.05, p = 0.001); however, the indirect effects mediated via median income and comorbidity scores were not statistically significant (p = 0.14 and p = 0.64 respectively). Among individuals identifying as Hispanic or Latino, both population density and income partially mediated the effect of ethnicity on SARS-CoV-2 positivity (OR, CI for population density: 1.02, 1.01 – 1.02, p < 0.001 and OR, CI for income: 1.04, 1.02 – 1.06, p < 0.001). Evaluation of comorbidities did not suggest a mediation influence for either NHB or Hispanic categories.

DISCUSSION

The underlying race and ethnic healthcare disparities have been painfully highlighted in the wake of the COVID-19 pandemic. Most reports indicate higher mortality or case fatality among minority racial groups (Black / African American) across major U.S. metropolitan areas.[11–13] However, robust insights on the racial differences for SARS-CoV-2 infection are limited. Furthermore, comprehensive data evaluating higher susceptibility to SARS-CoV-2 infection among Hispanic communities are also scarce. This is perhaps because of comparatively homogenous populations in non-U.S. regions of the world. Houston, as an exceptionally ethnically diverse population center,[16] is well suited for an investigation of racial, ethnic, and socioeconomic gradients in COVID-19 test positivity. We focus on highlighting the mechanisms of racial and ethnic disparities in susceptibility to SARS-CoV-2 infection and provide evidence of mediation of such disparities by novel social determinants of health (SDoH).

Our study adds to the current literature by analyzing emerging data for individuals being tested across one of the largest healthcare systems in the Greater Houston area. We report that racial and ethnic minorities (non-Hispanic Black and Hispanic individuals) are almost twice as likely to test positive for SARS-CoV-2 than the non-Hispanic White and non-Hispanic

population. These findings illuminate systematic racial / ethnic disparities in testing positive for SARS-CoV-2 infection. Though there are limited prior SARS-CoV-2 data, such racial and ethnic disparities have previously been described for the U.S. H1N1 influenza pandemic.[17] These data indicated that Spanish-speaking Hispanic and Black individuals were at a greater risk of H1N1 infection, primarily attributable to lack of healthcare access.

We explored three possible mechanisms of race disparities in our data. These included lower socio-economic status, residence in higher population dense areas, and higher level of comorbidities. We demonstrate that NHB race is significantly associated with all three potential disparity pathways, and in the traditional multivariable analyses, racial and ethnic disparities persisted even after controlling for these pathways. However, our mediation analyses highlighted the potential influence of residence in high population density areas as a viable pathway that at least partially explains the observed racial and ethnic disparity. Furthermore, residence in low income areas emerged as a significant mediation pathway for ethnic differences in SARS-CoV-2 positivity. Pathways mediating the influence of comorbidity status did not demonstrate a significant effect. We utilized population density as a marker for potential inability to maintain adequate social distancing as it has been indicated that maintaining the WHO recommended safe distance between people becomes challenging with high population densities.[18] Furthermore, overall effects of population density and disease spread has been previously described in literature.[19,20] In addition to lack of social distancing, higher population density may also be associated with several other behavioral and socio-demographic attributes that may predispose populations to both viral spread and increased susceptibility. For example, there are reports linking obesity, lack of physical activity, and higher mortality with residence in densely populated neighborhoods.[21,22]

As reported, our data also corroborate that older populations may be more susceptible to SARS-CoV-2 infection.[10] However, younger populations still have cause for concern as nearly 1 in 4 of the infected cases in our sample were between 36–50 years of age. Finally, our data demonstrate that males may be approximately 20% more likely to test positive for the SARS-CoV-2 infection. Potential sex differences in infectivity to SARS-CoV-2 and intersectionality with racial and ethnic socioeconomic factors need to be explored further in future analyses. Additional policy-oriented research should prioritize studying the intersectionality of these vulnerable economic statuses and racial disparities in COVID infection indicated by the present study.

Findings of our study need to be interpreted in the light of certain limitations. Our data are from a single center and may not be generalizable to the wider U.S. population. These findings need to be replicated in larger data sets across other large heterogenous U.S. metropolitans. However, the Houston metropolitan area is one of the most diverse and representative in the U.S.,[16] and our healthcare system is one of the largest systems providing care to COVID-19 patients in the Greater Houston area. Our sample was composed of 22% Black, 18% Hispanic, and 62% female population. We did not have information on certain demographic covariates such as education or household size. Educational status has been linked to healthcare awareness and may be important to adjust for in analyses of potential disparities, and household size may be used to provide more precise estimates of socio-economic status. However, we obtained and adjusted for zip code income data from the U.S. Census, as income has previously been shown to have strong correlation with educational attainment and socio-economic status.[23] Since testing was based on suspicion of infection and may have been influenced by factors such as access to care, the potential for selection bias cannot be ruled out.

Furthermore, lack of sensitivity of SARS-CoV-2 diagnostics tests have been reported; however, the three assays utilized for testing were cross validated for internal consistency. Finally, we did not have detailed information on comorbidities and their management in the study population. However, we did control for major comorbidities which are being reported as associated with COVID-19 outcomes.[24]

Conclusions

The strong association between racial and ethnic minorities and SARS-CoV-2 infection demonstrated in our data, even after adjustment for other important socio-demographic and comorbidity factors, highlight a potential catastrophe of inequality within the existential crisis of a global pandemic. Our data, representing a large heterogeneous U.S. metropolitan area, also provide preliminary evidence into the potential pathways for this disparity. It is highly likely that higher comorbidity burden and detrimental effects of adverse social determinants, including those that may not adequately permit safe practices of social distancing, mediate higher SARS-CoV-2 infectivity among racial and ethnic minorities.

As the pandemic continues to spread and evolve across the continental U.S., emerging data on association between SARS-CoV-2 infection and various socio-demographic factors will continue to enhance our understanding of targeted risks related to SARS-CoV-2 infection, and such data would enable us to comprehend healthcare services and access factors related to development and outcomes of COVID-19 among minority populations. Our findings substantiate prior calls for collection of robust data on race and ethnicity as a part of international collaborations,[25] and further drive home the critical importance of quantifying novel SDoH.

Acknowledgment: The authors thank Jacob M. Kolman, Senior Scientific Writer of the Houston Methodist Center for Outcomes Research, for reviewing the language and format of the manuscript.

IRB Approval: This work was carried out under an approved protocol for the Houston Methodist COVID-19 Surveillance and Outcomes Registry (HM CURATOR) by the Houston Methodist Research Institute Institutional Review Board (HMRI IRB).

Contribution Statement:

FV: design, data analysis and interpretation, drafting the manuscript, critical revision for important intellectual content, final approval

JCN: data acquisition, data analysis, drafting the manuscript, final approval

JRM: data acquisition, drafting the manuscript, final approval

OK: data acquisition, data analysis, drafting the manuscript, final approval

AP: data acquisition, data analysis, drafting the manuscript, final approval

SLJ: data acquisition, data interpretation, critical revision for important intellectual content, final approval

FNM: critical revision for important intellectual content, final approval

HDS: critical revision for important intellectual content, final approval

RAP: critical revision for important intellectual content, final approval

JDA: critical revision for important intellectual content, final approval

BK: critical revision for important intellectual content, final approval

KN: design, interpretation of data, critical revision for important intellectual content, final approval

Competing Interests: No authors declare a competing interest to this work

Funding: There are no funders to report for this submission.

Date sharing statement: All requests for de-identified data should be made to the corresponding author. All reasonable requests will be evaluated by the CURATOR Data Governance and

Sharing Committee comprising of FV, SLJ, BK and KN in the light of institutional policies and guidelines.

References

1. World Health Organization. WHO COVID-19 dashboard. https://covid19.who.int/. Published April 6, 2020. Accessed April 6, 2020.

2. Center for Systems Science and Engineering (CSSE) at Johns Hopkins University. COVID-19 map. Johns Hopkins Coronavirus Resource Center. https://coronavirus.jhu.edu/map.html. Published April 6, 2020. Accessed April 6, 2020.

3. Centers for Disease Control and Prevention. Coronavirus Disease 2019 (COVID-19) in the U.S. cdc.gov. https://www.cdc.gov/coronavirus/2019-ncov/cases-updates/cases-in-us.html. Published April 14, 2020. Accessed April 14, 2020.

4. Texas Health and Human Services. COVID-19 in Texas (dashboard). https://txdshs.maps.arcgis.com/apps/opsdashboard/index.html#/ed483ecd702b4298ab01e8b9cafc8b83. Published April 18, 2020. Accessed April 18, 2020.

5. McCann A. Most diverse cities in the U.S. Wallethub. https://wallethub.com/edu/most-diverse-cities/12690/. Published April 10, 2019. Accessed June 29, 2020.

6. Mejia B. How Houston has become the most diverse place in America. *Los Angeles Times*. https://www.latimes.com/nation/la-na-houston-diversity-2017-htmlstory.html. Published May 9, 2017. Accessed June 29, 2020.

7. Shi S, Qin M, Shen B, et al. Association of cardiac injury with mortality in hospitalized patients with COVID-19 in Wuhan, China. *JAMA Cardiol*. March 2020. doi:10.1001/jamacardio.2020.0950

8. Bonow RO, Fonarow GC, O'Gara PT, Yancy CW. Association of coronavirus disease 2019 (COVID-19) with myocardial injury and mortality. *JAMA Cardiol*. March 2020. doi:10.1001/jamacardio.2020.1105

9. 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*. April 2020. doi:10.1001/jama.2020.5394

10. Zhou F, Yu T, Du R, et al. Clinical course and risk factors for mortality of adult inpatients with COVID-19 in Wuhan, China: a retrospective cohort study. *Lancet*. 2020;395(10229):1054-1062. doi:10.1016/S0140-6736(20)30566-3

11. Deslatte M. Louisiana data: virus hits blacks, people with hypertension. *AP News*. https://apnews.com/cde897293f88f788d3f3bfdb3d46c466. Published April 7, 2020. Accessed April 20, 2020.

12. Reyes C, Husain N, Gutowski C, St. Clair S, Pratt G. Chicago's coronavirus disparity: Black Chicagoans are dying at nearly six times the rate of white residents, data show. *Chicago Tribune*. https://www.chicagotribune.com/coronavirus/ct-coronavirus-chicago-coronavirus-

deaths-demographics-lightfoot-20200406-77nlylhiavgjzb2wa4ckivh7mu-story.html. Published April 7, 2020. Accessed April 20, 2020.

13. Thebault R, Tran AB, Williams V. The coronavirus is infecting and killing black Americans at an alarmingly high rate. *Washington Post*. https://www.washingtonpost.com/nation/2020/04/07/coronavirus-is-infecting-killing-black-americans-an-alarmingly-high-rate-post-analysis-shows/. Published April 7, 2020. Accessed April 20, 2020.

14. United States Census Bureau. American Community Survey (ACS). https://www.census.gov/programs-surveys/acs. Published 2018. Accessed April 14, 2020.

15. United States Census Bureau. Gazetteer Files. https://www.census.gov/geographies/reference-files/time-series/geo/gazetteer-files.html. Accessed April 20, 2020.

16. Harris County Public Health. *Harris Cares: A 2020 Vision of Health in Harris County*. Houston, TX: Harris County Public Health; 2019. http://publichealth.harriscountytx.gov/Resources/Harris-Cares. Accessed November 21, 2019.

17. Quinn SC, Kumar S, Freimuth VS, Musa D, Casteneda-Angarita N, Kidwell K. Racial disparities in exposure, susceptibility, and access to health care in the US H1N1 influenza pandemic. *Am J Public Health*. 2011;101(2):285-293. doi:10.2105/AJPH.2009.188029

18. Rocklov J, Sjodin H. High population densities catalyze the spread of COVID-19. J Travel Med. 2020. doi: 10.1093/jtm/taaa038

19. Tarwater PM, Martin CF. Effects of population density on the spread of disease. *Complexity*. 2001;6(6):29-36. doi:10.1002/cplx.10003

20. Yang K, LeJeune J, Alsdorf D, Lu B, Shum CK, Liang S. Global distribution of outbreaks of water-associated infectious diseases. *PLoS Negl Trop Dis*. 2012;6(2):e1483. doi:10.1371/journal.pntd.0001483

21. Wang R, Feng Z, Xue D, Liu Y, Wu R. Exploring the links between population density, lifestyle, and being overweight: secondary data analyses of middle-aged and older Chinese adults. *Health Qual Life Outcomes*. 2019;17(1):100. doi:10.1186/s12955-019-1172-3

22. Beenackers MA, Oude Groeniger J, Kamphuis CBM, Van Lenthe FJ. Urban population density and mortality in a compact Dutch city: 23-year follow-up of the Dutch GLOBE study. *Health & Place*. 2018;53:79-85. doi:10.1016/j.healthplace.2018.06.010

23. U.S. Bureau of Labor Statistics. Unemployment rates and earnings by educational attainment. https://www.bls.gov/emp/chart-unemployment-earnings-education.htm. Published September 4, 2019. Accessed April 15, 2020.

24. Centers for Disease Control and Prevention. Groups at higher risk for severe illness. cdc.gov. https://www.cdc.gov/coronavirus/2019-ncov/need-extra-precautions/groups-at-higher-risk.html. Last updated April 17, 2020. Accessed April 20, 2020.

25. Abuelgasim E, Saw LJ, Shirke M, Zeinah M, Harky A. COVID-19: Unique public health issues facing Black, Asian and minority ethnic communities. *Curr Probl Cardiol*. 2020;100621. doi:10.1016/j.cpcardiol.2020.100621

Tables and Figures

Figure 1: Adjusted Probability and 95% Confidence Interval of Positive SARS-CoV-2 PCR in Non-Hispanic Black vs. Non-Hispanic White by increasing age

Figure 2: Adjusted Probability and 95% Confidence Interval of Positive SARS-CoV-2 PCR in Hispanic vs. Non-Hispanic by increasing age

Table 1: Summary measures and univariable association of socio-demographic characteristics with SARS-CoV-2 infection from HM CURATOR

Characteristics	Overall (n = 20,228)	SARS-CoV-2 Negative (n = 18,677)	SARS-CoV-2 Positive (n = 1,551)	OR (95% CI)[a]
Age, mean (SD)	51.1 (19.0)	51.0 (19.1)	52.1 (18.1)	1.00 (1.00–1.01)
Age Categories (%)				
Up to 35 years	25.1	25.5	20.9	Reference Category
36–50 years	25.0	24.9	27.0	1.32 (1.14 – 1.54)
51–75 years	38.7	38.5	40.8	1.29 (1.12 – 1.48)
>75 years	11.2	11.2	11.4	1.23 (1.02 – 1.49)
Females (%)	61.9	62.3	58.3	0.85 (0.76 – 0.94)
Race (%)[b]				
White	62.3	62.9	55.7	Reference Category
Black	21.6	21.0	28.8	1.55 (1.37 – 1.75)
Asian	9.2	9.1	9.7	1.20 (1.00 – 1.44)
Mixed / Other	1.3	1.2	1.5	1.36 (0.88 – 2.11)
Hispanic (%)[c]	17.8	16.8	29.2	2.02 (1.79 – 2.27)
Median Zip Household Income (IQR)[d]	70,658 (53,313–99,276)	70,758 (53,633–100,107)	66,523 (50,485–94,226)	-4,168[e] (-6463.2, -1872.8)
Median Zip Household Income Pentiles (%)				
I: 13,893 – 50,485	18.2	17.8	23.6	Reference Category
II: 50,642 – 65,805	18.4	18.1	21.2	0.89 (0.75 – 1.04)
III: 65,897 – 79,869	18.3	18.2	19.6	0.82 (0.70 – 0.96)
IV: 80,039 – 106,067	18.6	19.0	13.7	0.55 (0.46 – 0.65)
V: 106,415 – 240,417	17.2	17.3	16.1	0.71 (0.60 – 0.83)
Median (IQR) Population Density[f]	2797.2 (1439.1 – 4260.9)	2797.2 (1439.1 – 4211.4)	3320.3 (1904.4 – 4439.7)	523.1[e] (454.9 – 591.3)
Median Population Density Pentiles (%)				
I: 1.5 – 1026.6	18.2	18.6	13.5	Reference Category
II: 1034.6 – 2306.3	19.0	19.1	18.1	1.31 (1.09 – 1.58)
III: 2330.8 – 3328.9	17.4	17.4	18.3	1.45 (1.21 – 1.75)
IV: 3360.1 – 4665.6	18.1	17.8	22.1	1.71 (1.43 – 2.05)
V: 4742.6 – 98025.9	18.1	17.7	22.6	1.76 (1.47 – 2.10)
Insurance Status (%)				
Medicare	29.0	29.1	27.2	Reference Category
Medicaid	4.7	4.8	4.3	0.96 (0.73 – 1.26)
Pvt / Employer based	42.6	43.2	36.4	0.90 (0.79 – 1.03)
HC Exchange	1.7	1.6	3.0	2.02 (1.46 – 2.80)
Self-Pay	20.6	20.0	28.4	1.52 (1.33 – 1.75)
VA	1.3	1.4	0.7	0.54 (0.29 – 1.00)
Charlson Co-morbidity Index, Median (IQR)	2 (0 – 6)	2 (0 – 6)	2 (0 – 5)	0 (-0.36, 0.36)[f]
Charlson Co-morbidity Index (CCI) Categories (%)				
CCI: 0	33.1	33.4	30.4	Reference Category

Characteristics	Overall (n = 20,228)	SARS-CoV-2 Negative (n = 18,677)	SARS-CoV-2 Positive (n = 1,551)	OR (95% CI)[a]
CCI: 1 – 2	23.7	23.1	28.6	1.35 (1.18 – 1.54)
CCI: 3 – 6	20.3	20.2	20.4	0.10 (0.65 – 1.28)
CCI: > 6	22.9	23.1	20.6	0.98 (0.84 – 1.13)
Hypertension	47.2	47.1	48.4	1.06 (0.95 – 1.17)
Diabetes (without complications)	24.2	23.7	30.3	1.40 (1.24– 1.57)
Obesity	28.0	28.2	25.2	0.86 (0.76 – 0.96)

[a] Unadjusted Odds Ratios and 95% Confidence Intervals for association between individual co-variates in SRAS-CoV-2 positivity.
[b] Race: Missing, Unknown, Declined, n = 1157 (5.7%).
[c] Ethnicity: Missing, Unknown, Declined, n = 613 (3.0%).
[d] 2018 inflation adjusted USD. Missing n = 1,883 (9.3%). Pentiles were defined by categorizing the ordered distribution of median income into five categories.
[e] Difference in median and 95% CI of difference obtained via quantile regression.
[f] Population density Missing n = 1,854 (9.2%). Pentiles were defined by categorizing the ordered distribution of population density into five categories.
HM CURATOR: Houston Methodist, COVID-19 Surveillance and Outcomes Registry
SARS-CoV-2: Severe Acute Respiratory Syndrome Coronavirus 2

