

# BMJ Open

BMJ Open is committed to open peer review. As part of this commitment we make the peer review history of every article we publish publicly available.

When an article is published we post the peer reviewers' comments and the authors' responses online. We also post the versions of the paper that were used during peer review. These are the versions that the peer review comments apply to.

The versions of the paper that follow are the versions that were submitted during the peer review process. They are not the versions of record or the final published versions. They should not be cited or distributed as the published version of this manuscript.

BMJ Open is an open access journal and the full, final, typeset and author-corrected version of record of the manuscript is available on our site with no access controls, subscription charges or pay-per-view fees (<http://bmjopen.bmj.com>).

If you have any questions on BMJ Open's open peer review process please email [info.bmjopen@bmj.com](mailto:info.bmjopen@bmj.com)

# 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:	<i>BMJ Open</i>
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, Bitu 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

SCHOLARONE™  
Manuscripts



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.

1  
2  
3 **Racial and Ethnic Disparities in SARS-CoV-2 Pandemic: Analysis of a COVID-19**  
4 **Observational Registry for a Diverse U.S. Metropolitan Population**  
5  
6  
7

8 *Farhaan S. Vahidy PhD MBBS MPH,<sup>1,2</sup> Juan Carlos Nicolas BS,<sup>1</sup> Jennifer R. Meeks MS,<sup>1</sup>*  
9 *Osman Khan BS,<sup>1</sup> Stephen L. Jones MD MSHI,<sup>1,3,4</sup> Faisal N. Masud MD,<sup>4,5</sup>*  
10 *H. Dirk Sostman MD,<sup>4,6</sup> Robert A. Phillips MD PhD,<sup>1,4,7</sup> Julia D. Andrieni MD,<sup>4,8</sup>*  
11 *Bitu Kash PhD MBA,<sup>1,4,9</sup> Khurram Nasir MD MPH<sup>1,7</sup>*  
12  
13  
14  
15

- 16 1. Center for Outcomes Research, The Houston Methodist Research Institute, Houston TX
- 17 2. Houston Methodist Neurological Institute, Houston Methodist Hospital, Houston TX
- 18 3. Department of Surgery, Houston Methodist Hospital, Houston TX
- 19 4. Weill Cornell Medical College, New York NY
- 20 5. Department of Anesthesiology and Critical Care, Houston Methodist Hospital, Houston
- 21 TX
- 22 6. The Houston Methodist Research Institute and The Houston Methodist Academic
- 23 Institute, Houston TX
- 24 7. Department of Cardiology, DeBakey Heart and Vascular Center, Houston Methodist
- 25 Hospital, Houston TX
- 26 8. Department of Medicine, Houston Methodist Hospital, Houston TX
- 27 9. Department of Health Policy and Management, Texas A&M University School of Public
- 28 Health, College Station TX
- 29
- 30
- 31
- 32
- 33
- 34
- 35
- 36
- 37
- 38
- 39

40 **Corresponding Author**

41  
42 Farhaan S. Vahidy, PhD MBBS MPH  
43 Josie Roberts Building  
44 Houston Methodist  
45 7550 Greenbriar Drive  
46 Houston TX 77030  
47  
48  
49  
50  
51 [fvahidy@houstonmethodist.org](mailto:fvahidy@houstonmethodist.org)  
52  
53  
54

55 Word Count: 2,735  
56  
57  
58  
59  
60

## 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

1  
2  
3 of African American race on SARS-CoV-2 infection mediated via a pathway that included  
4  
5 residence in densely populated zip code.  
6

7  
8 **Conclusions:** There is strong evidence of race and ethnic disparities in the SARS-CoV-2  
9  
10 pandemic potentially mediated through unique social determinants of health.  
11  
12

### 13 14 15 **Strengths and limitations of this study**

- 16  
17 • One of the first studies to systematically evaluate race and ethnic disparities in  
18  
19 susceptibility to SARS-CoV-2 infection, while accounting for multiple socio-  
20  
21 demographic characteristics and comorbidities  
22
- 23  
24 • Study population represents a large and diverse metropolitan of the U.S. with data from  
25  
26 one of the largest healthcare providers across the greater metropolitan area  
27
- 28  
29 • Study evaluates potential mediation pathways for race disparities and demonstrates that  
30  
31 residence in areas with high population density may mediate race disparities in  
32  
33 susceptibility to SARS-CoV-2 infection  
34
- 35  
36 • Single center study with limited information about true burden of comorbidity and  
37  
38 lifestyle factors  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

## 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.<sup>1,2</sup> 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.<sup>3</sup> 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.<sup>4</sup> 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.<sup>5-8</sup> 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.<sup>9-11</sup> 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



1  
2  
3 infection among racial and ethnic minorities across diverse U.S. metropolitan areas outside of  
4 New York are limited. Furthermore, the mediators of SARS-CoV-2 infection among racial and  
5  
6 ethnic minorities have not been described.  
7  
8  
9

10 We explored socio-demographic characteristics such as age, sex, race, ethnicity, median  
11 household zip code income, population density of residents' zip codes, and health insurance  
12 status associated with positive SARS-CoV-2 testing in an urban and diverse population served  
13 by one of the leading healthcare systems of the greater Houston area. We further examined the  
14 association between pre-existing comorbidities and higher likelihood of SARS-CoV-2 infection  
15 in our study population. We hypothesized that older age, non-white race and ethnic minority  
16 status will be associated with significantly higher likelihood of SARS-CoV-2 infection, and  
17 factors such as low socio-economic status, residence in high population density areas (proxy for  
18 potential difficulties in social distancing) and higher comorbidity burden will mediate the effect  
19 of race on SARS-CoV-2 infection.  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32

### 33 **Methods**

34  
35 We analyzed data being contemporaneously collected since March 5, 2020 as a part of  
36 the COVID-19 Surveillance and Outcomes Registry (CURATOR) at the Houston Methodist  
37 Hospital system (HM). The Houston Methodist CURATOR has been approved by the HM  
38 Institutional Review Board (IRB) as an observational quality of care registry for all suspected  
39 and confirmed COVID-19 patients. CURATOR is populated from multiple data sources across  
40 the HM system such as electronic medical records, electronic databanks for laboratory and  
41 pharmacy, and electronic interactive patient interface tools. The HM system comprises a flagship  
42 tertiary care hospital in the Texas Medical Center, seven large community hospitals, a continuing  
43 care hospital, and multiple emergency centers and clinics throughout the Greater Houston area.  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

1  
2  
3 Data from various sources are curated into a harmonized format, assessed for quality and  
4 integrity, and stored on a secure institutional HIPAA-compliant server.  
5  
6