Table 2: Univariable comparison of socio-demographic and comorbidity factors between non-Hispanic Black (NHB) and non-Hispanic White (NHW) race categories

	Non-Hispanic Black n = 4,285	Non-Hispanic White n = 9,469	OR[a] / Median Difference (95% CI)	P value
Age: Mean (SD)	49.4 (17.8)	56.0 (19.2)	0.98 (0.98 – 0.98)	< 0.001
Age Category (%)				
Up to 35	24.3	18.8	Reference	
36 – 50	28.8	19.8	1.12 (1.01 – 1.25)	0.03
51 – 75	39.1	44.9	0.68 (0.61 – 0.74)	< 0.001
> 75	7.8	16.5	0.37 (0.32 – 0.42)	< 0.001
Females	68.1	58.1	1.54 (1.42 – 1.66)	< 0.001
Median (IQR) Zip Income	64,022 (47,303 – 79,658)	76,163 (60,130 – 102,019)	-12,141 (-14,018, -10,263)	< 0.001
Median Zip Income Pentiles (%) – Pentiles of increasing Income[b]				
Category I	33.5	13.1	Reference	
Category II	21.3	19.7	0.42 (0.38 – 0.47)	< 0.001
Category III	22.7	19.1	0.47 (0.42 – 0.52)	< 0.001
Category IV	10.5	26.2	0.16 (0.14 – 0.18)	< 0.001
Category V	11.9	22.0	0.21 (0.19 – 0.24)	< 0.001
Population Density for Zip: Median (IQR)	3217.6 (2040.7 – 4439.7)	2488.5 (812.2 – 4084.3)	729.1 (603.0 – 855.2)	< 0.001
Population Density for Zip Pentiles (%) – Pentiles of increasing population density[c]				
Category I	11.5	28.3	Reference	
Category II	21.7	20.9	2.52 (2.21 – 2.86)	< 0.001
Category III	21.8	16.9	3.20 (2.81 – 3.65)	< 0.001
Category IV	23.0	15.5	3.66 (3.21 – 4.17)	< 0.001
Category V	22.2	18.3	2.99 (2.63 – 3.41)	< 0.001
Charlson Comorbidity Index, median (IQR)	2 (0 – 6)	3 (1 – 7)	-1 (-1.22, -0.78)	< 0.001
Charlson Comorbidity Index (CCI) Categories (%)				
CCI: 0	30.8	24.9	Reference	
CCI: 1 – 2	25.1	22.3	0.91 (0.82 – 1.01)	0.07
CCI: 3 – 6	19.7	24.0	0.66 (0.60 – 0.74)	< 0.001
CCI: > 6	24.4	28.8	0.68 (0.62 – 0.75)	< 0.001
Hypertension	56.0	52.6	1.14 (1.07 – 1.23)	< 0.001
Diabetes (without complications)	29.6	21.9	1.50 (1.38 – 1.63)	< 0.001
Obesity	33.1	30.8	1.11 (1.03 – 1.20)	0.008

[a] *Unadjusted Odds Ratios and 95% Confidence Intervals for association between individual co-variates in SRAS-CoV-2 positivity or Difference in median and 95% CI of difference obtained via quantile regression.*

[b,c] *Pentiles were defined by categorizing the ordered distribution of median income and population density into five categories.*

Table 3: Univariable comparison of socio-demographic and comorbidity factors between Hispanic and non-Hispanic ethnicities

	Hispanic n = 3,590	Non-Hispanic n = 16,025	OR[a] / Median Difference[a] (95% CI)	P value
Age: Mean (SD)	45.1 (18.3)	52.8 (18.9)	0.98 (0.98 – 0.98)	< 0.001
Age Category (%)				
Up to 35	36.0	21.8	Reference	
36 – 50	26.1	24.6	0.64 (0.58 – 0.71)	< 0.001
51 – 75	32.0	40.1	0.47 (0.43 – 0.52)	< 0.001
> 75	5.9	12.7	0.28 (0.24 – 0.36)	< 0.001
Females	64.5	61.4	1.14 (1.06 – 1.23)	0.001
Median (IQR) Zip Income	65,742 (48,345 – 82,708)	73,742 (56,288 – 102,008)	-8,000 (-9,450.5, -6,549.5)	< 0.001
Median Zip Income Pentiles (%) – Pentiles of increasing Income[b]				
Category I	29.2	18.2	Reference	
Category II	23.3	19.6	0.74 (0.66 – 0.83)	< 0.001
Category III	21.8	19.8	0.68 (0.61 – 0.76)	< 0.001
Category IV	15.0	21.6	0.43 (0.38 – 0.49)	< 0.001
Category V	10.6	20.8	0.32 (0.28 – 0.36)	< 0.001
Population Density for Zip: Median (IQR)	3256.8 (1504.0 – 4299.7)	2741.6 (1408.1 – 4110.7)	515.2 (469.2 – 561.2)	< 0.001
Population Density for Zip Pentiles (%) – Pentiles of increasing population density[c]				
Category I	15.5	21.2	Reference	
Category II	21.1	20.9	1.39 (1.22 – 1.57)	< 0.001
Category III	15.3	20.0	1.05 (0.92 – 1.20)	0.48
Category IV	27.0	18.3	2.02 (1.79 – 2.28)	< 0.001
Category V	21.1	19.5	1.48 (1.31 – 1.68)	< 0.001
Charlson Comorbidity Index, median (IQR)	1 (0 – 4)	2 (0 – 6)	-1 (-1.21, -0.79)	< 0.001
Charlson Comorbidity Index (CCI) Categories (%)				
CCI: 0	41.8	29.7	Reference	
CCI: 1 – 2	25.4	23.6	0.76 (0.70 – 0.84)	< 0.001
CCI: 3 – 6	15.9	21.7	0.52 (0.47 – 0.58)	< 0.001
CCI: > 6	16.9	24.9	0.48 (0.43 – 0.54)	< 0.001
Hypertension	38.0	50.5	0.60 (0.56 – 0.65)	< 0.001
Diabetes (without complications)	27.4	23.6	1.22 (1.12 – 1.33)	< 0.001
Obesity	27.3	28.9	0.92 (0.85 – 1.00)	0.06

[a] *Unadjusted Odds Ratios and 95% Confidence Intervals for association between individual co-variates in SRAS-CoV-2 positivity or Difference in median and 95% CI of difference obtained via quantile regression.*
[b,c] *Pentiles were defined by categorizing the ordered distribution of median income and population density into five categories.*

Table 4: Adjusted Odds Ratios and 95% Confidence Intervals for likelihood of SARS-CoV-2 positivity among minority race and ethnic groups

Covariate	NHB vs. NHW, aOR (95% CI)[a]	Hispanic vs. Non-Hispanic, aOR (95% CI)[b]
Non-Hispanic Black (vs. Non-Hispanic White)	**2.23 (1.90 – 2.60)[c]**	
Hispanic (vs. Non-Hispanic)		**1.95 (1.72 – 2.20)[c]**
Age Categories		
Up to 35 years	*Reference Category*	
36–50 years	1.36 (1.07 – 1.72)	1.42 (1.20 – 1.68)
51–75 years	1.60 (1.21 – 2.11)	1.71 (1.39 – 2.11)
>75 years	2.20 (1.52 – 3.19)	2.08 (1.56 – 2.77)
Male (vs. Female)	1.20 (1.04 – 1.39)	1.17 (1.05 – 1.32)
Median Zip Household Income Categories (Pentiles of Increasing Income)		
Category I	*Reference Category*	
Category II	0.95 (0.76 – 1.18)	
Category III	1.02 (0.84 – 1.26)	
Category IV	0.74 (0.58 – 0.94)	
Category V	0.97 (0.77 – 1.22)	
Primary Insurance Type		
Medicare	*Reference Category*	
Medicaid	0.86 (0.55 – 1.33)	1.01 (0.74 – 1.37)
Private / Employer Based	1.11 (0.88 – 1.40)	0.93 (0.80 – 1.12)
Healthcare Exchange	2.06 (1.25 – 3.40)	1.80 (1.27 – 2.54)
Self-Pay	1.75 (1.33 – 2.30)	1.80 (1.27 – 2.54)
Veterans Affairs	0.72 (0.36 – 1.43)	0.58 (0.31 – 1.07)
Charlson Comorbidity Index Categories		
CCI: 0	*Reference Category*	
CCI: 1 – 2	1.26 (0.99 – 1.60)	1.03 (0.87 – 1.23)
CCI: 3 – 6	0.94 (0.68 – 1.28)	0.71 (0.56 – 0.91)
CCI: > 6	0.91 (0.65 – 1.30)	0.58 (0.44 – 0.76)
Hypertension	0.87 (0.73 – 1.05)	1.00 (0.87 – 1.15)
Diabetes (Without Complications)	1.42 (1.17 – 1.71)	1.62 (1.40 – 1.87)
Obesity	0.82 (0.70 – 0.97)	0.83 (0.73 – 0.94)

[a,b] Hosmer and Lemeshow goodness of fit p-value: 0.58 and 0.73 (H_0: Model fit is correct).

[c] ORs (95% CIs) represent the direct association adjusted for all other covariates in the model.

NHB: Non-Hispanic Black, NHW: Non-Hispanic White. Grayed out cells were not included in respective models either due to collinearity or demonstration of statistically significant mediation towards likelihood of SARS-CoV-2 positivity.

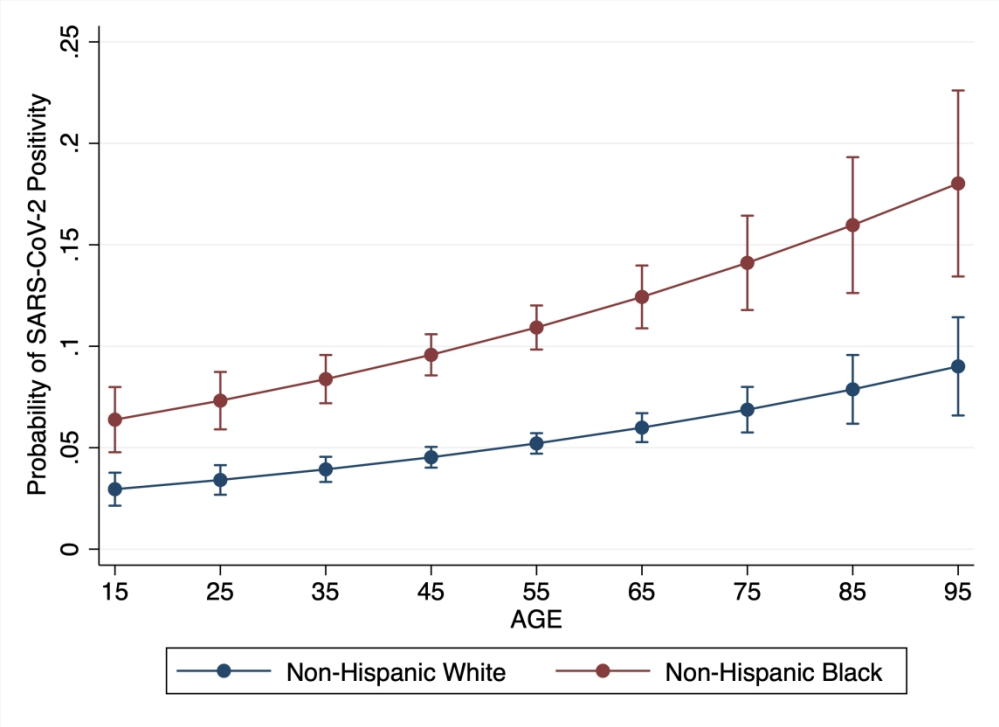


Figure 1: Adjusted Probability and 95% Confidence Interval of Positive SARS-CoV-2 PCR in Non-Hispanic Black vs. Non-Hispanic White by increasing age

467x339mm (144 x 144 DPI)

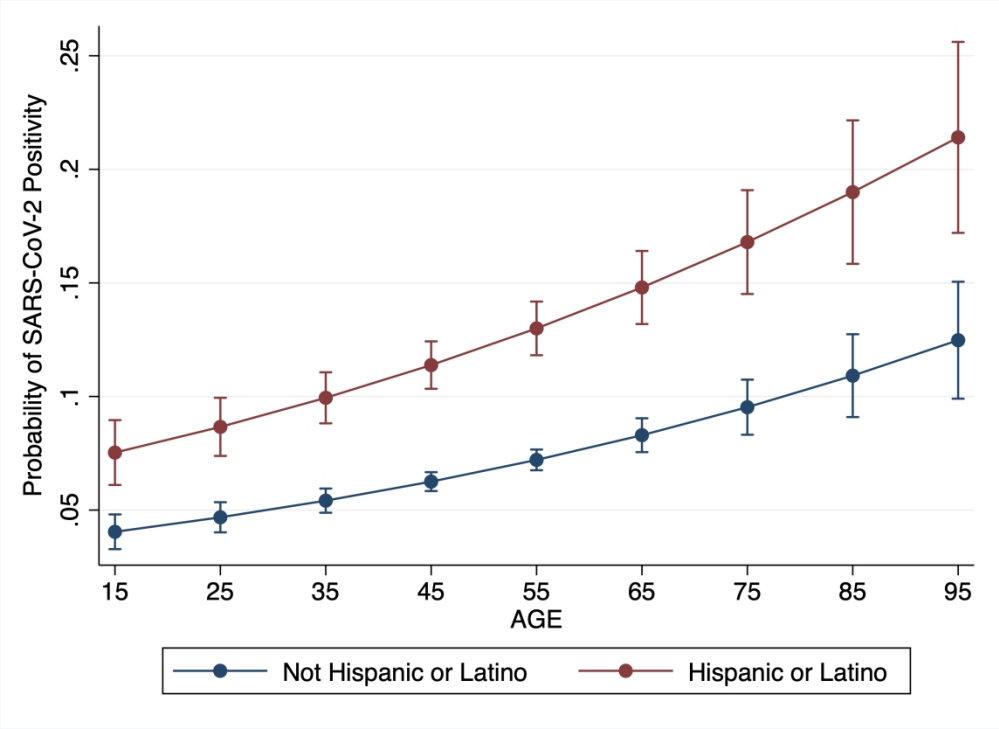


Figure 2: Adjusted Probability and 95% Confidence Interval of Positive SARS-CoV-2 PCR in Hispanic vs. Non-Hispanic by increasing age

467x339mm (144 x 144 DPI)

STROBE Statement—Checklist of items that should be included in reports of ***cross-sectional studies***

	Item No	Recommendation	Page No
Title and abstract	1	(*a*) Indicate the study's design with a commonly used term in the title or the abstract	2
		(*b*) Provide in the abstract an informative and balanced summary of what was done and what was found	2-3
Introduction			
Background/rationale	2	Explain the scientific background and rationale for the investigation being reported	4
Objectives	3	State specific objectives, including any prespecified hypotheses	5
Methods			
Study design	4	Present key elements of study design early in the paper	5
Setting	5	Describe the setting, locations, and relevant dates, including periods of recruitment, exposure, follow-up, and data collection	5-6
Participants	6	(*a*) Give the eligibility criteria, and the sources and methods of selection of participants	6
Variables	7	Clearly define all outcomes, exposures, predictors, potential confounders, and effect modifiers. Give diagnostic criteria, if applicable	6-7
Data sources/ measurement	8*	For each variable of interest, give sources of data and details of methods of assessment (measurement). Describe comparability of assessment methods if there is more than one group	6-7
Bias	9	Describe any efforts to address potential sources of bias	N/A
Study size	10	Explain how the study size was arrived at	8
Quantitative variables	11	Explain how quantitative variables were handled in the analyses. If applicable, describe which groupings were chosen and why	6-7
Statistical methods	12	(*a*) Describe all statistical methods, including those used to control for confounding	7
		(*b*) Describe any methods used to examine subgroups and interactions	7-8
		(*c*) Explain how missing data were addressed	N/A
		(*d*) If applicable, describe analytical methods taking account of sampling strategy	N/A
		(*e*) Describe any sensitivity analyses	N/A
Results			
Participants	13*	(a) Report numbers of individuals at each stage of study—eg numbers potentially eligible, examined for eligibility, confirmed eligible, included in the study, completing follow-up, and analysed	8
		(b) Give reasons for non-participation at each stage	N/A
		(c) Consider use of a flow diagram	N/A
Descriptive data	14*	(a) Give characteristics of study participants (eg demographic, clinical, social) and information on exposures and potential confounders	8, 21-22
		(b) Indicate number of participants with missing data for each variable of interest	21-22
Outcome data	15*	Report numbers of outcome events or summary measures	8
Main results	16	(*a*) Give unadjusted estimates and, if applicable, confounder-adjusted estimates and their precision (eg, 95% confidence interval). Make clear which confounders were adjusted for and why they were included	8-9, 10, 21-22

		(*b*) Report category boundaries when continuous variables were categorized	21-22
		(*c*) If relevant, consider translating estimates of relative risk into absolute risk for a meaningful time period	N/A
Other analyses	17	Report other analyses done—eg analyses of subgroups and interactions, and sensitivity analyses	9-10. 23-25
Discussion			
Key results	18	Summarise key results with reference to study objectives	11
Limitations	19	Discuss limitations of the study, taking into account sources of potential bias or imprecision. Discuss both direction and magnitude of any potential bias	13-14
Interpretation	20	Give a cautious overall interpretation of results considering objectives, limitations, multiplicity of analyses, results from similar studies, and other relevant evidence	12,14
Generalisability	21	Discuss the generalisability (external validity) of the study results	13
Other information			
Funding	22	Give the source of funding and the role of the funders for the present study and, if applicable, for the original study on which the present article is based	15

*Give information separately for exposed and unexposed groups.

Note: An Explanation and Elaboration article discusses each checklist item and gives methodological background and published examples of transparent reporting. The STROBE checklist is best used in conjunction with this article (freely available on the Web sites of PLoS Medicine at http://www.plosmedicine.org/, Annals of Internal Medicine at http://www.annals.org/, and Epidemiology at http://www.epidem.com/). Information on the STROBE Initiative is available at www.strobe-statement.org.