7  
8 We flagged all individuals who were tested for the SARS-CoV-2 using the real time  
9  
10 Reverse Transcriptase (RT) Polymerized Chain Reaction (PCR) diagnostic panels. These assays  
11  
12 were verified for quantitative detection of novel SARS-CoV-2 isolated and purified from  
13  
14 nasopharyngeal swab specimens obtained from individuals and immersed in universal transport  
15  
16 medium. Testing was carried out for symptomatic individuals or for individuals who had a self-  
17  
18 reported history of exposure to a COVID-19 case including recent travel to other countries with  
19  
20 high infection rates or hotspots within the U.S. Socio-demographic characteristics including age,  
21  
22 sex, race, ethnicity, and payer-status (insurance type) were obtained from the HM CURATOR  
23  
24 for analyses. We utilized the U.S. Census Bureau's American Community Survey (ACS) 5-year  
25  
26 data (2014–2018) to determine median household income by individual zip code tabulation areas  
27  
28 (ZCTA).<sup>12</sup> The median ZCTA household income was inflation-adjusted to 2018 USD. We also  
29  
30 utilized the same data source to obtain population estimates by ZCTA, and calculated ZCTA  
31  
32 level population density (population per mile square) by standardizing it for area measurements  
33  
34 of ZCTA. For the purpose of population density determination, land area estimates were  
35  
36 obtained from the Census Bureau's U.S. Gazetteer Files 2010.<sup>13</sup> In the absence of granular and  
37  
38 precise social distancing data, we have utilized population density as a proxy for potential  
39  
40 difficulties in social distancing among crowded communities.  
41  
42  
43  
44  
45

46  
47 We provide descriptive summary data as means (standard deviations) and proportions.  
48  
49 We fit univariable and multivariable logistic regression models to assess unadjusted and adjusted  
50  
51 association between socio-demographic characteristics and likelihood of being tested positive for  
52  
53 the SARS-CoV-2. We determined *a priori* to include all variables (age, sex, race, ethnicity, zip  
54  
55  
56  
57  
58  
59

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 percentiles 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

1  
2  
3 status, resided in ZCTAs with higher population density, and had high comorbidity burden for  
4 hypertension, diabetes, obesity and CAD / MI / CHF. Table 2 provides univariable comparison  
5 of African Americans vs. Non-African Americans across various socio-demographic and  
6 comorbidity characteristics.  
7  
8  
9  
10  
11

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

13

14 The significantly higher likelihood of SARS-CoV-2 infection among African Americans  
15 (compared to White) persisted after controlling for other demographics, insurance type, median  
16 household income, population density, and comorbidities. Adjusted odds ratios (CI) for African  
17 American vs. White: 1.84 (1.49–2.27). Our fully adjusted model estimated that Asians (vs.  
18 White) were also at a significantly higher risk of SARS-CoV-2 infection (aOR, CI: 1.46, 1.09–  
19 1.95). Furthermore, we also observed a statistically significant association between SARS-CoV-  
20 2 infection and Hispanic ethnicity, aOR (CI): 1.70 (1.35–2.14). Higher risk of infection among  
21 males (compared to females) and higher likelihood of SARS-CoV-2 infection among elderly also  
22 remained statistically significant. Detailed output of the fully adjusted logistic regression model  
23 is presented in Table 3. The influence of African American race (vs. White) and Hispanic (vs.  
24 Non-Hispanic) ethnicity was observed uniformly across the age spectrum of 10 – 80 years. In  
25 other words, we did not observe effect modification by age for relationship between race /  
26 ethnicity and SARS-CoV-2 infection. However older age in itself remains significantly  
27 associated with higher likelihood of SARS-CoV-2 infection. Based on the marginal probabilities  
28 obtained from our fully adjusted model, the probability of SARS-CoV-2 infection in a 40-year-  
29 old African American is 18.8% whereas it is 12.9% in a 40-year-old White individual, all other  
30 adjusted variables being constant. At the age of 75 this probability is 29.0% for an African  
31 American, and 20.8% for a Caucasian. A similar relationship differential was observed for  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

1  
2  
3 Hispanic vs. non-Hispanic. Probability of SARS-CoV-2 infection for African American vs.  
4 White, and for Hispanic vs. Non-Hispanic across age spectrum is presented in Figure 2 and  
5  
6  
7  
8 Figure 3.

9  
10 *Generalized Structural Equation Modeling for Mediation by Income, Population Density and*  
11  
12 *Comorbidity*

13  
14  
15 Utilizing the GSEM framework, we determined the direct and indirect effects of African  
16  
17 American race on SARS-CoV-2 infection with median income, population density and  
18  
19 comorbidity score modeled as mediators in three separate equations. The indirect effect of  
20  
21 African American race mediated through population density was statistically significant ( $p =$   
22  
23 0.008); however, the indirect effects mediated via median income and comorbidity scores were  
24  
25 not statistically significant ( $p = 0.31$  and  $p = 0.38$  respectively).  
26  
27  
28

29  
30 **Discussion**

31  
32 There is emerging evidence of race disparities in the evolving COVID-19 pandemic  
33  
34 across the continental U.S. Most reports indicate higher case fatality among African Americans  
35  
36 across major U.S. metropolitan areas.<sup>9–11</sup> However, robust insights on the racial and ethnic  
37  
38 differences for SARS-CoV-2 infection are limited. This is perhaps because of comparatively  
39  
40 homogenous populations in non-U.S. regions of the world. Houston, as an exceptionally  
41  
42 ethnically diverse population center,<sup>14</sup> is well suited for an investigation of racial, ethnic, and  
43  
44 socioeconomic gradients in COVID-19 test positivity.  
45  
46  
47

48  
49 Our study adds to the current literature by analyzing emerging data for individuals being  
50  
51 tested across one of the largest healthcare systems in the Greater Houston area; we report that  
52  
53 racial minorities (non-Hispanic African American and Asian) are approximately 50–80% more  
54  
55 likely to test positive for SARS-CoV-2 than the non-Hispanic White population. Our data also  
56  
57  
58  
59  
60

1  
2  
3 indicate that the Hispanic population is almost 70% more likely than non-Hispanics to be  
4 susceptible to SARS-CoV-2 infection. These findings illuminate systematic racial / ethnic  
5  
6 disparities in testing positive for SARS-CoV-2 infection. Though there are limited prior SARS-  
7  
8 CoV-2 data, such race and ethnic disparities have previously been described for the U.S. H1N1  
9  
10 influenza pandemic.<sup>15</sup> These data indicated that Spanish-speaking Hispanics were at a greater  
11  
12 risk of H1N1 infection primarily attributable to lack of healthcare access. Black people were also  
13  
14 more susceptible to complications of H1N1 infection.<sup>15</sup>  
15  
16  
17  
18