BMJ Open

Racial and Ethnic Disparities in SARS-CoV-2 Pandemic: Analysis of a COVID-19 Observational Registry for a Diverse U.S. Metropolitan Population

Journal:	*BMJ Open*
Manuscript ID	bmjopen-2020-039849.R3
Article Type:	Original research
Date Submitted by the Author:	22-Jul-2020
Complete List of Authors:	Vahidy, Farhaan S; Houston Methodist Research Institute, Center for Outcomes Research; Houston Methodist Neurological Institute Nicolas, Juan Carlos; Houston Methodist Research Institute Meeks, Jennifer R; Houston Methodist Research Institute Khan, Osman; Houston Methodist Research Institute Pan, Alan; Houston Methodist Research Institute, Center for Outcomes Research Jones, Stephen L.; Houston Methodist Research Institute; Weill Cornell Medicine Masud, Faisal; Houston Methodist Hospital; Weill Cornell Medicine Sostman, H Dirk; Houston Methodist Research Institute; Weill Cornell Medicine Phillips, Robert; Houston Methodist Hospital; Weill Cornell Medicine Andrieni, Julia D; Houston Methodist Hospital; Weill Cornell Medicine Kash, Bita A; Houston Methodist Research Institute; Texas A&M University School of Rural Public Health Nasir, Khurram; Houston Methodist Research Institute; Houston Methodist Hospital
Primary Subject Heading:	Epidemiology
Secondary Subject Heading:	Epidemiology, Infectious diseases, Public health
Keywords:	EPIDEMIOLOGY, Public health < INFECTIOUS DISEASES, INFECTIOUS DISEASES

I, the Submitting Author has the right to grant and does grant on behalf of all authors of the Work (as defined in the below author licence), an exclusive licence and/or a non-exclusive licence for contributions from authors who are: i) UK Crown employees; ii) where BMJ has agreed a CC-BY licence shall apply, and/or iii) in accordance with the terms applicable for US Federal Government officers or employees acting as part of their official duties; on a worldwide, perpetual, irrevocable, royalty-free basis to BMJ Publishing Group Ltd ("BMJ") its licensees and where the relevant Journal is co-owned by BMJ to the co-owners of the Journal, to publish the Work in this journal and any other BMJ products and to exploit all rights, as set out in our licence.

The Submitting Author accepts and understands that any supply made under these terms is made by BMJ to the Submitting Author unless you are acting as an employee on behalf of your employer or a postgraduate student of an affiliated institution which is paying any applicable article publishing charge ("APC") for Open Access articles. Where the Submitting Author wishes to make the Work available on an Open Access basis (and intends to pay the relevant APC), the terms of reuse of such Open Access shall be governed by a Creative Commons licence – details of these licences and which Creative Commons licence will apply to this Work are set out in our licence referred to above.

Other than as permitted in any relevant BMJ Author's Self Archiving Policies, I confirm this Work has not been accepted for publication elsewhere, is not being considered for publication elsewhere and does not duplicate material already published. I confirm all authors consent to publication of this Work and authorise the granting of this licence.

Racial and Ethnic Disparities in SARS-CoV-2 Pandemic: Analysis of a COVID-19 Observational Registry for a Diverse U.S. Metropolitan Population

Farhaan S. Vahidy PhD MBBS MPH,[1,2] Juan Carlos Nicolas BS,[1] Jennifer R. Meeks MS,[1] Osman Khan BS,[1] Alan Pan MS,[1] Stephen L. Jones MD MSHI,[1,3,4] Faisal N. Masud MD,[4,5] H. Dirk Sostman MD,[4,6] Robert A. Phillips MD PhD,[1,4,7] Julia D. Andrieni MD,[4,8] Bita Kash PhD MBA,[1,4,9] Khurram Nasir MD MPH[1,7]

1. Center for Outcomes Research, The Houston Methodist Research Institute, Houston TX
2. Houston Methodist Neurological Institute, Houston Methodist Hospital, Houston TX
3. Department of Surgery, Houston Methodist Hospital, Houston TX
4. Weill Cornell Medical College, New York NY
5. Department of Anesthesiology and Critical Care, Houston Methodist Hospital, Houston TX
6. The Houston Methodist Research Institute and The Houston Methodist Academic Institute, Houston TX
7. Department of Cardiology, DeBakey Heart and Vascular Center, Houston Methodist Hospital, Houston TX
8. Department of Medicine, Houston Methodist Hospital, Houston TX
9. Department of Health Policy and Management, Texas A&M University School of Public Health, College Station TX

Corresponding Author

Farhaan S. Vahidy, PhD MBBS MPH
Josie Roberts Building, Suite 4.123
Houston Methodist
7550 Greenbriar Drive
Houston TX 77030
fvahidy@houstonmethodist.org

Word Count: 3,093

Abstract

Introduction: Data on race and ethnic disparities for SARS-CoV-2 infection are limited. We analyzed socio-demographic factors associated with higher likelihood of SARS-CoV-2 infection and explore mediating pathways for race and ethnic disparities in the SARS-CoV-2 pandemic.

Methods: Cross-sectional analysis of COVID-19 Surveillance and Outcomes Registry (CURATOR), which captures data for a large healthcare system, comprising of one central tertiary-care hospital, seven large community hospitals, and an expansive ambulatory / emergency care network in the Greater Houston area. Nasopharyngeal samples for individuals inclusive of all ages, races, ethnicities and sex were tested for SARS-CoV-2. We analyzed socio-demographic (age, sex, race, ethnicity, household income, residence population density) and comorbidity (Charlson Comorbidity Index, hypertension, diabetes, obesity) factors. Multivariable logistic regression models were fitted to provide adjusted Odds Ratios (aOR) and 95% confidence intervals (CI) for likelihood of a positive SARS-CoV-2 test. Structural Equation Modeling (SEM) framework was utilized to explore three mediation pathways (low income, high population density, high comorbidity burden) for association between Non-Hispanic Black race (NHB), Hispanic ethnicity, and SARS-CoV-2 infection.

Results: Among 20,228 tested individuals, 1,551 (7.7%) tested positive. Overall mean (SD) age was 51.1 (19.0) years, 62% females, 22% Black and 18% were Hispanic. NHB and Hispanic ethnicity was associated with lower socio-economic status and higher population density residence. In the fully adjusted model, NHB (vs. Non-Hispanic White; aOR, CI: 2.23, 1.90-2.60) and Hispanic ethnicity (vs. non-Hispanic; aOR, CI: 1.95, 1.72-2.20) had a higher likelihood of infection. Older individuals and males were also at higher risk of infection. The SEM framework

demonstrated a significant indirect effect of NHB and Hispanic ethnicity on SARS-CoV-2 infection mediated via a pathway including residence in densely populated zip code.

Conclusions: There is strong evidence of race and ethnic disparities in the SARS-CoV-2 pandemic, that is potentially mediated through unique social determinants of health.

Strengths and limitations of this study

- One of the first studies to systematically evaluate race and ethnic disparities in susceptibility to SARS-CoV-2 infection, while accounting for multiple socio-demographic characteristics and comorbidities
- Study population represents a large and diverse metropolitan of the U.S. with data from one of the largest healthcare providers across the greater metropolitan area
- Study evaluates potential mediation pathways for race disparities and demonstrates that residence in areas with high population density may mediate race and ethnic disparities in susceptibility to SARS-CoV-2 infection
- Single center study with limited information about burden of comorbidity and lifestyle factors

INTRODUCTION

The Coronavirus disease (COVID-19), caused by infection with the Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2), is a pandemic that has thus far resulted in over 9.5 million cases globally in under 6 months. At the time of this reporting, the United States (U.S.) has approximately 25% of total global cases and has surpassed all countries in terms of absolute number of cases, cases per 1 million population, and fatalities.[1,2] Experts project these numbers to continue rising as widespread testing is instituted and newer patterns of infectivity emerge. The geographic distribution of cases across the U.S. demonstrates that the predominant pandemic burden hit major metropolitan areas. However, cases of COVID-19 have been reported across all 50 states, the District of Columbia, Guam, Puerto Rico, the Northern Mariana Islands, and the U.S. Virgin Islands.[3] As of May 31, 2020, the state of Texas had 64,287 reported cases of COVID-19, with about one-third in the Greater Houston area.[4] The Greater Houston area is home to approximately 7 million individuals, is the fourth-largest metropolitan area by population in the U.S., and is considered one of the nation's most diverse regions.[5-6]

Initial reports indicate that specific individuals such as the elderly; males; and people with comorbidities including hypertension, diabetes, obesity, coronary artery disease and heart failure have poor COVID-19 outcomes.[7–10] As the pandemic spread over the continental U.S. during the last four months, patterns of high-risk phenotypes started to emerge and reports of poor outcomes (particularly high case fatality) among racial minorities surfaced.[11–13] Though it is important to understand the determinants of poor outcomes among COVID-19 patients, it is equally imperative, from a public health perspective, to systematically examine the likelihood of SARS-CoV-2 infection across large diverse communities in the U.S. Data on higher likelihood of SARS-CoV-2 infection among racial and ethnic minorities across diverse U.S. metropolitan

areas are limited. Furthermore, the mediators of SARS-CoV-2 infection among racial and ethnic minorities have not been described.

We explored socio-demographic characteristics such as age, sex, race, ethnicity, median household income by zip codes, population density of residents' zip codes, and health insurance status associated with positive SARS-CoV-2 testing in an urban and diverse population served by one of the leading healthcare systems of the Greater Houston area. We further examined the association between pre-existing comorbidities and higher likelihood of SARS-CoV-2 infection in our study population. We hypothesized that older age, and racial and ethnic minorities will be associated with significantly higher likelihood of SARS-CoV-2 infection, and factors such as low socio-economic status, residence in high population density areas (proxy for potential difficulties in social distancing) and higher comorbidity burden will mediate the effect of race and ethnicity on SARS-CoV-2 infection.

METHODS

We analyzed data between March 5 and May 31, 2020 collected as a part of the COVID-19 Surveillance and Outcomes Registry (CURATOR) at Houston Methodist (HM). The HM CURATOR has been approved by the HM Institutional Review Board (IRB) as an observational quality of care registry for all suspected and confirmed COVID-19 patients. HM IRB granted CURATOR a waiver of informed consent and HIPAA (Health Insurance Portability and Accountability Act) authorization in accordance with current federal regulations. The CURATOR, designed and managed by the big data team at the Center for Outcomes Research (COR) at HM, is populated from multiple data sources across the HM system such as electronic medical records, electronic databanks for laboratory and pharmacy, and electronic interactive

patient interface tools. The HM system comprises a flagship tertiary care hospital in the Texas Medical Center, seven large community hospitals, a continuing care hospital, and multiple emergency centers and clinics throughout the Greater Houston area. Data from various sources are curated into a harmonized format, assessed for quality and integrity, and stored on a secure institutional HIPAA-compliant server.

We flagged all individuals who were tested for the SARS-CoV-2 using the real time Reverse Transcriptase (RT) Polymerized Chain Reaction (PCR) diagnostic panels. The three cross-validated PCR tests utilized were the World Health Organization (WHO) nucleic acid amplification test, Panther Fusion® SARS-CoV-2 Assay, and Cepheid Xpert® Xpress SARS-CoV-2 Assay. These assays were verified for quantitative detection of novel SARS-CoV-2 isolated and purified from nasopharyngeal swab specimens obtained from individuals and immersed in universal transport medium. Testing was carried out for symptomatic individuals or for individuals who had a self-reported history of exposure to a COVID-19 case including recent travel to other countries with high infection rates or hotspots within the U.S.

Socio-demographic characteristics including age, sex, race, ethnicity, and payer-status (insurance type) were obtained from the HM CURATOR for analyses. We also extracted information on presence of comorbidities comprising the Charlson Comorbidity Index (CCI) which include past history of myocardial infarction, congestive heart failure, peripheral vascular disease, cerebrovascular disease, dementia, chronic pulmonary disease, rheumatic disease, peptic ulcer disease, liver disease, diabetes with or without complications, hemiplegia, renal disease, any malignancy (excluding skin neoplasms), metastatic solid tumors, and AIDS/HIV. Data on hypertension and obesity were additionally obtained. We utilized the U.S. Census Bureau's American Community Survey (ACS) 5-year data (2014–2018) to determine median household

income by individual zip code tabulation areas (ZCTA).[14] The median ZCTA household income was inflation-adjusted to 2018 USD. We also utilized the same data source to obtain population estimates by ZCTA, and calculated ZCTA level population density (population per mile square) by standardizing it for area measurements of ZCTA. For the purpose of population density determination, land area estimates were obtained from the Census Bureau's U.S. Gazetteer Files 2010.[15] In the absence of granular and precise social distancing data, we have utilized population density as a proxy for potential difficulties in social distancing among crowded communities.

We provide descriptive summary data as means (standard deviations) and proportions. We fit univariable and multivariable logistic regression models to assess unadjusted and adjusted association between socio-demographic characteristics and likelihood of being tested positive for SARS-CoV-2. We additionally provide univariable comparison of various socio-demographic and comorbidity variables between non-Hispanic Black (NHB) and non-Hispanic White (NHW) race categories, as well as between Hispanic and Non-Hispanic ethnic groups. Age, income, population density and CCI were categorized for certain analyses. We determined *a priori* to include all variables (age, sex, race, ethnicity, zip code household income, insurance type, zip population density and CCI) in our initial multivariable model. Subsequently, factors demonstrating mediation were excluded from the final models. However, the factors that did not demonstrate mediation were included in the final models, as we believe that they continue to importantly inform the variance of estimates for direct effects.[16] We assessed the model fit utilizing the Hosmer-Lemeshow goodness of fit test, and crude and adjusted odds ratios (OR and aOR) and 95% confidence intervals (CI) are reported. Post-estimation marginal probabilities of SARS-CoV-2 infection were determined from the final adjusted model for major covariates (race, ethnicity and age). We explored the mediation influence of comorbidity burden (CCI),

socio-economic status (median income), and lack of social distancing (population density) on the relationship of Black race and Hispanic ethnicity with high likelihood of SARS-CoV-2 infection using the Generalized Structural Equation Modeling (GSEM) framework. The GSEM framework was set up to provide estimates of direct and indirect effect of Black race and Hispanic ethnicity on SARS-CoV-2 infectivity. Statistically significant ($p < 0.05$) indirect effects represent full or partial mediation by a tested covariate. We included all individuals tested for SARS-CoV-2 across our healthcare system and did not perform formal sample size calculations.

Patient and Public Involvement

There was no direct patient or public involvement in the design and conduct of this study.

RESULTS

Socio-demographic and comorbidity characteristics of the study population

Across the time period of analysis, we identified a total of 20,228 presumed cases tested for SARS-CoV-2, among whom 1,551 (7.7%, CI: 7.3-8.0) tested positive. Overall, the mean (SD) age of the study population was 51.1 (19.0) years; 61.9% were female and 62.3% were White (including Hispanic ethnicity). The study sample was comparable to the overall population of patients treated across HM, who have a mean (SD) age of 49.0 (22) years, are 56% female, and 53% White. The HM system metrics was derived from a sample of 3,216,290 patients managed across the system since May 22, 2016.

The overall median (IQR) household income was USD $70,658 ($53,313–$99,276), and 42.6% of the study population had private or employer-based insurance. In our univariate analysis, Black race (vs. White; OR, CI: 1.55, 1.37–1.75), Hispanic ethnicity (vs. non-Hispanic; OR, CI: 2.02, 1.79–2.27), and males (vs. females; OR, CI: 1.17, 1.06–1.31) were associated with

significantly higher likelihood of testing positive for SARS-CoV-2. Among the SARS-CoV-2 positive patients, 40.8% were in the age category of 51–75 years, and 11.4% were greater than 75 years. These proportions were significantly higher than the reference group (up to 35 years; OR, CI for 51-75 years vs. up to 35 years: 1.29, 1.12–1.48 and for >75 years vs. up to 35 years: 1.23, 1.02–1.49). Furthermore, individuals in higher pentiles of socio-economic status had significantly lower likelihood, whereas those residing in higher population density ZCTAs had higher likelihood of SARS-CoV-2 infection. We observed a significantly higher proportion of SARS-CoV-2 positive individuals in the CCI 1-2 category compared to CCI of 0 (OR, CI: 1.35, 1.18–1.54). However, similar differences for higher CCI categories were not observed. For specific comorbidities, a significantly greater proportion of diabetic individuals had SARS-CoV-2 positive results (OR, CI: 1.40, 0.17–1.68). The socio-demographic characteristics and comorbidity profiles for the overall and SARS-CoV-2 positive and negative patients are summarized in Table 1.

Socio-demographic and comorbidity characteristics associated with minority race and ethnicity

In our study sample comprising of 13,754 Non-Hispanic Black and White individuals, we compared the association between race and various socio-demographic and comorbidity characteristics (Table 2). Similarly, we also evaluated univariable differences for socio-demographic variables and co-morbidities between Hispanic and non-Hispanic individuals (Table 3). Minority race (NHB) and ethnicity (Hispanic) were both associated with younger age, higher proportion of females, and residence in low income and higher population density ZCTAs. However, NHB and Hispanic groups were both associated with an overall lower burden of comorbidities (as demonstrated by significantly lower median CCI) compared respectively to

NHW and non-Hispanic categories. A higher proportion of individuals among minority race and ethnicity were diabetic, and a higher proportion of NHB were also hypertensive compared to NHW.

Multivariable model and marginal probabilities for likelihood of SARS-CoV-2 infection and racial and ethnic minorities

The significantly higher likelihood of SARS-CoV-2 infection among minority race and ethnic groups persisted after controlling for other demographics, insurance type, median household income, population density, and comorbidities. Adjusted odds ratios (CI) for NHB vs. NHW was 2.23 (1.90 – 2.60) and for Hispanic vs. Non-Hispanic was 1.95 (1.72 – 2.20). Higher risk of infection among males (compared to females) and higher likelihood of SARS-CoV-2 infection among elderly also remained statistically significant. Detailed outputs of the fully adjusted logistic regression models for minority race and ethnic groups are presented in Table 4. Based on the marginal probabilities obtained from our fully adjusted model, the probability of SARS-CoV-2 infection in a 45-year-old NHB is 9.6% whereas it is 4.5% in a 45-year-old NHW individual, all other adjusted variables being constant. At the age of 75, this probability is 14.0% for an NHB and 6.9% for a NHW. A similar relationship differential was observed for Hispanic vs. non-Hispanic individuals. Multivariable model derived probabilities of SARS-CoV-2 infection for NHB vs. NHW and for Hispanic vs. Non-Hispanic across age spectrum are presented in Figure 1 and Figure 2.

Generalized Structural Equation Modeling for mediation by income, population density and Comorbidity Index

Utilizing the GSEM framework, we determined the direct and indirect effects of NHB and Hispanic ethnicity on SARS-CoV-2 infection with median income, population density and

CCI modeled as mediators in six separate equations adjusted for age and sex. The indirect effect of NHB mediated through population density was statistically significant (OR, CI: 1.03, 1.01 – 1.05, p = 0.001); however, the indirect effects mediated via median income and comorbidity scores were not statistically significant (p = 0.14 and p = 0.64 respectively). Among individuals identifying as Hispanic or Latino, both population density and income partially mediated the effect of ethnicity on SARS-CoV-2 positivity (OR, CI for population density: 1.02, 1.01 – 1.02, p < 0.001 and OR, CI for income: 1.04, 1.02 – 1.06, p < 0.001). Evaluation of comorbidities did not suggest a mediation influence for either NHB or Hispanic categories.