19 We explored three possible mechanisms of race disparities in our data. These included  
20  
21 lower socio-economic status, residence in higher population dense areas, and higher level of  
22  
23 comorbidities. We demonstrate that African American race is significantly associated with all  
24  
25 three potential disparity pathways, and in the traditional multivariable analyses, race and ethnic  
26  
27 disparities persisted even after controlling for these pathways. However, our mediation analyses  
28  
29 highlighted the potential influence of residence in high population density areas as a viable  
30  
31 pathway that at least partially explains race disparity. Pathways mediating the influence of  
32  
33 median income and comorbidity status did not demonstrate a significant effect. We utilized  
34  
35 population density as a marker for potential inability to maintain adequate social distancing as it  
36  
37 has been indicated that maintaining the WHO recommended safe distance between people  
38  
39 becomes challenging with high population densities.<sup>16</sup> Furthermore, overall effects of population  
40  
41 density and disease spread has been previously described in literature.<sup>17,18</sup> In addition to lack of  
42  
43 social distancing, higher population density may also be associated with several other behavioral  
44  
45 and socio-demographic attributes that may predispose to both viral spread and increased  
46  
47 susceptibility. For example, there are reports linking obesity, lack of physical activity, and higher  
48  
49 mortality with residence in densely populated neighborhoods.<sup>19,20</sup>  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

1  
2  
3 As reported, our data also corroborate that older populations may be more susceptible to  
4 SARS-CoV-2 infection;<sup>8</sup> however, younger populations still have cause for concern, as nearly 1  
5 in 4 of the infected cases in our sample were between 36–50 years. Finally, our data demonstrate  
6 that males may be approximately 30% more likely to test positive for the SARS-CoV-2  
7 infection. Potential sex differences in infectivity to SARS-CoV-2 and intersectionality with racial  
8 and ethnic socioeconomic factors need to be explored further in future analyses. Additional  
9 policy-oriented research should prioritize study on the intersectionality of these vulnerable  
10 economic statuses and racial disparities in COVID infection indicated by the present study.  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20

21 Findings of our study need to be interpreted in the light of certain limitations. First, our  
22 data are from a single center and may not be generalizable to the wider U.S. population. These  
23 findings need to be replicated in larger data sets across other large heterogenous U.S.  
24 metropolitans. However, the Houston metropolitan area is one of the most diverse and  
25 representative in the U.S.,<sup>14</sup> and our healthcare system is one of the largest systems providing  
26 care to COVID-19 patients in the Greater Houston area. Our sample was composed of 26%  
27 Black, 19% Hispanic, and 62% female population. Second, we did not have information on  
28 certain demographic covariates such as education. Educational status has been linked to  
29 healthcare awareness and may be important to adjust for in analyses of potential disparities.  
30 However, we obtained and adjusted for zip code income data from the U.S. Census, as income  
31 has previously been shown to have strong correlation with educational attainment.<sup>21</sup> Third, since  
32 testing was based on suspicion of infection and may have been influenced by factors such as  
33 access to care, the potential for selection bias cannot be ruled out. Finally, we did not have  
34 detailed information on comorbidities and their management in the study population. However,  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60



1  
2  
3 we did control for major comorbidities which are being reported as associated with COVID-19  
4  
5 outcomes.<sup>22</sup>  
6  
7  
8  
9

## 10 **Conclusions**

11  
12 The strong association between racial and ethnic minorities and SARS-CoV-2 infection  
13  
14 demonstrated in our data, even after adjustment for other important socio-demographic and  
15  
16 comorbidity factors, highlight a potential catastrophe of inequality within the existential crisis of  
17  
18 a global pandemic. Our data, representing a large heterogeneous U.S. metropolitan area, also  
19  
20 provide preliminary evidence into the potential pathways for this disparity. It is highly likely that  
21  
22 higher comorbidity burden and detrimental effects of adverse social determinants, including  
23  
24 those that may not adequately permit safe practices of social distancing, mediate higher SARS-  
25  
26 CoV-2 infectivity among racial and ethnic minorities.  
27  
28  
29

30  
31 As the pandemic continues to spread and evolve across the continental U.S., emerging  
32  
33 data on association between SARS-CoV-2 infection and various socio-demographic factors will  
34  
35 continue to enhance our understanding of targeted risks related to SARS-CoV-2 infection, and  
36  
37 such data would enable us to comprehend healthcare services and access factors related to  
38  
39 development and outcomes of COVID-19 among minority populations.  
40  
41  
42  
43

44 **Acknowledgment:** The authors thank Jacob M. Kolman, Senior Scientific Writer of the Houston  
45  
46 Methodist Center for Outcomes Research, for reviewing the language and format of the  
47  
48 manuscript.  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

1  
2  
3 **IRB Approval:** This work was carried out under an approved protocol for the Houston  
4  
5 Methodist COVID-19 Surveillance and Outcomes Registry (HM CURATOR) by the Houston  
6  
7 Methodist Research Institute Institutional Review Board (HMRI IRB).  
8  
9

10 **Contribution Statement:**

11  
12 FV: design, data analysis and interpretation, drafting the manuscript, critical revision for  
13  
14 important intellectual content, final approval  
15

16  
17 JCN: data acquisition, data analysis, drafting the manuscript, final approval  
18

19  
20 OK: data acquisition, data analysis, drafting the manuscript, final approval  
21

22  
23 SLJ: data acquisition, data interpretation, critical revision for important intellectual content, final  
24  
25 approval  
26

27  
28 FNM: critical revision for important intellectual content, final approval  
29

30  
31 HDS: critical revision for important intellectual content, final approval  
32

33  
34 RAP: critical revision for important intellectual content, final approval  
35

36  
37 JDA: critical revision for important intellectual content, final approval  
38

39  
40 BK: critical revision for important intellectual content, final approval  
41

42  
43 KN: design, interpretation of data, critical revision for important intellectual content, final  
44  
45 approval  
46

47 **Competing Interests:** No authors declare a competing interest to this work  
48

49 **Funding:** This work was supported internally by the Houston Methodist Academic Institute.  
50

51 **Date sharing statement:** All requests for de-identified data should be made to the corresponding  
52  
53 author. All reasonable requests will be evaluated by the CURATOR Data Governance and  
54  
55 Sharing Committee comprising of FV, SLJ, BK and KN in the light of institutional policies and  
56  
57 guidelines.  
58  
59  
60

## 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](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.harriscountytexas.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. <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

For peer review only

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

	<b>African American n = 1,159</b>	<b>Non-African American n = 3,354</b>	<b>OR / Median Difference (95% CI)</b>	<b>P value</b>
Age: Mean (SD)	48.8 (17.7)	51.2 (19.2)	0.99 (0.99 – 0.99)	< 0.001
<b>Age Category (%)</b>				
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
<b>Females</b>	66.4	60.6	1.28 (1.12 – 1.48)	<0.001
<b>Median (IQR) Zip Income</b>	60,765 (46,801 – 76,163)	75,793 (57,252 – 102,008)	-15,028 (-1,667, -12,388)	< 0.001
<b>Median Zip Income Pentiles (%) – Pentiles of increasing Income</b>				
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
<b>Population Density for Zip: Median (IQR)</b>	3256.8 (2123.6 – 4439.7)	2814.2 (1439.1 – 4260.9)	442.6 (306.7 – 578.5)	< 0.001
<b>Population Density for Zip Pentiles (%) – Pentiles of increasing population density</b>				
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

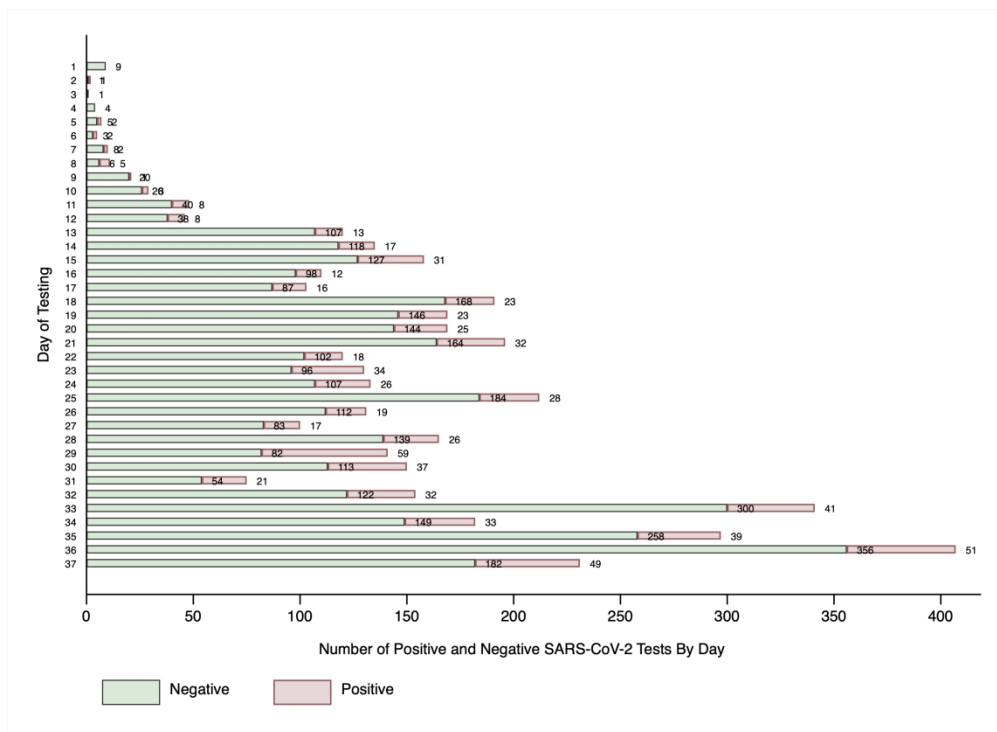


**Table 3:** Adjusted Odds Ratios and 95% Confidence Intervals for socio-demographic and comorbidity factors associated with SARS-CoV-2 infection

Covariate	Adjusted Odds Ratio	95% Confidence Interval	P value
<b>Age Categories</b>			
Up to 35 years	<i>Reference Category</i>		
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
<b>Male (vs. Female)</b>	1.32	1.12 – 1.57	0.001
<b>Race Categories (Non-Hispanic)</b>			
White	<i>Reference Category</i>		
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
<b>Hispanic (vs. Non-Hispanic)</b>	1.70	1.35 – 2.14	<0.001
<b>Median Zip Household Income Categories (Pentiles of Increasing Income)</b>			
Category I	<i>Reference Category</i>		
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
<b>Primary Insurance Type</b>			
Medicare	<i>Reference Category</i>		
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
<b>Zip Population Density (Pentiles of Increasing Density)</b>			
Category I	<i>Reference Category</i>		
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)

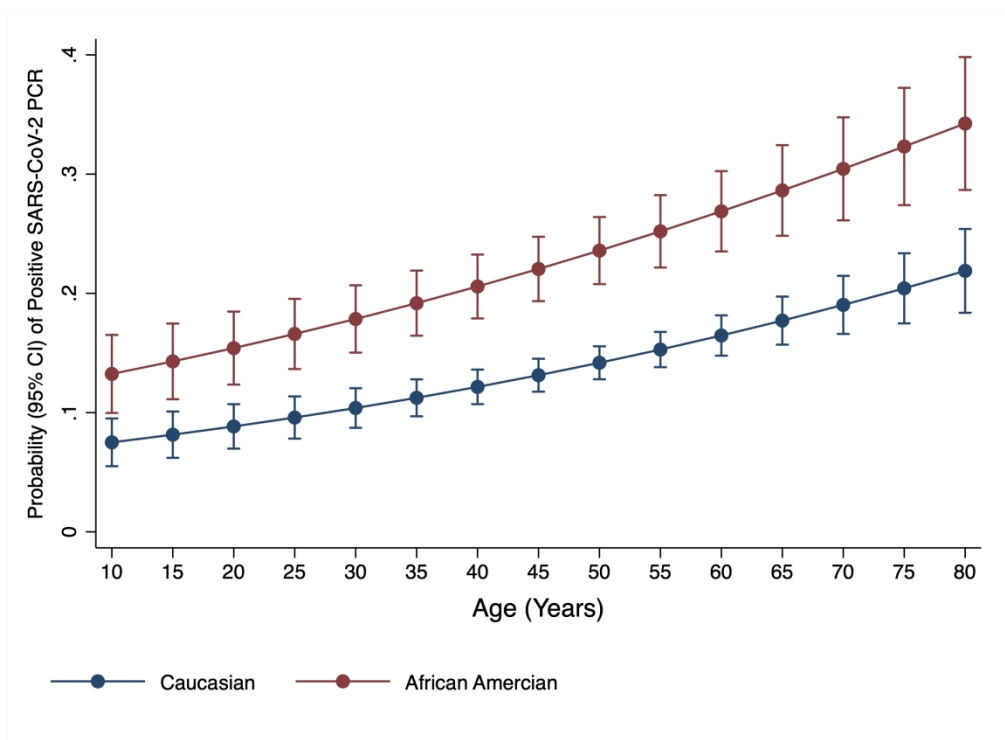
1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60