DISCUSSION

The underlying race and ethnic healthcare disparities have been painfully highlighted in the wake of the COVID-19 pandemic. Most reports indicate higher mortality or case fatality among minority racial groups (Black / African American) across major U.S. metropolitan areas.[11–13] However, robust insights on the racial differences for SARS-CoV-2 infection are limited. Furthermore, comprehensive data evaluating higher susceptibility to SARS-CoV-2 infection among Hispanic communities are also scarce. This is perhaps because of comparatively homogenous populations in non-U.S. regions of the world. Houston, as an exceptionally ethnically diverse population center,[17] is well suited for an investigation of racial, ethnic, and socioeconomic gradients in COVID-19 test positivity. We focus on highlighting the mechanisms of racial and ethnic disparities in susceptibility to SARS-CoV-2 infection and provide evidence of mediation of such disparities by novel social determinants of health (SDoH).

Our study adds to the current literature by analyzing emerging data for individuals being tested across one of the largest healthcare systems in the Greater Houston area. We report that racial and ethnic minorities (non-Hispanic Black and Hispanic individuals) are almost twice as

likely to test positive for SARS-CoV-2 than the non-Hispanic White and non-Hispanic population. These findings illuminate systematic racial / ethnic disparities in testing positive for SARS-CoV-2 infection. Though there are limited prior SARS-CoV-2 data, such racial and ethnic disparities have previously been described for the U.S. H1N1 influenza pandemic.[18] These data indicated that Spanish-speaking Hispanic and Black individuals were at a greater risk of H1N1 infection, primarily attributable to lack of healthcare access.

We explored three possible mechanisms of race disparities in our data. These included lower socio-economic status, residence in higher population dense areas, and higher level of comorbidities. We demonstrate that NHB race is significantly associated with all three potential disparity pathways, and in the traditional multivariable analyses, racial and ethnic disparities persisted even after controlling for these pathways. However, our mediation analyses highlighted the potential influence of residence in high population density areas as a viable pathway that at least partially explains the observed racial and ethnic disparity. Furthermore, residence in low income areas emerged as a significant mediation pathway for ethnic differences in SARS-CoV-2 positivity. Pathways mediating the influence of comorbidity status did not demonstrate a significant effect. We utilized population density as a marker for potential inability to maintain adequate social distancing as it has been indicated that maintaining the WHO recommended safe distance between people becomes challenging with high population densities.[19] Furthermore, overall effects of population density and disease spread has been previously described in literature.[20,21] In addition to lack of social distancing, higher population density may also be associated with several other behavioral and socio-demographic attributes that may predispose populations to both viral spread and increased susceptibility. For example, there are reports

linking obesity, lack of physical activity, and higher mortality with residence in densely populated neighborhoods.[22,23]

As reported, our data also corroborate that older populations may be more susceptible to SARS-CoV-2 infection.[10] However, younger populations still have cause for concern as nearly 1 in 4 of the infected cases in our sample were between 36–50 years of age. Finally, our data demonstrate that males may be approximately 20% more likely to test positive for the SARS-CoV-2 infection. Potential sex differences in infectivity to SARS-CoV-2 and intersectionality with racial and ethnic socioeconomic factors need to be explored further in future analyses. Additional policy-oriented research should prioritize studying the intersectionality of these vulnerable economic statuses and racial disparities in COVID infection indicated by the present study.

Findings of our study need to be interpreted in the light of certain limitations. Our data are from a single center and may not be generalizable to the wider U.S. population. These findings need to be replicated in larger data sets across other large heterogenous U.S. metropolitans. However, the Houston metropolitan area is one of the most diverse and representative in the U.S.[17] and our healthcare system is one of the largest systems providing care to COVID-19 patients in the Greater Houston area. Our sample was composed of 22% Black, 18% Hispanic, and 62% female population. We did not have information on certain demographic covariates such as education or household size. Educational status has been linked to healthcare awareness and may be important to adjust for in analyses of potential disparities, and household size may be used to provide more precise estimates of socio-economic status. However, we obtained and adjusted for zip code income data from the U.S. Census, as income has previously been shown to have strong correlation with educational attainment and socio-

economic status.[24] Since testing was based on suspicion of infection and may have been influenced by factors such as access to care, the potential for selection bias cannot be ruled out. Furthermore, lack of sensitivity of SARS-CoV-2 diagnostics tests have been reported; however, the three assays utilized for testing were cross validated for internal consistency. Finally, we did not have detailed information on comorbidities and their management in the study population. However, we did control for major comorbidities which are being reported as associated with COVID-19 outcomes.[25]

Conclusions

The strong association between racial and ethnic minorities and SARS-CoV-2 infection demonstrated in our data, even after adjustment for other important socio-demographic and comorbidity factors, highlight a potential catastrophe of inequality within the existential crisis of a global pandemic. Our data, representing a large heterogeneous U.S. metropolitan area, also provide preliminary evidence into the potential pathways for this disparity. It is highly likely that higher comorbidity burden and detrimental effects of adverse social determinants, including those that may not adequately permit safe practices of social distancing, mediate higher SARS-CoV-2 infectivity among racial and ethnic minorities.

As the pandemic continues to spread and evolve across the continental U.S., emerging data on association between SARS-CoV-2 infection and various socio-demographic factors will continue to enhance our understanding of targeted risks related to SARS-CoV-2 infection, and such data would enable us to comprehend healthcare services and access factors related to development and outcomes of COVID-19 among minority populations. Our findings substantiate

prior calls for collection of robust data on race and ethnicity as a part of international collaborations,[26] and further drive home the critical importance of quantifying novel SDoH.

Acknowledgment: The authors thank Jacob M. Kolman, Senior Scientific Writer of the Houston Methodist Center for Outcomes Research, for reviewing the language and format of the manuscript.

IRB Approval: This work was carried out under an approved protocol for the Houston Methodist COVID-19 Surveillance and Outcomes Registry (HM CURATOR) by the Houston Methodist Research Institute Institutional Review Board (HMRI IRB).

Contribution Statement:

FV: design, data analysis and interpretation, drafting the manuscript, critical revision for important intellectual content, final approval

JCN: data acquisition, data analysis, drafting the manuscript, final approval

JRM: data acquisition, drafting the manuscript, final approval

OK: data acquisition, data analysis, drafting the manuscript, final approval

AP: data acquisition, data analysis, drafting the manuscript, final approval

SLJ: data acquisition, data interpretation, critical revision for important intellectual content, final approval

FNM: critical revision for important intellectual content, final approval

HDS: critical revision for important intellectual content, final approval

RAP: critical revision for important intellectual content, final approval

JDA: critical revision for important intellectual content, final approval

BK: critical revision for important intellectual content, final approval

KN: design, interpretation of data, critical revision for important intellectual content, final approval

Competing Interests: No authors declare a competing interest to this work

Funding: There are no funders to report for this submission.

Date sharing statement: All requests for de-identified data should be made to the corresponding author. All reasonable requests will be evaluated by the CURATOR Data Governance and Sharing Committee comprising of FV, SLJ, BK and KN in the light of institutional policies and guidelines.

References

1. World Health Organization. WHO COVID-19 dashboard. https://covid19.who.int/. Published April 6, 2020. Accessed April 6, 2020.

2. Center for Systems Science and Engineering (CSSE) at Johns Hopkins University. COVID-19 map. Johns Hopkins Coronavirus Resource Center. https://coronavirus.jhu.edu/map.html. Published April 6, 2020. Accessed April 6, 2020.

3. Centers for Disease Control and Prevention. Coronavirus Disease 2019 (COVID-19) in the U.S. cdc.gov. https://www.cdc.gov/coronavirus/2019-ncov/cases-updates/cases-in-us.html. Published April 14, 2020. Accessed April 14, 2020.

4. Texas Health and Human Services. COVID-19 in Texas (dashboard). https://txdshs.maps.arcgis.com/apps/opsdashboard/index.html#/ed483ecd702b4298ab01e8b9cafc8b83. Published April 18, 2020. Accessed April 18, 2020.

5. McCann A. Most diverse cities in the U.S. Wallethub. https://wallethub.com/edu/most-diverse-cities/12690/. Published April 10, 2019. Accessed June 29, 2020.

6. Mejia B. How Houston has become the most diverse place in America. *Los Angeles Times*. https://www.latimes.com/nation/la-na-houston-diversity-2017-htmlstory.html. Published May 9, 2017. Accessed June 29, 2020.

7. Shi S, Qin M, Shen B, et al. Association of cardiac injury with mortality in hospitalized patients with COVID-19 in Wuhan, China. *JAMA Cardiol*. March 2020. doi:10.1001/jamacardio.2020.0950

8. Bonow RO, Fonarow GC, O'Gara PT, Yancy CW. Association of coronavirus disease 2019 (COVID-19) with myocardial injury and mortality. *JAMA Cardiol*. March 2020. doi:10.1001/jamacardio.2020.1105

9. 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*. April 2020. doi:10.1001/jama.2020.5394

10. Zhou F, Yu T, Du R, et al. Clinical course and risk factors for mortality of adult inpatients with COVID-19 in Wuhan, China: a retrospective cohort study. *Lancet*. 2020;395(10229):1054-1062. doi:10.1016/S0140-6736(20)30566-3

11. Deslatte M. Louisiana data: virus hits blacks, people with hypertension. *AP News*. https://apnews.com/cde897293f88f788d3f3bfdb3d46c466. Published April 7, 2020. Accessed April 20, 2020.

12. Reyes C, Husain N, Gutowski C, St. Clair S, Pratt G. Chicago's coronavirus disparity: Black Chicagoans are dying at nearly six times the rate of white residents, data show. *Chicago Tribune*. https://www.chicagotribune.com/coronavirus/ct-coronavirus-chicago-coronavirus-

deaths-demographics-lightfoot-20200406-77nlylhiavgjzb2wa4ckivh7mu-story.html. Published April 7, 2020. Accessed April 20, 2020.

13. Thebault R, Tran AB, Williams V. The coronavirus is infecting and killing black Americans at an alarmingly high rate. *Washington Post*. https://www.washingtonpost.com/nation/2020/04/07/coronavirus-is-infecting-killing-black-americans-an-alarmingly-high-rate-post-analysis-shows/. Published April 7, 2020. Accessed April 20, 2020.

14. United States Census Bureau. American Community Survey (ACS). https://www.census.gov/programs-surveys/acs. Published 2018. Accessed April 14, 2020.

15. United States Census Bureau. Gazetteer Files. https://www.census.gov/geographies/reference-files/time-series/geo/gazetteer-files.html. Accessed April 20, 2020.

16. Richiardi L, Bellocco R, Zugna D. Mediation analysis in epidemiology: methods, interpretation and bias. *Int J Epidemiol*. 2013;42(5):1511-1519. doi:10.1093/ije/dyt127

17. Harris County Public Health. *Harris Cares: A 2020 Vision of Health in Harris County*. Houston, TX: Harris County Public Health; 2019. http://publichealth.harriscountytx.gov/Resources/Harris-Cares. Accessed November 21, 2019.

18. Quinn SC, Kumar S, Freimuth VS, Musa D, Casteneda-Angarita N, Kidwell K. Racial disparities in exposure, susceptibility, and access to health care in the US H1N1 influenza pandemic. *Am J Public Health*. 2011;101(2):285-293. doi:10.2105/AJPH.2009.188029

19. Rocklov J, Sjodin H. High population densities catalyze the spread of COVID-19. J Travel Med. 2020. doi: 10.1093/jtm/taaa038

20. Tarwater PM, Martin CF. Effects of population density on the spread of disease. *Complexity*. 2001;6(6):29-36. doi:10.1002/cplx.10003

21. Yang K, LeJeune J, Alsdorf D, Lu B, Shum CK, Liang S. Global distribution of outbreaks of water-associated infectious diseases. *PLoS Negl Trop Dis*. 2012;6(2):e1483. doi:10.1371/journal.pntd.0001483

22. Wang R, Feng Z, Xue D, Liu Y, Wu R. Exploring the links between population density, lifestyle, and being overweight: secondary data analyses of middle-aged and older Chinese adults. *Health Qual Life Outcomes*. 2019;17(1):100. doi:10.1186/s12955-019-1172-3

23. Beenackers MA, Oude Groeniger J, Kamphuis CBM, Van Lenthe FJ. Urban population density and mortality in a compact Dutch city: 23-year follow-up of the Dutch GLOBE study. *Health & Place*. 2018;53:79-85. doi:10.1016/j.healthplace.2018.06.010

24. U.S. Bureau of Labor Statistics. Unemployment rates and earnings by educational attainment. https://www.bls.gov/emp/chart-unemployment-earnings-education.htm. Published September 4, 2019. Accessed April 15, 2020.

25. Centers for Disease Control and Prevention. Groups at higher risk for severe illness. cdc.gov. https://www.cdc.gov/coronavirus/2019-ncov/need-extra-precautions/groups-at-higher-risk.html. Last updated April 17, 2020. Accessed April 20, 2020.

26. Abuelgasim E, Saw LJ, Shirke M, Zeinah M, Harky A. COVID-19: Unique public health issues facing Black, Asian and minority ethnic communities. *Curr Probl Cardiol*. 2020;100621. doi:10.1016/j.cpcardiol.2020.100621

Tables and Figures

Figure 1: Adjusted Probability and 95% Confidence Interval of Positive SARS-CoV-2 PCR in Non-Hispanic Black vs. Non-Hispanic White by increasing age

Figure 2: Adjusted Probability and 95% Confidence Interval of Positive SARS-CoV-2 PCR in Hispanic vs. Non-Hispanic by increasing age

Table 1: Summary measures and univariable association of socio-demographic characteristics with SARS-CoV-2 infection from HM CURATOR

Characteristics	Overall (n = 20,228)	SARS-CoV-2 Negative (n = 18,677)	SARS-CoV-2 Positive (n = 1,551)	OR (95% CI)[a]
Age, mean (SD)	51.1 (19.0)	51.0 (19.1)	52.1 (18.1)	1.00 (1.00–1.01)
Age Categories (%)				
Up to 35 years	25.1	25.5	20.9	Reference Category
36–50 years	25.0	24.9	27.0	1.32 (1.14 – 1.54)
51–75 years	38.7	38.5	40.8	1.29 (1.12 – 1.48)
>75 years	11.2	11.2	11.4	1.23 (1.02 – 1.49)
Females (%)	61.9	62.3	58.3	0.85 (0.76 – 0.94)
Race (%)[b]				
White	62.3	62.9	55.7	Reference Category
Black	21.6	21.0	28.8	1.55 (1.37 – 1.75)
Asian	9.2	9.1	9.7	1.20 (1.00 – 1.44)
Mixed / Other	1.3	1.2	1.5	1.36 (0.88 – 2.11)
Hispanic (%)[c]	17.8	16.8	29.2	2.02 (1.79 – 2.27)
Median Zip Household Income (IQR)[d]	70,658 (53,313–99,276)	70,758 (53,633–100,107)	66,523 (50,485–94,226)	-4,168[e] (-6463.2, -1872.8)
Median Zip Household Income Pentiles (%)				
I: 13,893 – 50,485	18.2	17.8	23.6	Reference Category
II: 50,642 – 65,805	18.4	18.1	21.2	0.89 (0.75 – 1.04)
III: 65,897 – 79,869	18.3	18.2	19.6	0.82 (0.70 – 0.96)
IV: 80,039 – 106,067	18.6	19.0	13.7	0.55 (0.46 – 0.65)
V: 106,415 – 240,417	17.2	17.3	16.1	0.71 (0.60 – 0.83)
Median (IQR) Population Density[f]	2797.2 (1439.1 – 4260.9)	2797.2 (1439.1 – 4211.4)	3320.3 (1904.4 – 4439.7)	523.1[e] (454.9 – 591.3)
Median Population Density Pentiles (%)				
I: 1.5 – 1026.6	18.2	18.6	13.5	Reference Category
II: 1034.6 – 2306.3	19.0	19.1	18.1	1.31 (1.09 – 1.58)
III: 2330.8 – 3328.9	17.4	17.4	18.3	1.45 (1.21 – 1.75)
IV: 3360.1 – 4665.6	18.1	17.8	22.1	1.71 (1.43 – 2.05)
V: 4742.6 – 98025.9	18.1	17.7	22.6	1.76 (1.47 – 2.10)
Insurance Status (%)				
Medicare	29.0	29.1	27.2	Reference Category
Medicaid	4.7	4.8	4.3	0.96 (0.73 – 1.26)
Pvt / Employer based	42.6	43.2	36.4	0.90 (0.79 – 1.03)
HC Exchange	1.7	1.6	3.0	2.02 (1.46 – 2.80)
Self-Pay	20.6	20.0	28.4	1.52 (1.33 – 1.75)
VA	1.3	1.4	0.7	0.54 (0.29 – 1.00)
Charlson Co-morbidity Index, Median (IQR)	2 (0 – 6)	2 (0 – 6)	2 (0 – 5)	0 (-0.36, 0.36)[f]
Charlson Co-morbidity Index (CCI) Categories (%)				
CCI: 0	33.1	33.4	30.4	Reference Category

Characteristics	Overall (n = 20,228)	SARS-CoV-2 Negative (n = 18,677)	SARS-CoV-2 Positive (n = 1,551)	OR (95% CI)[a]
CCI: 1 – 2	23.7	23.1	28.6	1.35 (1.18 – 1.54)
CCI: 3 – 6	20.3	20.2	20.4	0.10 (0.65 – 1.28)
CCI: > 6	22.9	23.1	20.6	0.98 (0.84 – 1.13)
Hypertension	47.2	47.1	48.4	1.06 (0.95 – 1.17)
Diabetes (without complications)	24.2	23.7	30.3	1.40 (1.24– 1.57)
Obesity	28.0	28.2	25.2	0.86 (0.76 – 0.96)

[a] Unadjusted Odds Ratios and 95% Confidence Intervals for association between individual co-variates in SRAS-CoV-2 positivity.
[b] Race: Missing, Unknown, Declined, n = 1157 (5.7%).
[c] Ethnicity: Missing, Unknown, Declined, n = 613 (3.0%).
[d] 2018 inflation adjusted USD. Missing n = 1,883 (9.3%). Pentiles were defined by categorizing the ordered distribution of median income into five categories.
[e] Difference in median and 95% CI of difference obtained via quantile regression.
[f] Population density Missing n = 1,854 (9.2%). Pentiles were defined by categorizing the ordered distribution of population density into five categories.
HM CURATOR: Houston Methodist, COVID-19 Surveillance and Outcomes Registry
SARS-CoV-2: Severe Acute Respiratory Syndrome Coronavirus 2