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

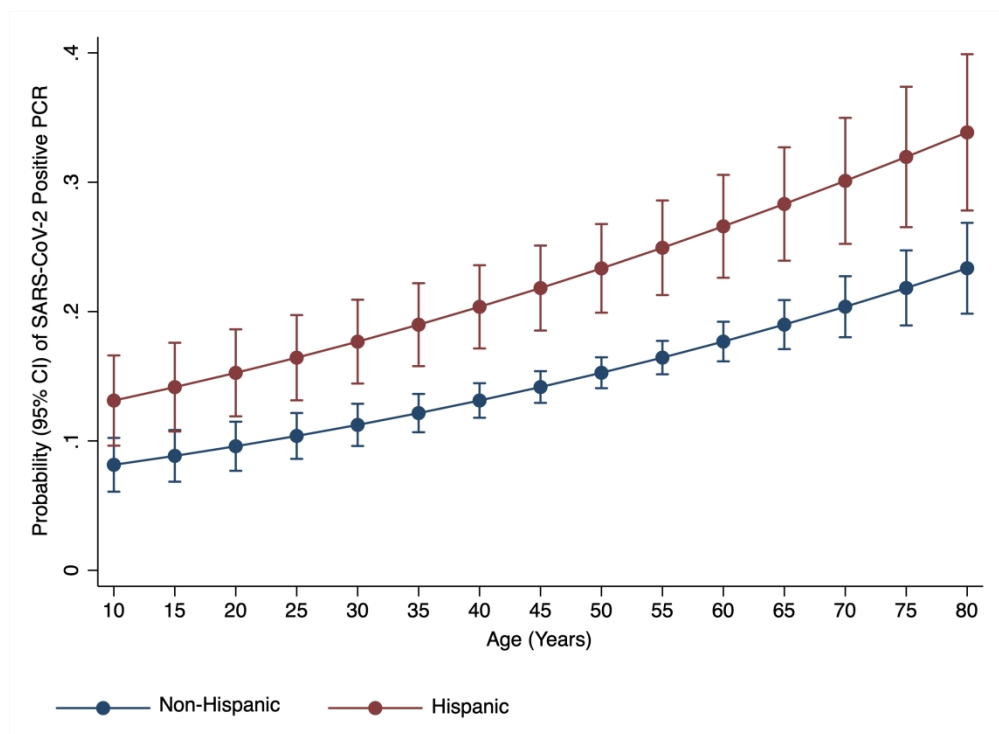
489x355mm (144 x 144 DPI)

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60



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
<b>Title and abstract</b>	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
<b>Introduction</b>			
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
<b>Methods</b>			
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
<b>Results</b>			
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
<b>Discussion</b>			
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
<b>Other information</b>			
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](http://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:	<i>BMJ Open</i>
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, Bit A; Houston Methodist Research Institute; Texas A&M University School of Rural Public Health Nasir, Khurram; Houston Methodist Research Institute; Houston Methodist Hospital
<b>Primary Subject Heading</b>:	Epidemiology
Secondary Subject Heading:	Epidemiology, Infectious diseases, Public health
Keywords:	EPIDEMIOLOGY, Public health < INFECTIOUS DISEASES, INFECTIOUS DISEASES

SCHOLARONE™  
Manuscripts



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.



1  
2  
3 **Racial and Ethnic Disparities in SARS-CoV-2 Pandemic: Analysis of a COVID-19**  
4 **Observational Registry for a Diverse U.S. Metropolitan Population**  
5  
6  
7

8 *Farhaan S. Vahidy PhD MBBS MPH,<sup>1,2</sup> Juan Carlos Nicolas BS,<sup>1</sup> Jennifer R. Meeks MS,<sup>1</sup>*  
9 *Osman Khan BS,<sup>1</sup> Alan Pan MS,<sup>1</sup> Stephen L. Jones MD MSHI,<sup>1,3,4</sup> Faisal N. Masud MD,<sup>4,5</sup>*  
10 *H. Dirk Sostman MD,<sup>4,6</sup> Robert A. Phillips MD PhD,<sup>1,4,7</sup> Julia D. Andrieni MD,<sup>4,8</sup>*  
11 *Bitu Kash PhD MBA,<sup>1,4,9</sup> Khurram Nasir MD MPH<sup>1,7</sup>*  
12  
13  
14  
15

- 16 1. Center for Outcomes Research, The Houston Methodist Research Institute, Houston TX
- 17 2. Houston Methodist Neurological Institute, Houston Methodist Hospital, Houston TX
- 18 3. Department of Surgery, Houston Methodist Hospital, Houston TX
- 19 4. Weill Cornell Medical College, New York NY
- 20 5. Department of Anesthesiology and Critical Care, Houston Methodist Hospital, Houston
- 21 TX
- 22 6. The Houston Methodist Research Institute and The Houston Methodist Academic
- 23 Institute, Houston TX
- 24 7. Department of Cardiology, DeBakey Heart and Vascular Center, Houston Methodist
- 25 Hospital, Houston TX
- 26 8. Department of Medicine, Houston Methodist Hospital, Houston TX
- 27 9. Department of Health Policy and Management, Texas A&M University School of Public
- 28 Health, College Station TX
- 29
- 30
- 31
- 32
- 33
- 34
- 35
- 36
- 37
- 38
- 39

40 **Corresponding Author**

41  
42 **Farhaan S. Vahidy, PhD MBBS MPH**

43 Josie Roberts Building, Suite 4.123

44 Houston Methodist

45 7550 Greenbriar Drive

46 Houston TX 77030

47 [fvahidy@houstonmethodist.org](mailto:fvahidy@houstonmethodist.org)  
48  
49  
50

51  
52  
53  
54  
55 Word Count: 3,062  
56  
57  
58  
59  
60

## 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

1  
2  
3 statistically significant indirect effect of NHB and Hispanic ethnicity on SARS-CoV-2 infection  
4 mediated via a pathway including residence in densely populated zip code.  
5  
6

7 **Conclusions:** There is strong evidence of race and ethnic disparities in the SARS-CoV-2  
8 pandemic, potentially mediated through unique social determinants of health.  
9  
10  
11  
12

### 13 **Strengths and limitations of this study**

- 14  
15  
16 • One of the first studies to systematically evaluate race and ethnic disparities in  
17 susceptibility to SARS-CoV-2 infection, while accounting for multiple socio-  
18 demographic characteristics and comorbidities  
19  
20  
21
- 22  
23 • Study population represents a large and diverse metropolitan of the U.S. with data from  
24 one of the largest healthcare providers across the greater metropolitan area  
25  
26
- 27  
28 • Study evaluates potential mediation pathways for race disparities and demonstrates that  
29 residence in areas with high population density may mediate race and ethnic disparities in  
30 susceptibility to SARS-CoV-2 infection  
31  
32  
33
- 34  
35 • Single center study with limited information about burden of comorbidity and lifestyle  
36 factors  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

## 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.<sup>1,2</sup> 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.<sup>3</sup> 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.<sup>4</sup> 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.<sup>5-6</sup>

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.<sup>7-10</sup> 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.<sup>11-13</sup> 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

1  
2  
3 areas are limited. Furthermore, the mediators of SARS-CoV-2 infection among racial and ethnic  
4  
5 minorities have not been described.  
6

7  
8 We explored socio-demographic characteristics such as age, sex, race, ethnicity, median  
9  
10 household income by zip codes, population density of residents' zip codes, and health insurance  
11  
12 status associated with positive SARS-CoV-2 testing in an urban and diverse population served  
13  
14 by one of the leading healthcare systems of the Greater Houston area. We further examined the  
15  
16 association between pre-existing comorbidities and higher likelihood of SARS-CoV-2 infection  
17  
18 in our study population. We hypothesized that older age, and racial and ethnic minorities will be  
19  
20 associated with significantly higher likelihood of SARS-CoV-2 infection, and factors such as  
21  
22 low socio-economic status, residence in high population density areas (proxy for potential  
23  
24 difficulties in social distancing) and higher comorbidity burden will mediate the effect of race  
25  
26 and ethnicity on SARS-CoV-2 infection.  
27  
28  
29  
30  
31  
32

### 33 **METHODS**

34  
35 We analyzed data between March 5 and May 31, 2020 collected as a part of the COVID-  
36  
37 19 Surveillance and Outcomes Registry (CURATOR) at Houston Methodist (HM). The HM  
38  
39 CURATOR has been approved by the HM Institutional Review Board (IRB) as an observational  
40  
41 quality of care registry for all suspected and confirmed COVID-19 patients. HM IRB granted  
42  
43 CURATOR a waiver of informed consent and HIPAA (Health Insurance Portability and  
44  
45 Accountability Act) authorization in accordance with current federal regulations. The  
46  
47 CURATOR, designed and managed by the big data team at the Center for Outcomes Research  
48  
49 (COR) at HM, is populated from multiple data sources across the HM system such as electronic  
50  
51 medical records, electronic databanks for laboratory and pharmacy, and electronic interactive  
52  
53  
54  
55  
56  
57  
58  
59  
60

1  
2  
3 patient interface tools. The HM system comprises a flagship tertiary care hospital in the Texas  
4  
5 Medical Center, seven large community hospitals, a continuing care hospital, and multiple  
6  
7 emergency centers and clinics throughout the Greater Houston area. Data from various sources  
8  
9 are curated into a harmonized format, assessed for quality and integrity, and stored on a secure  
10  
11 institutional HIPAA-compliant server.  
12  
13

14  
15 We flagged all individuals who were tested for the SARS-CoV-2 using the real time  
16  
17 Reverse Transcriptase (RT) Polymerized Chain Reaction (PCR) diagnostic panels. The three  
18  
19 cross-validated PCR tests utilized were the World Health Organization (WHO) nucleic acid  
20  
21 amplification test, Panther Fusion<sup>®</sup> SARS-CoV-2 Assay, and Cepheid Xpert<sup>®</sup> Xpress SARS-  
22  
23 CoV-2 Assay. These assays were verified for quantitative detection of novel SARS-CoV-2  
24  
25 isolated and purified from nasopharyngeal swab specimens obtained from individuals and  
26  
27 immersed in universal transport medium. Testing was carried out for symptomatic individuals or  
28  
29 for individuals who had a self-reported history of exposure to a COVID-19 case including recent  
30  
31 travel to other countries with high infection rates or hotspots within the U.S.  
32  
33

34  
35 Socio-demographic characteristics including age, sex, race, ethnicity, and payer-status  
36  
37 (insurance type) were obtained from the HM CURATOR for analyses. We also extracted  
38  
39 information on presence of comorbidities comprising the Charlson Comorbidity Index (CCI)  
40  
41 which include past history of myocardial infarction, congestive heart failure, peripheral vascular  
42  
43 disease, cerebrovascular disease, dementia, chronic pulmonary disease, rheumatic disease, peptic  
44  
45 ulcer disease, liver disease, diabetes with or without complications, hemiplegia, renal disease,  
46  
47 any malignancy (excluding skin neoplasms), metastatic solid tumors, and AIDS/HIV. Data on  
48  
49 hypertension and obesity were additionally obtained. We utilized the U.S. Census Bureau's  
50  
51 American Community Survey (ACS) 5-year data (2014–2018) to determine median household  
52  
53  
54  
55  
56  
57  
58  
59

1  
2  
3 income by individual zip code tabulation areas (ZCTA).<sup>14</sup> The median ZCTA household income  
4 was inflation-adjusted to 2018 USD. We also utilized the same data source to obtain population  
5 estimates by ZCTA, and calculated ZCTA level population density (population per mile square)  
6 by standardizing it for area measurements of ZCTA. For the purpose of population density  
7 determination, land area estimates were obtained from the Census Bureau's U.S. Gazetteer Files  
8 2010.<sup>15</sup> In the absence of granular and precise social distancing data, we have utilized population  
9 density as a proxy for potential difficulties in social distancing among crowded communities.  
10  
11  
12  
13  
14  
15  
16  
17  
18

19 We provide descriptive summary data as means (standard deviations) and proportions.  
20  
21 We fit univariable and multivariable logistic regression models to assess unadjusted and adjusted  
22 association between socio-demographic characteristics and likelihood of being tested positive for  
23 SARS-CoV-2. We additionally provide univariable comparison of various socio-demographic  
24 and comorbidity variables between non-Hispanic Black (NHB) and non-Hispanic White (NHW)  
25 race categories, as well as between Hispanic and Non-Hispanic ethnic groups. Age, income,  
26 population density and CCI were categorized for certain analyses. We determined *a priori* to  
27 include all variables (age, sex, race, ethnicity, zip code household income, insurance type, zip  
28 population density and CCI) in our initial multivariable model. Factors demonstrating mediation  
29 were excluded from the final model. We assessed the model fit utilizing the Hosmer-Lemeshow  
30 goodness of fit test, and crude and adjusted odds ratios (OR and aOR) and 95% confidence  
31 intervals (CI) are reported. Post-estimation marginal probabilities of SARS-CoV-2 infection  
32 were determined from the fully adjusted model for major covariates (race, ethnicity and age). We  
33 explored the mediation influence of comorbidity burden (CCI), socio-economic status (median  
34 income), and lack of social distancing (population density) on the relationship of Black race and  
35 Hispanic ethnicity with high likelihood of SARS-CoV-2 infection using the Generalized  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

1  
2  
3 Structural Equation Modeling (GSEM) framework. The GSEM framework was set up to provide  
4 estimates of direct and indirect effect of Black race and Hispanic ethnicity on SARS-CoV-2  
5 infectivity. Statistically significant ( $p < 0.05$ ) indirect effects represent full or partial mediation  
6 by a tested covariate. We included all individuals tested for SARS-CoV-2 across our healthcare  
7 system and did not perform formal sample size calculations.  
8  
9  
10  
11  
12  
13

### 14 **Patient and Public Involvement**

15  
16  
17 There was no direct patient or public involvement in the design and conduct of this study.  
18  
19  
20