Table 2: Univariable comparison of socio-demographic and comorbidity factors between non-Hispanic Black (NHB) and non-Hispanic White (NHW) race categories

	Non-Hispanic Black n = 4,285	Non-Hispanic White n = 9,469	OR[a] / Median Difference (95% CI)	P value
Age: Mean (SD)	49.4 (17.8)	56.0 (19.2)	0.98 (0.98 – 0.98)	< 0.001
Age Category (%)				
Up to 35	24.3	18.8	Reference	
36 – 50	28.8	19.8	1.12 (1.01 – 1.25)	0.03
51 – 75	39.1	44.9	0.68 (0.61 – 0.74)	< 0.001
> 75	7.8	16.5	0.37 (0.32 – 0.42)	< 0.001
Females	68.1	58.1	1.54 (1.42 – 1.66)	< 0.001
Median (IQR) Zip Income	64,022 (47,303 – 79,658)	76,163 (60,130 – 102,019)	-12,141 (-14,018, -10,263)	< 0.001
Median Zip Income Pentiles (%) – Pentiles of increasing Income[b]				
Category I	33.5	13.1	Reference	
Category II	21.3	19.7	0.42 (0.38 – 0.47)	< 0.001
Category III	22.7	19.1	0.47 (0.42 – 0.52)	< 0.001
Category IV	10.5	26.2	0.16 (0.14 – 0.18)	< 0.001
Category V	11.9	22.0	0.21 (0.19 – 0.24)	< 0.001
Population Density for Zip: Median (IQR)	3217.6 (2040.7 – 4439.7)	2488.5 (812.2 – 4084.3)	729.1 (603.0 – 855.2)	< 0.001
Population Density for Zip Pentiles (%) – Pentiles of increasing population density[c]				
Category I	11.5	28.3	Reference	
Category II	21.7	20.9	2.52 (2.21 – 2.86)	< 0.001
Category III	21.8	16.9	3.20 (2.81 – 3.65)	< 0.001
Category IV	23.0	15.5	3.66 (3.21 – 4.17)	< 0.001
Category V	22.2	18.3	2.99 (2.63 – 3.41)	< 0.001
Charlson Comorbidity Index, median (IQR)	2 (0 – 6)	3 (1 – 7)	-1 (-1.22, -0.78)	< 0.001
Charlson Comorbidity Index (CCI) Categories (%)				
CCI: 0	30.8	24.9	Reference	
CCI: 1 – 2	25.1	22.3	0.91 (0.82 – 1.01)	0.07
CCI: 3 – 6	19.7	24.0	0.66 (0.60 – 0.74)	< 0.001
CCI: > 6	24.4	28.8	0.68 (0.62 – 0.75)	< 0.001
Hypertension	56.0	52.6	1.14 (1.07 – 1.23)	< 0.001
Diabetes (without complications)	29.6	21.9	1.50 (1.38 – 1.63)	< 0.001
Obesity	33.1	30.8	1.11 (1.03 – 1.20)	0.008

[a] *Unadjusted Odds Ratios and 95% Confidence Intervals for association between individual co-variates in SRAS-CoV-2 positivity or Difference in median and 95% CI of difference obtained via quantile regression.*

[b,c] *Pentiles were defined by categorizing the ordered distribution of median income and population density into five categories.*

Table 3: Univariable comparison of socio-demographic and comorbidity factors between Hispanic and non-Hispanic ethnicities

	Hispanic n = 3,590	Non-Hispanic n = 16,025	OR[a] / Median Difference[a] (95% CI)	P value
Age: Mean (SD)	45.1 (18.3)	52.8 (18.9)	0.98 (0.98 – 0.98)	< 0.001
Age Category (%)				
Up to 35	36.0	21.8	Reference	
36 – 50	26.1	24.6	0.64 (0.58 – 0.71)	< 0.001
51 – 75	32.0	40.1	0.47 (0.43 – 0.52)	< 0.001
> 75	5.9	12.7	0.28 (0.24 – 0.36)	< 0.001
Females	64.5	61.4	1.14 (1.06 – 1.23)	0.001
Median (IQR) Zip Income	65,742 (48,345 – 82,708)	73,742 (56,288 – 102,008)	-8,000 (-9,450.5, -6,549.5)	< 0.001
Median Zip Income Pentiles (%) – Pentiles of increasing Income[b]				
Category I	29.2	18.2	Reference	
Category II	23.3	19.6	0.74 (0.66 – 0.83)	< 0.001
Category III	21.8	19.8	0.68 (0.61 – 0.76)	< 0.001
Category IV	15.0	21.6	0.43 (0.38 – 0.49)	< 0.001
Category V	10.6	20.8	0.32 (0.28 – 0.36)	< 0.001
Population Density for Zip: Median (IQR)	3256.8 (1504.0 – 4299.7)	2741.6 (1408.1 – 4110.7)	515.2 (469.2 – 561.2)	< 0.001
Population Density for Zip Pentiles (%) – Pentiles of increasing population density[c]				
Category I	15.5	21.2	Reference	
Category II	21.1	20.9	1.39 (1.22 – 1.57)	< 0.001
Category III	15.3	20.0	1.05 (0.92 – 1.20)	0.48
Category IV	27.0	18.3	2.02 (1.79 – 2.28)	< 0.001
Category V	21.1	19.5	1.48 (1.31 – 1.68)	< 0.001
Charlson Comorbidity Index, median (IQR)	1 (0 – 4)	2 (0 – 6)	-1 (-1.21, -0.79)	< 0.001
Charlson Comorbidity Index (CCI) Categories (%)				
CCI: 0	41.8	29.7	Reference	
CCI: 1 – 2	25.4	23.6	0.76 (0.70 – 0.84)	< 0.001
CCI: 3 – 6	15.9	21.7	0.52 (0.47 – 0.58)	< 0.001
CCI: > 6	16.9	24.9	0.48 (0.43 – 0.54)	< 0.001
Hypertension	38.0	50.5	0.60 (0.56 – 0.65)	< 0.001
Diabetes (without complications)	27.4	23.6	1.22 (1.12 – 1.33)	< 0.001
Obesity	27.3	28.9	0.92 (0.85 – 1.00)	0.06

[a] *Unadjusted Odds Ratios and 95% Confidence Intervals for association between individual co-variates in SRAS-CoV-2 positivity or Difference in median and 95% CI of difference obtained via quantile regression.*
[b,c] *Pentiles were defined by categorizing the ordered distribution of median income and population density into five categories.*

Table 4: Adjusted Odds Ratios and 95% Confidence Intervals for likelihood of SARS-CoV-2 positivity among minority race and ethnic groups

Covariate	NHB vs. NHW, aOR (95% CI)[a]	Hispanic vs. Non-Hispanic, aOR (95% CI)[b]
Non-Hispanic Black (vs. Non-Hispanic White)	**2.23 (1.90 – 2.60)[c]**	
Hispanic (vs. Non-Hispanic)		**1.95 (1.72 – 2.20)[c]**
Age Categories		
Up to 35 years	*Reference Category*	
36–50 years	1.36 (1.07 – 1.72)	1.42 (1.20 – 1.68)
51–75 years	1.60 (1.21 – 2.11)	1.71 (1.39 – 2.11)
>75 years	2.20 (1.52 – 3.19)	2.08 (1.56 – 2.77)
Male (vs. Female)	1.20 (1.04 – 1.39)	1.17 (1.05 – 1.32)
Median Zip Household Income Categories (Pentiles of Increasing Income)		
Category I	*Reference Category*	
Category II	0.95 (0.76 – 1.18)	
Category III	1.02 (0.84 – 1.26)	
Category IV	0.74 (0.58 – 0.94)	
Category V	0.97 (0.77 – 1.22)	
Primary Insurance Type		
Medicare	*Reference Category*	
Medicaid	0.86 (0.55 – 1.33)	1.01 (0.74 – 1.37)
Private / Employer Based	1.11 (0.88 – 1.40)	0.93 (0.80 – 1.12)
Healthcare Exchange	2.06 (1.25 – 3.40)	1.80 (1.27 – 2.54)
Self-Pay	1.75 (1.33 – 2.30)	1.80 (1.27 – 2.54)
Veterans Affairs	0.72 (0.36 – 1.43)	0.58 (0.31 – 1.07)
Charlson Comorbidity Index Categories		
CCI: 0	*Reference Category*	
CCI: 1 – 2	1.26 (0.99 – 1.60)	1.03 (0.87 – 1.23)
CCI: 3 – 6	0.94 (0.68 – 1.28)	0.71 (0.56 – 0.91)
CCI: > 6	0.91 (0.65 – 1.30)	0.58 (0.44 – 0.76)
Hypertension	0.87 (0.73 – 1.05)	1.00 (0.87 – 1.15)
Diabetes (Without Complications)	1.42 (1.17 – 1.71)	1.62 (1.40 – 1.87)
Obesity	0.82 (0.70 – 0.97)	0.83 (0.73 – 0.94)

[a,b] Hosmer and Lemeshow goodness of fit p-value: 0.58 and 0.73 (H_0: Model fit is correct).

[c] ORs (95% CIs) represent the direct association adjusted for all other covariates in the model.

NHB: Non-Hispanic Black, NHW: Non-Hispanic White. Grayed out cells were not included in respective models either due to collinearity or demonstration of statistically significant mediation towards likelihood of SARS-CoV-2 positivity.

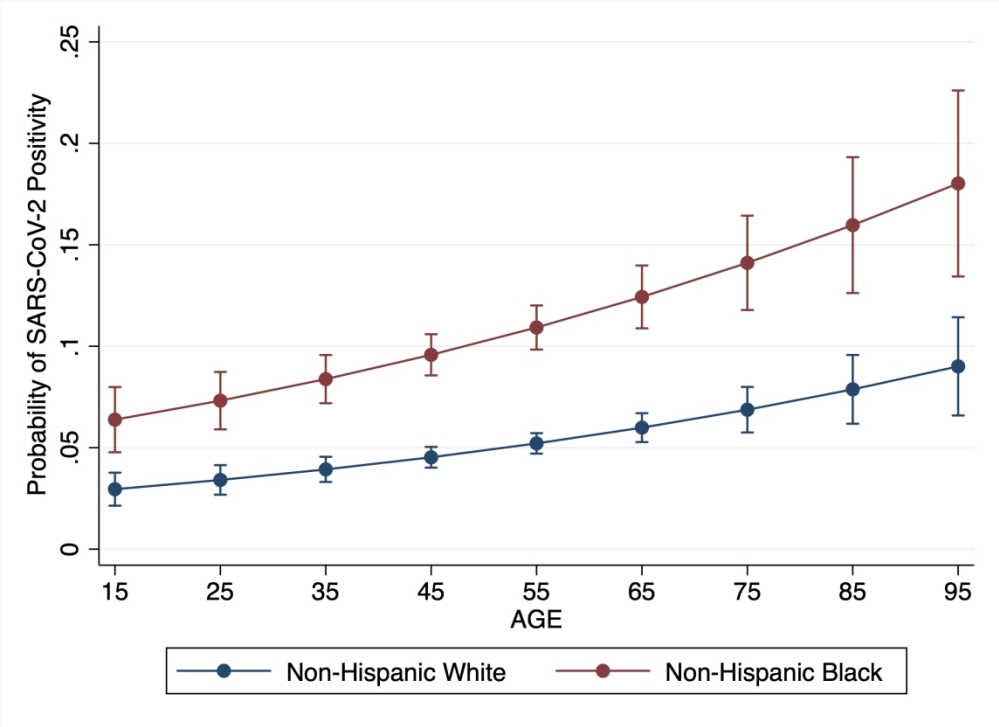


Figure 1: Adjusted Probability and 95% Confidence Interval of Positive SARS-CoV-2 PCR in Non-Hispanic Black vs. Non-Hispanic White by increasing age

467x339mm (144 x 144 DPI)

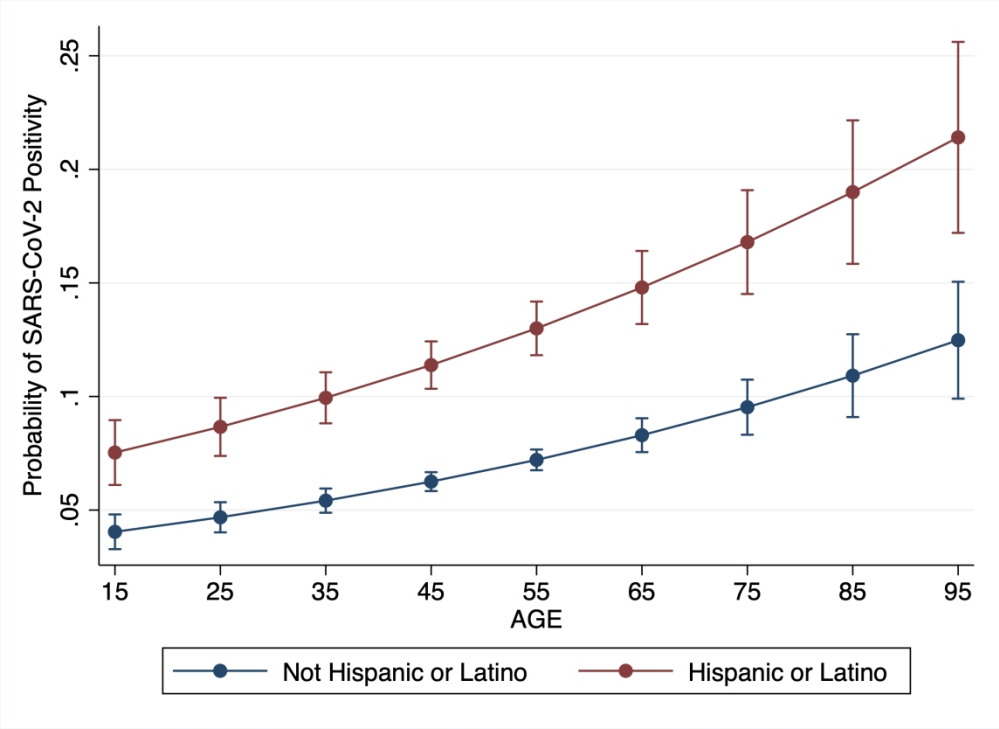


Figure 2: Adjusted Probability and 95% Confidence Interval of Positive SARS-CoV-2 PCR in Hispanic vs. Non-Hispanic by increasing age

467x339mm (144 x 144 DPI)

STROBE Statement—Checklist of items that should be included in reports of ***cross-sectional studies***

	Item No	Recommendation	Page No
Title and abstract	1	(*a*) Indicate the study's design with a commonly used term in the title or the abstract	2
		(*b*) Provide in the abstract an informative and balanced summary of what was done and what was found	2-3
Introduction			
Background/rationale	2	Explain the scientific background and rationale for the investigation being reported	4
Objectives	3	State specific objectives, including any prespecified hypotheses	5
Methods			
Study design	4	Present key elements of study design early in the paper	5
Setting	5	Describe the setting, locations, and relevant dates, including periods of recruitment, exposure, follow-up, and data collection	5-6
Participants	6	(*a*) Give the eligibility criteria, and the sources and methods of selection of participants	6
Variables	7	Clearly define all outcomes, exposures, predictors, potential confounders, and effect modifiers. Give diagnostic criteria, if applicable	6-7
Data sources/ measurement	8*	For each variable of interest, give sources of data and details of methods of assessment (measurement). Describe comparability of assessment methods if there is more than one group	6-7
Bias	9	Describe any efforts to address potential sources of bias	N/A
Study size	10	Explain how the study size was arrived at	8
Quantitative variables	11	Explain how quantitative variables were handled in the analyses. If applicable, describe which groupings were chosen and why	6-7
Statistical methods	12	(*a*) Describe all statistical methods, including those used to control for confounding	7
		(*b*) Describe any methods used to examine subgroups and interactions	7-8
		(*c*) Explain how missing data were addressed	N/A
		(*d*) If applicable, describe analytical methods taking account of sampling strategy	N/A
		(*e*) Describe any sensitivity analyses	N/A
Results			
Participants	13*	(a) Report numbers of individuals at each stage of study—eg numbers potentially eligible, examined for eligibility, confirmed eligible, included in the study, completing follow-up, and analysed	8
		(b) Give reasons for non-participation at each stage	N/A
		(c) Consider use of a flow diagram	N/A
Descriptive data	14*	(a) Give characteristics of study participants (eg demographic, clinical, social) and information on exposures and potential confounders	8, 21-22
		(b) Indicate number of participants with missing data for each variable of interest	21-22
Outcome data	15*	Report numbers of outcome events or summary measures	8
Main results	16	(*a*) Give unadjusted estimates and, if applicable, confounder-adjusted estimates and their precision (eg, 95% confidence interval). Make clear which confounders were adjusted for and why they were included	8-9, 10, 21-22

		(*b*) Report category boundaries when continuous variables were categorized	21-22
		(*c*) If relevant, consider translating estimates of relative risk into absolute risk for a meaningful time period	N/A
Other analyses	17	Report other analyses done—eg analyses of subgroups and interactions, and sensitivity analyses	9-10. 23-25
Discussion			
Key results	18	Summarise key results with reference to study objectives	11
Limitations	19	Discuss limitations of the study, taking into account sources of potential bias or imprecision. Discuss both direction and magnitude of any potential bias	13-14
Interpretation	20	Give a cautious overall interpretation of results considering objectives, limitations, multiplicity of analyses, results from similar studies, and other relevant evidence	12,14
Generalisability	21	Discuss the generalisability (external validity) of the study results	13
Other information			
Funding	22	Give the source of funding and the role of the funders for the present study and, if applicable, for the original study on which the present article is based	15

*Give information separately for exposed and unexposed groups.

Note: An Explanation and Elaboration article discusses each checklist item and gives methodological background and published examples of transparent reporting. The STROBE checklist is best used in conjunction with this article (freely available on the Web sites of PLoS Medicine at http://www.plosmedicine.org/, Annals of Internal Medicine at http://www.annals.org/, and Epidemiology at http://www.epidem.com/). Information on the STROBE Initiative is available at www.strobe-statement.org.

BMJ Open

Racial and Ethnic Disparities in SARS-CoV-2 Pandemic: Analysis of a COVID-19 Observational Registry for a Diverse U.S. Metropolitan Population

Journal:	*BMJ Open*
Manuscript ID	bmjopen-2020-039849.R4
Article Type:	Original research
Date Submitted by the Author:	23-Jul-2020
Complete List of Authors:	Vahidy, Farhaan S; Houston Methodist Research Institute, Center for Outcomes Research; Houston Methodist Neurological Institute Nicolas, Juan Carlos; Houston Methodist Research Institute Meeks, Jennifer R; Houston Methodist Research Institute Khan, Osman; Houston Methodist Research Institute Pan, Alan; Houston Methodist Research Institute, Center for Outcomes Research Jones, Stephen L.; Houston Methodist Research Institute; Weill Cornell Medicine Masud, Faisal; Houston Methodist Hospital; Weill Cornell Medicine Sostman, H Dirk; Houston Methodist Research Institute; Weill Cornell Medicine Phillips, Robert; Houston Methodist Hospital; Weill Cornell Medicine Andrieni, Julia D; Houston Methodist Hospital; Weill Cornell Medicine Kash, Bita A; Houston Methodist Research Institute; Texas A&M University School of Rural Public Health Nasir, Khurram; Houston Methodist Research Institute; Houston Methodist Hospital
Primary Subject Heading:	Epidemiology
Secondary Subject Heading:	Epidemiology, Infectious diseases, Public health
Keywords:	EPIDEMIOLOGY, Public health < INFECTIOUS DISEASES, INFECTIOUS DISEASES

I, the Submitting Author has the right to grant and does grant on behalf of all authors of the Work (as defined in the below author licence), an exclusive licence and/or a non-exclusive licence for contributions from authors who are: i) UK Crown employees; ii) where BMJ has agreed a CC-BY licence shall apply, and/or iii) in accordance with the terms applicable for US Federal Government officers or employees acting as part of their official duties; on a worldwide, perpetual, irrevocable, royalty-free basis to BMJ Publishing Group Ltd ("BMJ") its licensees and where the relevant Journal is co-owned by BMJ to the co-owners of the Journal, to publish the Work in this journal and any other BMJ products and to exploit all rights, as set out in our licence.

The Submitting Author accepts and understands that any supply made under these terms is made by BMJ to the Submitting Author unless you are acting as an employee on behalf of your employer or a postgraduate student of an affiliated institution which is paying any applicable article publishing charge ("APC") for Open Access articles. Where the Submitting Author wishes to make the Work available on an Open Access basis (and intends to pay the relevant APC), the terms of reuse of such Open Access shall be governed by a Creative Commons licence – details of these licences and which Creative Commons licence will apply to this Work are set out in our licence referred to above.

Other than as permitted in any relevant BMJ Author's Self Archiving Policies, I confirm this Work has not been accepted for publication elsewhere, is not being considered for publication elsewhere and does not duplicate material already published. I confirm all authors consent to publication of this Work and authorise the granting of this licence.

Racial and Ethnic Disparities in SARS-CoV-2 Pandemic: Analysis of a COVID-19 Observational Registry for a Diverse U.S. Metropolitan Population

Farhaan S. Vahidy PhD MBBS MPH,[1,2] Juan Carlos Nicolas BS,[1] Jennifer R. Meeks MS,[1] Osman Khan BS,[1] Alan Pan MS,[1] Stephen L. Jones MD MSHI,[1,3,4] Faisal N. Masud MD,[4,5] H. Dirk Sostman MD,[4,6] Robert A. Phillips MD PhD,[1,4,7] Julia D. Andrieni MD,[4,8] Bita Kash PhD MBA,[1,4,9] Khurram Nasir MD MPH[1,7]

1. Center for Outcomes Research, The Houston Methodist Research Institute, Houston TX
2. Houston Methodist Neurological Institute, Houston Methodist Hospital, Houston TX
3. Department of Surgery, Houston Methodist Hospital, Houston TX
4. Weill Cornell Medical College, New York NY
5. Department of Anesthesiology and Critical Care, Houston Methodist Hospital, Houston TX
6. The Houston Methodist Research Institute and The Houston Methodist Academic Institute, Houston TX
7. Department of Cardiology, DeBakey Heart and Vascular Center, Houston Methodist Hospital, Houston TX
8. Department of Medicine, Houston Methodist Hospital, Houston TX
9. Department of Health Policy and Management, Texas A&M University School of Public Health, College Station TX

Corresponding Author

Farhaan S. Vahidy, PhD MBBS MPH
Josie Roberts Building, Suite 4.123
Houston Methodist
7550 Greenbriar Drive
Houston TX 77030
fvahidy@houstonmethodist.org

Word Count: 3,126

Abstract

Introduction: Data on race and ethnic disparities for SARS-CoV-2 infection are limited. We analyzed socio-demographic factors associated with higher likelihood of SARS-CoV-2 infection and explore mediating pathways for race and ethnic disparities in the SARS-CoV-2 pandemic.

Methods: Cross-sectional analysis of COVID-19 Surveillance and Outcomes Registry (CURATOR), which captures data for a large healthcare system, comprising of one central tertiary-care hospital, seven large community hospitals, and an expansive ambulatory / emergency care network in the Greater Houston area. Nasopharyngeal samples for individuals inclusive of all ages, races, ethnicities and sex were tested for SARS-CoV-2. We analyzed socio-demographic (age, sex, race, ethnicity, household income, residence population density) and comorbidity (Charlson Comorbidity Index, hypertension, diabetes, obesity) factors. Multivariable logistic regression models were fitted to provide adjusted Odds Ratios (aOR) and 95% confidence intervals (CI) for likelihood of a positive SARS-CoV-2 test. Structural Equation Modeling (SEM) framework was utilized to explore three mediation pathways (low income, high population density, high comorbidity burden) for association between Non-Hispanic Black race (NHB), Hispanic ethnicity, and SARS-CoV-2 infection.

Results: Among 20,228 tested individuals, 1,551 (7.7%) tested positive. Overall mean (SD) age was 51.1 (19.0) years, 62% females, 22% Black and 18% were Hispanic. NHB and Hispanic ethnicity was associated with lower socio-economic status and higher population density residence. In the fully adjusted model, NHB (vs. Non-Hispanic White; aOR, CI: 2.23, 1.90-2.60) and Hispanic ethnicity (vs. non-Hispanic; aOR, CI: 1.95, 1.72-2.20) had a higher likelihood of infection. Older individuals and males were also at higher risk of infection. The SEM framework

demonstrated a significant indirect effect of NHB and Hispanic ethnicity on SARS-CoV-2 infection mediated via a pathway including residence in densely populated zip code.

Conclusions: There is strong evidence of race and ethnic disparities in the SARS-CoV-2 pandemic, that is potentially mediated through unique social determinants of health.

Strengths and limitations of this study

- One of the first studies to systematically evaluate race and ethnic disparities in susceptibility to SARS-CoV-2 infection, while accounting for multiple socio-demographic characteristics and comorbidities
- Study population represents a large and diverse metropolitan of the U.S. with data from one of the largest healthcare providers across the greater metropolitan area
- Study evaluates potential mediation pathways for race disparities and demonstrates that residence in areas with high population density may mediate race and ethnic disparities in susceptibility to SARS-CoV-2 infection
- Single center study with limited information about burden of comorbidity and lifestyle factors

INTRODUCTION

The Coronavirus disease (COVID-19), caused by infection with the Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2), is a pandemic that has thus far resulted in over 9.5 million cases globally in under 6 months. At the time of this reporting, the United States (U.S.) has approximately 25% of total global cases and has surpassed all countries in terms of absolute number of cases, cases per 1 million population, and fatalities.[1,2] Experts project these numbers to continue rising as widespread testing is instituted and newer patterns of infectivity emerge. The geographic distribution of cases across the U.S. demonstrates that the predominant pandemic burden hit major metropolitan areas. However, cases of COVID-19 have been reported across all 50 states, the District of Columbia, Guam, Puerto Rico, the Northern Mariana Islands, and the U.S. Virgin Islands.[3] As of May 31, 2020, the state of Texas had 64,287 reported cases of COVID-19, with about one-third in the Greater Houston area.[4] The Greater Houston area is home to approximately 7 million individuals, is the fourth-largest metropolitan area by population in the U.S., and is considered one of the nation's most diverse regions.[5-6]

Initial reports indicate that specific individuals such as the elderly; males; and people with comorbidities including hypertension, diabetes, obesity, coronary artery disease and heart failure have poor COVID-19 outcomes.[7–10] As the pandemic spread over the continental U.S. during the last four months, patterns of high-risk phenotypes started to emerge and reports of poor outcomes (particularly high case fatality) among racial minorities surfaced.[11–13] Though it is important to understand the determinants of poor outcomes among COVID-19 patients, it is equally imperative, from a public health perspective, to systematically examine the likelihood of SARS-CoV-2 infection across large diverse communities in the U.S. Data on higher likelihood of SARS-CoV-2 infection among racial and ethnic minorities across diverse U.S. metropolitan

areas are limited. Furthermore, the mediators of SARS-CoV-2 infection among racial and ethnic minorities have not been described.

We explored socio-demographic characteristics such as age, sex, race, ethnicity, median household income by zip codes, population density of residents' zip codes, and health insurance status associated with positive SARS-CoV-2 testing in an urban and diverse population served by one of the leading healthcare systems of the Greater Houston area. We further examined the association between pre-existing comorbidities and higher likelihood of SARS-CoV-2 infection in our study population. We hypothesized that older age, and racial and ethnic minorities will be associated with significantly higher likelihood of SARS-CoV-2 infection, and factors such as low socio-economic status, residence in high population density areas (proxy for potential difficulties in social distancing) and higher comorbidity burden will mediate the effect of race and ethnicity on SARS-CoV-2 infection.

METHODS

We analyzed data between March 5 and May 31, 2020 collected as a part of the COVID-19 Surveillance and Outcomes Registry (CURATOR) at Houston Methodist (HM). The HM CURATOR has been approved by the HM Institutional Review Board (IRB) as an observational quality of care registry for all suspected and confirmed COVID-19 patients. HM IRB granted CURATOR a waiver of informed consent and HIPAA (Health Insurance Portability and Accountability Act) authorization in accordance with current federal regulations. The CURATOR, designed and managed by the big data team at the Center for Outcomes Research (COR) at HM, is populated from multiple data sources across the HM system such as electronic medical records, electronic databanks for laboratory and pharmacy, and electronic interactive

patient interface tools. The HM system comprises a flagship tertiary care hospital in the Texas Medical Center, seven large community hospitals, a continuing care hospital, and multiple emergency centers and clinics throughout the Greater Houston area. Data from various sources are curated into a harmonized format, assessed for quality and integrity, and stored on a secure institutional HIPAA-compliant server.

We flagged all individuals who were tested for the SARS-CoV-2 using the real time Reverse Transcriptase (RT) Polymerized Chain Reaction (PCR) diagnostic panels. The three cross-validated PCR tests utilized were the World Health Organization (WHO) nucleic acid amplification test, Panther Fusion® SARS-CoV-2 Assay, and Cepheid Xpert® Xpress SARS-CoV-2 Assay. These assays were verified for quantitative detection of novel SARS-CoV-2 isolated and purified from nasopharyngeal swab specimens obtained from individuals and immersed in universal transport medium. Testing was carried out for symptomatic individuals or for individuals who had a self-reported history of exposure to a COVID-19 case including recent travel to other countries with high infection rates or hotspots within the U.S.

Socio-demographic characteristics including age, sex, race, ethnicity, and payer-status (insurance type) were obtained from the HM CURATOR for analyses. We also extracted information on presence of comorbidities comprising the Charlson Comorbidity Index (CCI) which include past history of myocardial infarction, congestive heart failure, peripheral vascular disease, cerebrovascular disease, dementia, chronic pulmonary disease, rheumatic disease, peptic ulcer disease, liver disease, diabetes with or without complications, hemiplegia, renal disease, any malignancy (excluding skin neoplasms), metastatic solid tumors, and AIDS/HIV. Data on hypertension and obesity were additionally obtained. We utilized the U.S. Census Bureau's American Community Survey (ACS) 5-year data (2014–2018) to determine median household

income by individual zip code tabulation areas (ZCTA).[14] The median ZCTA household income was inflation-adjusted to 2018 USD. We also utilized the same data source to obtain population estimates by ZCTA, and calculated ZCTA level population density (population per mile square) by standardizing it for area measurements of ZCTA. For the purpose of population density determination, land area estimates were obtained from the Census Bureau's U.S. Gazetteer Files 2010.[15] In the absence of granular and precise social distancing data, we have utilized population density as a proxy for potential difficulties in social distancing among crowded communities.

We provide descriptive summary data as means (standard deviations) and proportions. We fit univariable and multivariable logistic regression models to assess unadjusted and adjusted association between socio-demographic characteristics and likelihood of being tested positive for SARS-CoV-2. We additionally provide univariable comparison of various socio-demographic and comorbidity variables between non-Hispanic Black (NHB) and non-Hispanic White (NHW) race categories, as well as between Hispanic and Non-Hispanic ethnic groups. Age, income, population density and CCI were categorized for certain analyses. We included age, sex, race, ethnicity, zip code household income, insurance type, zip population density and CCI in our initial multivariable model. Zip code household income, zip population density and CCI were evaluated as mediators. Factors demonstrating mediation were excluded from the final models. However, the factors that did not demonstrate mediation were included in the final models, as we believe that they continue to importantly inform the variance of estimates for direct effects.[16] We assessed the model fit utilizing the Hosmer-Lemeshow goodness of fit test, and crude and adjusted odds ratios (OR and aOR) and 95% confidence intervals (CI) are reported. Post-estimation marginal probabilities of SARS-CoV-2 infection were determined from the final adjusted model for major covariates (race, ethnicity and age). We explored the mediation

influence of comorbidity burden (CCI), socio-economic status (median income), and lack of social distancing (population density) on the relationship of Black race and Hispanic ethnicity with high likelihood of SARS-CoV-2 infection using the Generalized Structural Equation Modeling (GSEM) framework. The GSEM framework was set up to provide estimates of direct and indirect effect of Black race and Hispanic ethnicity on SARS-CoV-2 infectivity. Statistically significant ($p < 0.05$) indirect effects represent full or partial mediation by a tested covariate. We included all individuals tested for SARS-CoV-2 across our healthcare system and did not perform formal sample size calculations.

Patient and Public Involvement

There was no direct patient or public involvement in the design and conduct of this study.

RESULTS

Socio-demographic and comorbidity characteristics of the study population

Across the time period of analysis, we identified a total of 20,228 presumed cases tested for SARS-CoV-2, among whom 1,551 (7.7%, CI: 7.3-8.0) tested positive. Overall, the mean (SD) age of the study population was 51.1 (19.0) years; 61.9% were female and 62.3% were White (including Hispanic ethnicity). The study sample was comparable to the overall population of patients treated across HM, who have a mean (SD) age of 49.0 (22) years, are 56% female, and 53% White. The HM system metrics was derived from a sample of 3,216,290 patients managed across the system since May 22, 2016.

The overall median (IQR) household income was USD $70,658 ($53,313–$99,276), and 42.6% of the study population had private or employer-based insurance. In our univariate analysis, Black race (vs. White; OR, CI: 1.55, 1.37–1.75), Hispanic ethnicity (vs. non-Hispanic;

OR, CI: 2.02, 1.79–2.27), and males (vs. females; OR, CI: 1.17, 1.06–1.31) were associated with significantly higher likelihood of testing positive for SARS-CoV-2. Among the SARS-CoV-2 positive patients, 40.8% were in the age category of 51–75 years, and 11.4% were greater than 75 years. These proportions were significantly higher than the reference group (up to 35 years; OR, CI for 51-75 years vs. up to 35 years: 1.29, 1.12–1.48 and for >75 years vs. up to 35 years: 1.23, 1.02–1.49). Furthermore, individuals in higher pentiles of socio-economic status had significantly lower likelihood, whereas those residing in higher population density ZCTAs had higher likelihood of SARS-CoV-2 infection. We observed a significantly higher proportion of SARS-CoV-2 positive individuals in the CCI 1-2 category compared to CCI of 0 (OR, CI: 1.35, 1.18–1.54). However, similar differences for higher CCI categories were not observed. For specific comorbidities, a significantly greater proportion of diabetic individuals had SARS-CoV-2 positive results (OR, CI: 1.40, 0.17–1.68). The socio-demographic characteristics and comorbidity profiles for the overall and SARS-CoV-2 positive and negative patients are summarized in Table 1.

Socio-demographic and comorbidity characteristics associated with minority race and ethnicity

In our study sample comprising of 13,754 Non-Hispanic Black and White individuals, we compared the association between race and various socio-demographic and comorbidity characteristics (Table 2). Similarly, we also evaluated univariable differences for socio-demographic variables and co-morbidities between Hispanic and non-Hispanic individuals (Table 3). Minority race (NHB) and ethnicity (Hispanic) were both associated with younger age, higher proportion of females, and residence in low income and higher population density ZCTAs. However, NHB and Hispanic groups were both associated with an overall lower burden

of comorbidities (as demonstrated by significantly lower median CCI) compared respectively to NHW and non-Hispanic categories. A higher proportion of individuals among minority race and ethnicity were diabetic, and a higher proportion of NHB were also hypertensive compared to NHW.

Multivariable model and marginal probabilities for likelihood of SARS-CoV-2 infection and racial and ethnic minorities

The significantly higher likelihood of SARS-CoV-2 infection among minority race and ethnic groups persisted after controlling for other demographics, insurance type, median household income, population density, and comorbidities. Adjusted odds ratios (CI) for NHB vs. NHW was 2.23 (1.90 – 2.60) and for Hispanic vs. Non-Hispanic was 1.95 (1.72 – 2.20). Higher risk of infection among males (compared to females) and higher likelihood of SARS-CoV-2 infection among elderly also remained statistically significant. Detailed outputs of the fully adjusted logistic regression models for minority race and ethnic groups are presented in Table 4. Based on the marginal probabilities obtained from our fully adjusted model, the probability of SARS-CoV-2 infection in a 45-year-old NHB is 9.6% whereas it is 4.5% in a 45-year-old NHW individual, all other adjusted variables being constant. At the age of 75, this probability is 14.0% for an NHB and 6.9% for a NHW. A similar relationship differential was observed for Hispanic vs. non-Hispanic individuals. Multivariable model derived probabilities of SARS-CoV-2 infection for NHB vs. NHW and for Hispanic vs. Non-Hispanic across age spectrum are presented in Figure 1 and Figure 2.