## 21 **RESULTS**

### 22 **Socio-demographic and comorbidity characteristics of the study population**

23  
24  
25  
26 Across the time period of analysis, we identified a total of 20,228 presumed cases tested  
27 for SARS-CoV-2, among whom 1,551 (7.7%, CI: 7.3-8.0) tested positive. Overall, the mean  
28 (SD) age of the study population was 51.1 (19.0) years; 61.9% were female and 62.3% were  
29 White (including Hispanic ethnicity). The study sample was comparable to the overall  
30 population of patients treated across HM, who have a mean (SD) age of 49.0 (22) years, are 56%  
31 female, and 53% White. The HM system metrics was derived from a sample of 3,216,290  
32 patients managed across the system since May 22, 2016.  
33  
34  
35  
36  
37  
38  
39  
40  
41

42 The overall median (IQR) household income was USD \$70,658 (\$53,313–\$99,276), and  
43 42.6% of the study population had private or employer-based insurance. In our univariate  
44 analysis, Black race (vs. White; OR, CI: 1.55, 1.37–1.75), Hispanic ethnicity (vs. non-Hispanic;  
45 OR, CI: 2.02, 1.79–2.27), and males (vs. females; OR, CI: 1.17, 1.06–1.31) were associated with  
46 significantly higher likelihood of testing positive for SARS-CoV-2. Among the SARS-CoV-2  
47 positive patients, 40.8% were in the age category of 51–75 years, and 11.4% were greater than  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60



1  
2  
3 75 years. These proportions were significantly higher than the reference group (up to 35 years;  
4  
5 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:  
6  
7 1.23, 1.02–1.49). Furthermore, individuals in higher pentiles of socio-economic status had  
8  
9 significantly lower likelihood, whereas those residing in higher population density ZCTAs had  
10  
11 higher likelihood of SARS-CoV-2 infection. We observed a significantly higher proportion of  
12  
13 SARS-CoV-2 positive individuals in the CCI 1-2 category compared to CCI of 0 (OR, CI: 1.35,  
14  
15 1.18–1.54). However, similar differences for higher CCI categories were not observed. For  
16  
17 specific comorbidities, a significantly greater proportion of diabetic individuals had SARS-CoV-  
18  
19 2 positive results (OR, CI: 1.40, 0.17–1.68). The socio-demographic characteristics and  
20  
21 comorbidity profiles for the overall and SARS-CoV-2 positive and negative patients are  
22  
23 summarized in Table 1.  
24  
25  
26  
27

### 28 **Socio-demographic and comorbidity characteristics associated with minority race and** 29 30 **ethnicity** 31 32

33 In our study sample comprising of 13,754 Non-Hispanic Black and White individuals, we  
34  
35 compared the association between race and various socio-demographic and comorbidity  
36  
37 characteristics (Table 2). Similarly, we also evaluated univariable differences for socio-  
38  
39 demographic variables and co-morbidities between Hispanic and non-Hispanic individuals  
40  
41 (Table 3). Minority race (NHB) and ethnicity (Hispanic) were both associated with younger age,  
42  
43 higher proportion of females, and residence in low income and higher population density  
44  
45 ZCTAs. However, NHB and Hispanic groups were both associated with an overall lower burden  
46  
47 of comorbidities (as demonstrated by significantly lower median CCI) compared respectively to  
48  
49 NHW and non-Hispanic categories. A higher proportion of individuals among minority race and  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

1  
2  
3 ethnicity were diabetic, and a higher proportion of NHB were also hypertensive compared to  
4  
5  
6 NHW.

### 7 **Multivariable model and marginal probabilities for likelihood of SARS-CoV-2 infection** 8 9 **and racial and ethnic minorities**

10  
11  
12 The significantly higher likelihood of SARS-CoV-2 infection among minority race and  
13  
14 ethnic groups persisted after controlling for other demographics, insurance type, median  
15  
16 household income, population density, and comorbidities. Adjusted odds ratios (CI) for NHB vs.  
17  
18 NHW was 2.23 (1.90 – 2.60) and for Hispanic vs. Non-Hispanic was 1.95 (1.72 – 2.20). Higher  
19  
20 risk of infection among males (compared to females) and higher likelihood of SARS-CoV-2  
21  
22 infection among elderly also remained statistically significant. Detailed outputs of the fully  
23  
24 adjusted logistic regression models for minority race and ethnic groups are presented in Table 4.  
25  
26 Based on the marginal probabilities obtained from our fully adjusted model, the probability of  
27  
28 SARS-CoV-2 infection in a 45-year-old NHB is 9.6% whereas it is 4.5% in a 45-year-old NHW  
29  
30 individual, all other adjusted variables being constant. At the age of 75, this probability is 14.0%  
31  
32 for an NHB and 6.9% for a NHW. A similar relationship differential was observed for Hispanic  
33  
34 vs. non-Hispanic individuals. Multivariable model derived probabilities of SARS-CoV-2  
35  
36 infection for NHB vs. NHW and for Hispanic vs. Non-Hispanic across age spectrum are  
37  
38 presented in Figure 1 and Figure 2.  
39  
40  
41  
42  
43

### 44 **Generalized Structural Equation Modeling for mediation by income, population density** 45 46 **and Comorbidity Index**

47  
48  
49 Utilizing the GSEM framework, we determined the direct and indirect effects of NHB  
50  
51 and Hispanic ethnicity on SARS-CoV-2 infection with median income, population density and  
52  
53 CCI modeled as mediators in six separate equations adjusted for age and sex. The indirect effect  
54  
55  
56  
57  
58  
59  
60

1  
2  
3 of NHB mediated through population density was statistically significant (OR, CI: 1.03, 1.01 –  
4 1.05,  $p = 0.001$ ); however, the indirect effects mediated via median income and comorbidity  
5 scores were not statistically significant ( $p = 0.14$  and  $p = 0.64$  respectively). Among individuals  
6 identifying as Hispanic or Latino, both population density and income partially mediated the  
7 effect of ethnicity on SARS-CoV-2 positivity (OR, CI for population density: 1.02, 1.01 – 1.02,  
8  $p < 0.001$  and OR, CI for income: 1.04, 1.02 – 1.06,  $p < 0.001$ ). Evaluation of comorbidities did  
9 not suggest a mediation influence for either NHB or Hispanic categories.