Generalized Structural Equation Modeling for mediation by income, population density and Comorbidity Index

Utilizing the GSEM framework, we determined the direct and indirect effects of NHB and Hispanic ethnicity on SARS-CoV-2 infection with median income, population density and CCI modeled as mediators in six separate equations adjusted for age and sex. The indirect effect of NHB mediated through population density was statistically significant (OR, CI: 1.03, 1.01 – 1.05, p = 0.001); however, the indirect effects mediated via median income and comorbidity scores were not statistically significant (p = 0.14 and p = 0.64 respectively). Among individuals identifying as Hispanic or Latino, both population density and income partially mediated the effect of ethnicity on SARS-CoV-2 positivity (OR, CI for population density: 1.02, 1.01 – 1.02, p < 0.001 and OR, CI for income: 1.04, 1.02 – 1.06, p < 0.001). Evaluation of comorbidities did not suggest a mediation influence for either NHB or Hispanic categories.

DISCUSSION

The underlying race and ethnic healthcare disparities have been painfully highlighted in the wake of the COVID-19 pandemic. Most reports indicate higher mortality or case fatality among minority racial groups (Black / African American) across major U.S. metropolitan areas.[11–13] However, robust insights on the racial differences for SARS-CoV-2 infection are limited. Furthermore, comprehensive data evaluating higher susceptibility to SARS-CoV-2 infection among Hispanic communities are also scarce. This is perhaps because of comparatively homogenous populations in non-U.S. regions of the world. Houston, as an exceptionally ethnically diverse population center,[17] is well suited for an investigation of racial, ethnic, and socioeconomic gradients in COVID-19 test positivity. We focus on highlighting the mechanisms of racial and ethnic disparities in susceptibility to SARS-CoV-2 infection and provide evidence of mediation of such disparities by novel social determinants of health (SDoH).

Our study adds to the current literature by analyzing emerging data for individuals being tested across one of the largest healthcare systems in the Greater Houston area. We report that racial and ethnic minorities (non-Hispanic Black and Hispanic individuals) are almost twice as likely to test positive for SARS-CoV-2 than the non-Hispanic White and non-Hispanic population. These findings illuminate systematic racial / ethnic disparities in testing positive for SARS-CoV-2 infection. Though there are limited prior SARS-CoV-2 data, such racial and ethnic disparities have previously been described for the U.S. H1N1 influenza pandemic.[18] These data indicated that Spanish-speaking Hispanic and Black individuals were at a greater risk of H1N1 infection, primarily attributable to lack of healthcare access.

We explored three possible mechanisms of race disparities in our data. These included lower socio-economic status, residence in higher population dense areas, and higher level of comorbidities. We demonstrate that NHB race is significantly associated with all three potential disparity pathways, and in the traditional multivariable analyses, racial and ethnic disparities persisted even after controlling for these pathways. However, our mediation analyses highlighted the potential influence of residence in high population density areas as a viable pathway that at least partially explains the observed racial and ethnic disparity. Furthermore, residence in low income areas emerged as a significant mediation pathway for ethnic differences in SARS-CoV-2 positivity. Pathways mediating the influence of comorbidity status did not demonstrate a significant effect. We utilized population density as a marker for potential inability to maintain adequate social distancing as it has been indicated that maintaining the WHO recommended safe distance between people becomes challenging with high population densities.[19] Furthermore, overall effects of population density and disease spread has been previously described in literature.[20,21] In addition to lack of social distancing, higher population density may also be

associated with several other behavioral and socio-demographic attributes that may predispose populations to both viral spread and increased susceptibility. For example, there are reports linking obesity, lack of physical activity, and higher mortality with residence in densely populated neighborhoods.[22,23]

As reported, our data also corroborate that older populations may be more susceptible to SARS-CoV-2 infection.[10] However, younger populations still have cause for concern as nearly 1 in 4 of the infected cases in our sample were between 36–50 years of age. Finally, our data demonstrate that males may be approximately 20% more likely to test positive for the SARS-CoV-2 infection. Potential sex differences in infectivity to SARS-CoV-2 and intersectionality with racial and ethnic socioeconomic factors need to be explored further in future analyses. Additional policy-oriented research should prioritize studying the intersectionality of these vulnerable economic statuses and racial disparities in COVID infection indicated by the present study.

Findings of our study need to be interpreted in the light of certain limitations. Our data are from a single center and may not be generalizable to the wider U.S. population. These findings need to be replicated in larger data sets across other large heterogenous U.S. metropolitans. However, the Houston metropolitan area is one of the most diverse and representative in the U.S.[17] and our healthcare system is one of the largest systems providing care to COVID-19 patients in the Greater Houston area. Our sample was composed of 22% Black, 18% Hispanic, and 62% female population. Our final multivariable models included potential mediators which may produce biased estimates.[16] However, these potential mediators did not demonstrate a statistically significant indirect effect in our analyses. We did not have information on certain demographic covariates such as education or household size. Educational

status has been linked to healthcare awareness and may be important to adjust for in analyses of potential disparities, and household size may be used to provide more precise estimates of socio-economic status. However, we obtained and adjusted for zip code income data from the U.S. Census, as income has previously been shown to have strong correlation with educational attainment and socio-economic status.[24] Since testing was based on suspicion of infection and may have been influenced by factors such as access to care, the potential for selection bias cannot be ruled out. Furthermore, lack of sensitivity of SARS-CoV-2 diagnostics tests have been reported; however, the three assays utilized for testing were cross validated for internal consistency. Finally, we did not have detailed information on comorbidities and their management in the study population. However, we did control for major comorbidities which are being reported as associated with COVID-19 outcomes.[25]

Conclusions

The strong association between racial and ethnic minorities and SARS-CoV-2 infection demonstrated in our data, even after adjustment for other important socio-demographic and comorbidity factors, highlight a potential catastrophe of inequality within the existential crisis of a global pandemic. Our data, representing a large heterogeneous U.S. metropolitan area, also provide preliminary evidence into the potential pathways for this disparity. It is highly likely that higher comorbidity burden and detrimental effects of adverse social determinants, including those that may not adequately permit safe practices of social distancing, mediate higher SARS-CoV-2 infectivity among racial and ethnic minorities.

As the pandemic continues to spread and evolve across the continental U.S., emerging data on association between SARS-CoV-2 infection and various socio-demographic factors will

continue to enhance our understanding of targeted risks related to SARS-CoV-2 infection, and such data would enable us to comprehend healthcare services and access factors related to development and outcomes of COVID-19 among minority populations. Our findings substantiate prior calls for collection of robust data on race and ethnicity as a part of international collaborations,[26] and further drive home the critical importance of quantifying novel SDoH.

Acknowledgment: The authors thank Jacob M. Kolman, Senior Scientific Writer of the Houston Methodist Center for Outcomes Research, for reviewing the language and format of the manuscript.

IRB Approval: This work was carried out under an approved protocol for the Houston Methodist COVID-19 Surveillance and Outcomes Registry (HM CURATOR) by the Houston Methodist Research Institute Institutional Review Board (HMRI IRB).

Contribution Statement:

FV: design, data analysis and interpretation, drafting the manuscript, critical revision for important intellectual content, final approval

JCN: data acquisition, data analysis, drafting the manuscript, final approval

JRM: data acquisition, drafting the manuscript, final approval

OK: data acquisition, data analysis, drafting the manuscript, final approval

AP: data acquisition, data analysis, drafting the manuscript, final approval

SLJ: data acquisition, data interpretation, critical revision for important intellectual content, final approval

FNM: critical revision for important intellectual content, final approval

HDS: critical revision for important intellectual content, final approval

RAP: critical revision for important intellectual content, final approval

JDA: critical revision for important intellectual content, final approval

BK: critical revision for important intellectual content, final approval

KN: design, interpretation of data, critical revision for important intellectual content, final approval

Competing Interests: No authors declare a competing interest to this work

Funding: There are no funders to report for this submission.

Date sharing statement: All requests for de-identified data should be made to the corresponding author. All reasonable requests will be evaluated by the CURATOR Data Governance and Sharing Committee comprising of FV, SLJ, BK and KN in the light of institutional policies and guidelines.

References

1. World Health Organization. WHO COVID-19 dashboard. https://covid19.who.int/. Published April 6, 2020. Accessed April 6, 2020.

2. Center for Systems Science and Engineering (CSSE) at Johns Hopkins University. COVID-19 map. Johns Hopkins Coronavirus Resource Center. https://coronavirus.jhu.edu/map.html. Published April 6, 2020. Accessed April 6, 2020.

3. Centers for Disease Control and Prevention. Coronavirus Disease 2019 (COVID-19) in the U.S. cdc.gov. https://www.cdc.gov/coronavirus/2019-ncov/cases-updates/cases-in-us.html. Published April 14, 2020. Accessed April 14, 2020.

4. Texas Health and Human Services. COVID-19 in Texas (dashboard). https://txdshs.maps.arcgis.com/apps/opsdashboard/index.html#/ed483ecd702b4298ab01e8b9cafc8b83. Published April 18, 2020. Accessed April 18, 2020.

5. McCann A. Most diverse cities in the U.S. Wallethub. https://wallethub.com/edu/most-diverse-cities/12690/. Published April 10, 2019. Accessed June 29, 2020.

6. Mejia B. How Houston has become the most diverse place in America. *Los Angeles Times*. https://www.latimes.com/nation/la-na-houston-diversity-2017-htmlstory.html. Published May 9, 2017. Accessed June 29, 2020.

7. Shi S, Qin M, Shen B, et al. Association of cardiac injury with mortality in hospitalized patients with COVID-19 in Wuhan, China. *JAMA Cardiol*. March 2020. doi:10.1001/jamacardio.2020.0950

8. Bonow RO, Fonarow GC, O'Gara PT, Yancy CW. Association of coronavirus disease 2019 (COVID-19) with myocardial injury and mortality. *JAMA Cardiol*. March 2020. doi:10.1001/jamacardio.2020.1105

9. 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*. April 2020. doi:10.1001/jama.2020.5394

10. Zhou F, Yu T, Du R, et al. Clinical course and risk factors for mortality of adult inpatients with COVID-19 in Wuhan, China: a retrospective cohort study. *Lancet*. 2020;395(10229):1054-1062. doi:10.1016/S0140-6736(20)30566-3

11. Deslatte M. Louisiana data: virus hits blacks, people with hypertension. *AP News*. https://apnews.com/cde897293f88f788d3f3bfdb3d46c466. Published April 7, 2020. Accessed April 20, 2020.

12. Reyes C, Husain N, Gutowski C, St. Clair S, Pratt G. Chicago's coronavirus disparity: Black Chicagoans are dying at nearly six times the rate of white residents, data show. *Chicago Tribune*. https://www.chicagotribune.com/coronavirus/ct-coronavirus-chicago-coronavirus-

deaths-demographics-lightfoot-20200406-77nlylhiavgjzb2wa4ckivh7mu-story.html. Published April 7, 2020. Accessed April 20, 2020.

13. Thebault R, Tran AB, Williams V. The coronavirus is infecting and killing black Americans at an alarmingly high rate. *Washington Post*. https://www.washingtonpost.com/nation/2020/04/07/coronavirus-is-infecting-killing-black-americans-an-alarmingly-high-rate-post-analysis-shows/. Published April 7, 2020. Accessed April 20, 2020.

14. United States Census Bureau. American Community Survey (ACS). https://www.census.gov/programs-surveys/acs. Published 2018. Accessed April 14, 2020.

15. United States Census Bureau. Gazetteer Files. https://www.census.gov/geographies/reference-files/time-series/geo/gazetteer-files.html. Accessed April 20, 2020.

16. Richiardi L, Bellocco R, Zugna D. Mediation analysis in epidemiology: methods, interpretation and bias. *Int J Epidemiol*. 2013;42(5):1511-1519. doi:10.1093/ije/dyt127

17. Harris County Public Health. *Harris Cares: A 2020 Vision of Health in Harris County*. Houston, TX: Harris County Public Health; 2019. http://publichealth.harriscountytx.gov/Resources/Harris-Cares. Accessed November 21, 2019.

18. Quinn SC, Kumar S, Freimuth VS, Musa D, Casteneda-Angarita N, Kidwell K. Racial disparities in exposure, susceptibility, and access to health care in the US H1N1 influenza pandemic. *Am J Public Health*. 2011;101(2):285-293. doi:10.2105/AJPH.2009.188029

19. Rocklov J, Sjodin H. High population densities catalyze the spread of COVID-19. J Travel Med. 2020. doi: 10.1093/jtm/taaa038

20. Tarwater PM, Martin CF. Effects of population density on the spread of disease. *Complexity*. 2001;6(6):29-36. doi:10.1002/cplx.10003

21. Yang K, LeJeune J, Alsdorf D, Lu B, Shum CK, Liang S. Global distribution of outbreaks of water-associated infectious diseases. *PLoS Negl Trop Dis*. 2012;6(2):e1483. doi:10.1371/journal.pntd.0001483

22. Wang R, Feng Z, Xue D, Liu Y, Wu R. Exploring the links between population density, lifestyle, and being overweight: secondary data analyses of middle-aged and older Chinese adults. *Health Qual Life Outcomes*. 2019;17(1):100. doi:10.1186/s12955-019-1172-3

23. Beenackers MA, Oude Groeniger J, Kamphuis CBM, Van Lenthe FJ. Urban population density and mortality in a compact Dutch city: 23-year follow-up of the Dutch GLOBE study. *Health & Place*. 2018;53:79-85. doi:10.1016/j.healthplace.2018.06.010

24. U.S. Bureau of Labor Statistics. Unemployment rates and earnings by educational attainment. https://www.bls.gov/emp/chart-unemployment-earnings-education.htm. Published September 4, 2019. Accessed April 15, 2020.

25. Centers for Disease Control and Prevention. Groups at higher risk for severe illness. cdc.gov. https://www.cdc.gov/coronavirus/2019-ncov/need-extra-precautions/groups-at-higher-risk.html. Last updated April 17, 2020. Accessed April 20, 2020.

26. Abuelgasim E, Saw LJ, Shirke M, Zeinah M, Harky A. COVID-19: Unique public health issues facing Black, Asian and minority ethnic communities. *Curr Probl Cardiol*. 2020;100621. doi:10.1016/j.cpcardiol.2020.100621

Tables and Figures

Figure 1: Adjusted Probability and 95% Confidence Interval of Positive SARS-CoV-2 PCR in Non-Hispanic Black vs. Non-Hispanic White by increasing age

Figure 2: Adjusted Probability and 95% Confidence Interval of Positive SARS-CoV-2 PCR in Hispanic vs. Non-Hispanic by increasing age

Table 1: Summary measures and univariable association of socio-demographic characteristics with SARS-CoV-2 infection from HM CURATOR

Characteristics	Overall (n = 20,228)	SARS-CoV-2 Negative (n = 18,677)	SARS-CoV-2 Positive (n = 1,551)	OR (95% CI)[a]
Age, mean (SD)	51.1 (19.0)	51.0 (19.1)	52.1 (18.1)	1.00 (1.00–1.01)
Age Categories (%)				
Up to 35 years	25.1	25.5	20.9	Reference Category
36–50 years	25.0	24.9	27.0	1.32 (1.14 – 1.54)
51–75 years	38.7	38.5	40.8	1.29 (1.12 – 1.48)
>75 years	11.2	11.2	11.4	1.23 (1.02 – 1.49)
Females (%)	61.9	62.3	58.3	0.85 (0.76 – 0.94)
Race (%)[b]				
White	62.3	62.9	55.7	Reference Category
Black	21.6	21.0	28.8	1.55 (1.37 – 1.75)
Asian	9.2	9.1	9.7	1.20 (1.00 – 1.44)
Mixed / Other	1.3	1.2	1.5	1.36 (0.88 – 2.11)
Hispanic (%)[c]	17.8	16.8	29.2	2.02 (1.79 – 2.27)
Median Zip Household Income (IQR)[d]	70,658 (53,313–99,276)	70,758 (53,633–100,107)	66,523 (50,485–94,226)	-4,168[e] (-6463.2, -1872.8)
Median Zip Household Income Pentiles (%)				
I: 13,893 – 50,485	18.2	17.8	23.6	Reference Category
II: 50,642 – 65,805	18.4	18.1	21.2	0.89 (0.75 – 1.04)
III: 65,897 – 79,869	18.3	18.2	19.6	0.82 (0.70 – 0.96)
IV: 80,039 – 106,067	18.6	19.0	13.7	0.55 (0.46 – 0.65)
V: 106,415 – 240,417	17.2	17.3	16.1	0.71 (0.60 – 0.83)
Median (IQR) Population Density[f]	2797.2 (1439.1 – 4260.9)	2797.2 (1439.1 – 4211.4)	3320.3 (1904.4 – 4439.7)	523.1[e] (454.9 – 591.3)
Median Population Density Pentiles (%)				
I: 1.5 – 1026.6	18.2	18.6	13.5	Reference Category
II: 1034.6 – 2306.3	19.0	19.1	18.1	1.31 (1.09 – 1.58)
III: 2330.8 – 3328.9	17.4	17.4	18.3	1.45 (1.21 – 1.75)
IV: 3360.1 – 4665.6	18.1	17.8	22.1	1.71 (1.43 – 2.05)
V: 4742.6 – 98025.9	18.1	17.7	22.6	1.76 (1.47 – 2.10)
Insurance Status (%)				
Medicare	29.0	29.1	27.2	Reference Category
Medicaid	4.7	4.8	4.3	0.96 (0.73 – 1.26)
Pvt / Employer based	42.6	43.2	36.4	0.90 (0.79 – 1.03)
HC Exchange	1.7	1.6	3.0	2.02 (1.46 – 2.80)
Self-Pay	20.6	20.0	28.4	1.52 (1.33 – 1.75)
VA	1.3	1.4	0.7	0.54 (0.29 – 1.00)
Charlson Co-morbidity Index, Median (IQR)	2 (0 – 6)	2 (0 – 6)	2 (0 – 5)	0 (-0.36, 0.36)[f]
Charlson Co-morbidity Index (CCI) Categories (%)				
CCI: 0	33.1	33.4	30.4	Reference Category

Characteristics	Overall (n = 20,228)	SARS-CoV-2 Negative (n = 18,677)	SARS-CoV-2 Positive (n = 1,551)	OR (95% CI)[a]
CCI: 1 – 2	23.7	23.1	28.6	1.35 (1.18 – 1.54)
CCI: 3 – 6	20.3	20.2	20.4	0.10 (0.65 – 1.28)
CCI: > 6	22.9	23.1	20.6	0.98 (0.84 – 1.13)
Hypertension	47.2	47.1	48.4	1.06 (0.95 – 1.17)
Diabetes (without complications)	24.2	23.7	30.3	1.40 (1.24– 1.57)
Obesity	28.0	28.2	25.2	0.86 (0.76 – 0.96)

[a] Unadjusted Odds Ratios and 95% Confidence Intervals for association between individual co-variates in SRAS-CoV-2 positivity.
[b] Race: Missing, Unknown, Declined, n = 1157 (5.7%).
[c] Ethnicity: Missing, Unknown, Declined, n = 613 (3.0%).
[d] 2018 inflation adjusted USD. Missing n = 1,883 (9.3%). Pentiles were defined by categorizing the ordered distribution of median income into five categories.
[e] Difference in median and 95% CI of difference obtained via quantile regression.
[f] Population density Missing n = 1,854 (9.2%). Pentiles were defined by categorizing the ordered distribution of population density into five categories.
HM CURATOR: Houston Methodist, COVID-19 Surveillance and Outcomes Registry
SARS-CoV-2: Severe Acute Respiratory Syndrome Coronavirus 2

Table 2: Univariable comparison of socio-demographic and comorbidity factors between non-Hispanic Black (NHB) and non-Hispanic White (NHW) race categories

	Non-Hispanic Black n = 4,285	Non-Hispanic White n = 9,469	OR[a] / Median Difference (95% CI)	P value
Age: Mean (SD)	49.4 (17.8)	56.0 (19.2)	0.98 (0.98 – 0.98)	< 0.001
Age Category (%)				
Up to 35	24.3	18.8	Reference	
36 – 50	28.8	19.8	1.12 (1.01 – 1.25)	0.03
51 – 75	39.1	44.9	0.68 (0.61 – 0.74)	< 0.001
> 75	7.8	16.5	0.37 (0.32 – 0.42)	< 0.001
Females	68.1	58.1	1.54 (1.42 – 1.66)	< 0.001
Median (IQR) Zip Income	64,022 (47,303 – 79,658)	76,163 (60,130 – 102,019)	-12,141 (-14,018, -10,263)	< 0.001
Median Zip Income Pentiles (%) – Pentiles of increasing Income[b]				
Category I	33.5	13.1	Reference	
Category II	21.3	19.7	0.42 (0.38 – 0.47)	< 0.001
Category III	22.7	19.1	0.47 (0.42 – 0.52)	< 0.001
Category IV	10.5	26.2	0.16 (0.14 – 0.18)	< 0.001
Category V	11.9	22.0	0.21 (0.19 – 0.24)	< 0.001
Population Density for Zip: Median (IQR)	3217.6 (2040.7 – 4439.7)	2488.5 (812.2 – 4084.3)	729.1 (603.0 – 855.2)	< 0.001
Population Density for Zip Pentiles (%) – Pentiles of increasing population density[c]				
Category I	11.5	28.3	Reference	
Category II	21.7	20.9	2.52 (2.21 – 2.86)	< 0.001
Category III	21.8	16.9	3.20 (2.81 – 3.65)	< 0.001
Category IV	23.0	15.5	3.66 (3.21 – 4.17)	< 0.001
Category V	22.2	18.3	2.99 (2.63 – 3.41)	< 0.001
Charlson Comorbidity Index, median (IQR)	2 (0 – 6)	3 (1 – 7)	-1 (-1.22, -0.78)	< 0.001
Charlson Comorbidity Index (CCI) Categories (%)				
CCI: 0	30.8	24.9	Reference	
CCI: 1 – 2	25.1	22.3	0.91 (0.82 – 1.01)	0.07
CCI: 3 – 6	19.7	24.0	0.66 (0.60 – 0.74)	< 0.001
CCI: > 6	24.4	28.8	0.68 (0.62 – 0.75)	< 0.001
Hypertension	56.0	52.6	1.14 (1.07 – 1.23)	< 0.001
Diabetes (without complications)	29.6	21.9	1.50 (1.38 – 1.63)	< 0.001
Obesity	33.1	30.8	1.11 (1.03 – 1.20)	0.008

[a] *Unadjusted Odds Ratios and 95% Confidence Intervals for association between individual co-variates in SRAS-CoV-2 positivity or Difference in median and 95% CI of difference obtained via quantile regression.*

[b,c] *Pentiles were defined by categorizing the ordered distribution of median income and population density into five categories.*

Table 3: Univariable comparison of socio-demographic and comorbidity factors between Hispanic and non-Hispanic ethnicities

	Hispanic n = 3,590	Non-Hispanic n = 16,025	OR[a] / Median Difference[a] (95% CI)	P value
Age: Mean (SD)	45.1 (18.3)	52.8 (18.9)	0.98 (0.98 – 0.98)	< 0.001
Age Category (%)				
Up to 35	36.0	21.8	Reference	
36 – 50	26.1	24.6	0.64 (0.58 – 0.71)	< 0.001
51 – 75	32.0	40.1	0.47 (0.43 – 0.52)	< 0.001
> 75	5.9	12.7	0.28 (0.24 – 0.36)	< 0.001
Females	64.5	61.4	1.14 (1.06 – 1.23)	0.001
Median (IQR) Zip Income	65,742 (48,345 – 82,708)	73,742 (56,288 – 102,008)	-8,000 (-9,450.5, -6,549.5)	< 0.001
Median Zip Income Pentiles (%) – Pentiles of increasing Income[b]				
Category I	29.2	18.2	Reference	
Category II	23.3	19.6	0.74 (0.66 – 0.83)	< 0.001
Category III	21.8	19.8	0.68 (0.61 – 0.76)	< 0.001
Category IV	15.0	21.6	0.43 (0.38 – 0.49)	< 0.001
Category V	10.6	20.8	0.32 (0.28 – 0.36)	< 0.001
Population Density for Zip: Median (IQR)	3256.8 (1504.0 – 4299.7)	2741.6 (1408.1 – 4110.7)	515.2 (469.2 – 561.2)	< 0.001
Population Density for Zip Pentiles (%) – Pentiles of increasing population density[c]				
Category I	15.5	21.2	Reference	
Category II	21.1	20.9	1.39 (1.22 – 1.57)	< 0.001
Category III	15.3	20.0	1.05 (0.92 – 1.20)	0.48
Category IV	27.0	18.3	2.02 (1.79 – 2.28)	< 0.001
Category V	21.1	19.5	1.48 (1.31 – 1.68)	< 0.001
Charlson Comorbidity Index, median (IQR)	1 (0 – 4)	2 (0 – 6)	-1 (-1.21, -0.79)	< 0.001
Charlson Comorbidity Index (CCI) Categories (%)				
CCI: 0	41.8	29.7	Reference	
CCI: 1 – 2	25.4	23.6	0.76 (0.70 – 0.84)	< 0.001
CCI: 3 – 6	15.9	21.7	0.52 (0.47 – 0.58)	< 0.001
CCI: > 6	16.9	24.9	0.48 (0.43 – 0.54)	< 0.001
Hypertension	38.0	50.5	0.60 (0.56 – 0.65)	< 0.001
Diabetes (without complications)	27.4	23.6	1.22 (1.12 – 1.33)	< 0.001
Obesity	27.3	28.9	0.92 (0.85 – 1.00)	0.06

[a] *Unadjusted Odds Ratios and 95% Confidence Intervals for association between individual co-variates in SRAS-CoV-2 positivity or Difference in median and 95% CI of difference obtained via quantile regression.*
[b,c] *Pentiles were defined by categorizing the ordered distribution of median income and population density into five categories.*

Table 4: Adjusted Odds Ratios and 95% Confidence Intervals for likelihood of SARS-CoV-2 positivity among minority race and ethnic groups

Covariate	NHB vs. NHW, aOR (95% CI)[a]	Hispanic vs. Non-Hispanic, aOR (95% CI)[b]
Non-Hispanic Black (vs. Non-Hispanic White)	**2.23 (1.90 – 2.60)[c]**	
Hispanic (vs. Non-Hispanic)		**1.95 (1.72 – 2.20)[c]**
Age Categories		
Up to 35 years	*Reference Category*	
36–50 years	1.36 (1.07 – 1.72)	1.42 (1.20 – 1.68)
51–75 years	1.60 (1.21 – 2.11)	1.71 (1.39 – 2.11)
>75 years	2.20 (1.52 – 3.19)	2.08 (1.56 – 2.77)
Male (vs. Female)	1.20 (1.04 – 1.39)	1.17 (1.05 – 1.32)
Median Zip Household Income Categories (Pentiles of Increasing Income)		
Category I	*Reference Category*	
Category II	0.95 (0.76 – 1.18)	
Category III	1.02 (0.84 – 1.26)	
Category IV	0.74 (0.58 – 0.94)	
Category V	0.97 (0.77 – 1.22)	
Primary Insurance Type		
Medicare	*Reference Category*	
Medicaid	0.86 (0.55 – 1.33)	1.01 (0.74 – 1.37)
Private / Employer Based	1.11 (0.88 – 1.40)	0.93 (0.80 – 1.12)
Healthcare Exchange	2.06 (1.25 – 3.40)	1.80 (1.27 – 2.54)
Self-Pay	1.75 (1.33 – 2.30)	1.80 (1.27 – 2.54)
Veterans Affairs	0.72 (0.36 – 1.43)	0.58 (0.31 – 1.07)
Charlson Comorbidity Index Categories		
CCI: 0	*Reference Category*	
CCI: 1 – 2	1.26 (0.99 – 1.60)	1.03 (0.87 – 1.23)
CCI: 3 – 6	0.94 (0.68 – 1.28)	0.71 (0.56 – 0.91)
CCI: > 6	0.91 (0.65 – 1.30)	0.58 (0.44 – 0.76)
Hypertension	0.87 (0.73 – 1.05)	1.00 (0.87 – 1.15)
Diabetes (Without Complications)	1.42 (1.17 – 1.71)	1.62 (1.40 – 1.87)
Obesity	0.82 (0.70 – 0.97)	0.83 (0.73 – 0.94)

[a,b] Hosmer and Lemeshow goodness of fit p-value: 0.58 and 0.73 (H_0: Model fit is correct).

[c] ORs (95% CIs) represent the direct association adjusted for all other covariates in the model.

NHB: Non-Hispanic Black, NHW: Non-Hispanic White. Grayed out cells were not included in respective models either due to collinearity or demonstration of statistically significant mediation towards likelihood of SARS-CoV-2 positivity.

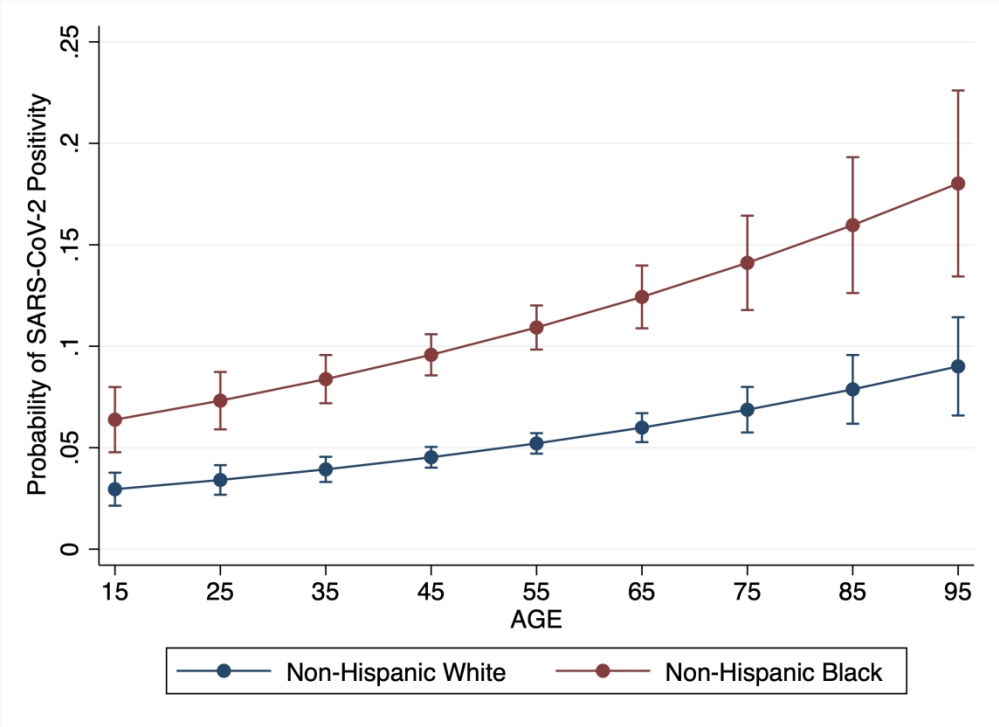


Figure 1: Adjusted Probability and 95% Confidence Interval of Positive SARS-CoV-2 PCR in Non-Hispanic Black vs. Non-Hispanic White by increasing age

467x339mm (144 x 144 DPI)

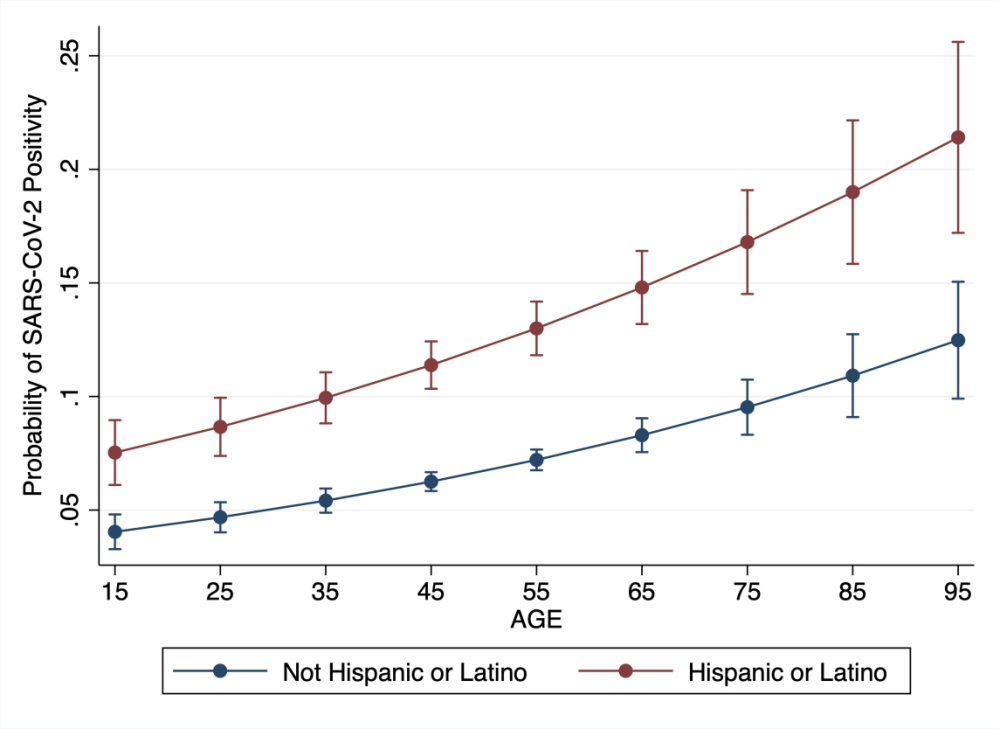


Figure 2: Adjusted Probability and 95% Confidence Interval of Positive SARS-CoV-2 PCR in Hispanic vs. Non-Hispanic by increasing age

467x339mm (144 x 144 DPI)

STROBE Statement—Checklist of items that should be included in reports of ***cross-sectional studies***

	Item No	Recommendation	Page No
Title and abstract	1	(*a*) Indicate the study's design with a commonly used term in the title or the abstract	2
		(*b*) Provide in the abstract an informative and balanced summary of what was done and what was found	2-3
Introduction			
Background/rationale	2	Explain the scientific background and rationale for the investigation being reported	4
Objectives	3	State specific objectives, including any prespecified hypotheses	5
Methods			
Study design	4	Present key elements of study design early in the paper	5
Setting	5	Describe the setting, locations, and relevant dates, including periods of recruitment, exposure, follow-up, and data collection	5-6
Participants	6	(*a*) Give the eligibility criteria, and the sources and methods of selection of participants	6
Variables	7	Clearly define all outcomes, exposures, predictors, potential confounders, and effect modifiers. Give diagnostic criteria, if applicable	6-7
Data sources/ measurement	8*	For each variable of interest, give sources of data and details of methods of assessment (measurement). Describe comparability of assessment methods if there is more than one group	6-7
Bias	9	Describe any efforts to address potential sources of bias	N/A
Study size	10	Explain how the study size was arrived at	8
Quantitative variables	11	Explain how quantitative variables were handled in the analyses. If applicable, describe which groupings were chosen and why	6-7
Statistical methods	12	(*a*) Describe all statistical methods, including those used to control for confounding	7
		(*b*) Describe any methods used to examine subgroups and interactions	7-8
		(*c*) Explain how missing data were addressed	N/A
		(*d*) If applicable, describe analytical methods taking account of sampling strategy	N/A
		(*e*) Describe any sensitivity analyses	N/A
Results			
Participants	13*	(a) Report numbers of individuals at each stage of study—eg numbers potentially eligible, examined for eligibility, confirmed eligible, included in the study, completing follow-up, and analysed	8
		(b) Give reasons for non-participation at each stage	N/A
		(c) Consider use of a flow diagram	N/A
Descriptive data	14*	(a) Give characteristics of study participants (eg demographic, clinical, social) and information on exposures and potential confounders	8, 21-22
		(b) Indicate number of participants with missing data for each variable of interest	21-22
Outcome data	15*	Report numbers of outcome events or summary measures	8
Main results	16	(*a*) Give unadjusted estimates and, if applicable, confounder-adjusted estimates and their precision (eg, 95% confidence interval). Make clear which confounders were adjusted for and why they were included	8-9, 10, 21-22

		(*b*) Report category boundaries when continuous variables were categorized	21-22
		(*c*) If relevant, consider translating estimates of relative risk into absolute risk for a meaningful time period	N/A
Other analyses	17	Report other analyses done—eg analyses of subgroups and interactions, and sensitivity analyses	9-10. 23-25
Discussion			
Key results	18	Summarise key results with reference to study objectives	11
Limitations	19	Discuss limitations of the study, taking into account sources of potential bias or imprecision. Discuss both direction and magnitude of any potential bias	13-14
Interpretation	20	Give a cautious overall interpretation of results considering objectives, limitations, multiplicity of analyses, results from similar studies, and other relevant evidence	12,14
Generalisability	21	Discuss the generalisability (external validity) of the study results	13
Other information			
Funding	22	Give the source of funding and the role of the funders for the present study and, if applicable, for the original study on which the present article is based	15

*Give information separately for exposed and unexposed groups.

Note: An Explanation and Elaboration article discusses each checklist item and gives methodological background and published examples of transparent reporting. The STROBE checklist is best used in conjunction with this article (freely available on the Web sites of PLoS Medicine at http://www.plosmedicine.org/, Annals of Internal Medicine at http://www.annals.org/, and Epidemiology at http://www.epidem.com/). Information on the STROBE Initiative is available at www.strobe-statement.org.