## 20 **DISCUSSION**

21  
22  
23 The underlying race and ethnic healthcare disparities have been painfully highlighted in  
24 the wake of the COVID-19 pandemic. Most reports indicate higher mortality or case fatality  
25 among minority racial groups (Black / African American) across major U.S. metropolitan  
26 areas.<sup>11–13</sup> However, robust insights on the racial differences for SARS-CoV-2 infection are  
27 limited. Furthermore, comprehensive data evaluating higher susceptibility to SARS-CoV-2  
28 infection among Hispanic communities are also scarce. This is perhaps because of comparatively  
29 homogenous populations in non-U.S. regions of the world. Houston, as an exceptionally  
30 ethnically diverse population center,<sup>16</sup> is well suited for an investigation of racial, ethnic, and  
31 socioeconomic gradients in COVID-19 test positivity. We focus on highlighting the mechanisms  
32 of racial and ethnic disparities in susceptibility to SARS-CoV-2 infection and provide evidence  
33 of mediation of such disparities by novel social determinants of health (SDoH).  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46

47  
48 Our study adds to the current literature by analyzing emerging data for individuals being  
49 tested across one of the largest healthcare systems in the Greater Houston area. We report that  
50 racial and ethnic minorities (non-Hispanic Black and Hispanic individuals) are almost twice as  
51 likely to test positive for SARS-CoV-2 than the non-Hispanic White and non-Hispanic  
52  
53  
54  
55  
56  
57  
58  
59  
60

1  
2  
3 population. These findings illuminate systematic racial / ethnic disparities in testing positive for  
4 SARS-CoV-2 infection. Though there are limited prior SARS-CoV-2 data, such racial and ethnic  
5 disparities have previously been described for the U.S. H1N1 influenza pandemic.<sup>17</sup> These data  
6 indicated that Spanish-speaking Hispanic and Black individuals were at a greater risk of H1N1  
7 infection, primarily attributable to lack of healthcare access.  
8  
9

10  
11 We explored three possible mechanisms of race disparities in our data. These included  
12 lower socio-economic status, residence in higher population dense areas, and higher level of  
13 comorbidities. We demonstrate that NHB race is significantly associated with all three potential  
14 disparity pathways, and in the traditional multivariable analyses, racial and ethnic disparities  
15 persisted even after controlling for these pathways. However, our mediation analyses highlighted  
16 the potential influence of residence in high population density areas as a viable pathway that at  
17 least partially explains the observed racial and ethnic disparity. Furthermore, residence in low  
18 income areas emerged as a significant mediation pathway for ethnic differences in SARS-CoV-2  
19 positivity. Pathways mediating the influence of comorbidity status did not demonstrate a  
20 significant effect. We utilized population density as a marker for potential inability to maintain  
21 adequate social distancing as it has been indicated that maintaining the WHO recommended safe  
22 distance between people becomes challenging with high population densities.<sup>18</sup> Furthermore,  
23 overall effects of population density and disease spread has been previously described in  
24 literature.<sup>19,20</sup> In addition to lack of social distancing, higher population density may also be  
25 associated with several other behavioral and socio-demographic attributes that may predispose  
26 populations to both viral spread and increased susceptibility. For example, there are reports  
27 linking obesity, lack of physical activity, and higher mortality with residence in densely  
28 populated neighborhoods.<sup>21,22</sup>  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

1  
2  
3 As reported, our data also corroborate that older populations may be more susceptible to  
4 SARS-CoV-2 infection.<sup>10</sup> However, younger populations still have cause for concern as nearly 1  
5  
6 in 4 of the infected cases in our sample were between 36–50 years of age. Finally, our data  
7  
8 demonstrate that males may be approximately 20% more likely to test positive for the SARS-  
9  
10 CoV-2 infection. Potential sex differences in infectivity to SARS-CoV-2 and intersectionality  
11  
12 with racial and ethnic socioeconomic factors need to be explored further in future analyses.  
13  
14 Additional policy-oriented research should prioritize studying the intersectionality of these  
15  
16 vulnerable economic statuses and racial disparities in COVID infection indicated by the present  
17  
18 study.  
19  
20  
21  
22  
23

24 Findings of our study need to be interpreted in the light of certain limitations. Our data  
25  
26 are from a single center and may not be generalizable to the wider U.S. population. These  
27  
28 findings need to be replicated in larger data sets across other large heterogenous U.S.  
29  
30 metropolitans. However, the Houston metropolitan area is one of the most diverse and  
31  
32 representative in the U.S.,<sup>16</sup> and our healthcare system is one of the largest systems providing  
33  
34 care to COVID-19 patients in the Greater Houston area. Our sample was composed of 22%  
35  
36 Black, 18% Hispanic, and 62% female population. We did not have information on certain  
37  
38 demographic covariates such as education or household size. Educational status has been linked  
39  
40 to healthcare awareness and may be important to adjust for in analyses of potential disparities,  
41  
42 and household size may be used to provide more precise estimates of socio-economic status.  
43  
44 However, we obtained and adjusted for zip code income data from the U.S. Census, as income  
45  
46 has previously been shown to have strong correlation with educational attainment and socio-  
47  
48 economic status.<sup>23</sup> Since testing was based on suspicion of infection and may have been  
49  
50 influenced by factors such as access to care, the potential for selection bias cannot be ruled out.  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

1  
2  
3 Furthermore, lack of sensitivity of SARS-CoV-2 diagnostics tests have been reported; however,  
4 the three assays utilized for testing were cross validated for internal consistency. Finally, we did  
5 not have detailed information on comorbidities and their management in the study population.  
6  
7 However, we did control for major comorbidities which are being reported as associated with  
8 COVID-19 outcomes.<sup>24</sup>  
9  
10  
11  
12  
13  
14  
15  
16

## 17 **Conclusions**

18  
19 The strong association between racial and ethnic minorities and SARS-CoV-2 infection  
20 demonstrated in our data, even after adjustment for other important socio-demographic and  
21 comorbidity factors, highlight a potential catastrophe of inequality within the existential crisis of  
22 a global pandemic. Our data, representing a large heterogeneous U.S. metropolitan area, also  
23 provide preliminary evidence into the potential pathways for this disparity. It is highly likely that  
24 higher comorbidity burden and detrimental effects of adverse social determinants, including  
25 those that may not adequately permit safe practices of social distancing, mediate higher SARS-  
26 CoV-2 infectivity among racial and ethnic minorities.  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36

37  
38 As the pandemic continues to spread and evolve across the continental U.S., emerging  
39 data on association between SARS-CoV-2 infection and various socio-demographic factors will  
40 continue to enhance our understanding of targeted risks related to SARS-CoV-2 infection, and  
41 such data would enable us to comprehend healthcare services and access factors related to  
42 development and outcomes of COVID-19 among minority populations. Our findings substantiate  
43 prior calls for collection of robust data on race and ethnicity as a part of international  
44 collaborations,<sup>25</sup> and further drive home the critical importance of quantifying novel SDoH.  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

1  
2  
3 **Acknowledgment:** The authors thank Jacob M. Kolman, Senior Scientific Writer of the Houston  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

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

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

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

For peer review only



## 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](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->

- 1  
2  
3 deaths-demographics-lightfoot-20200406-77nlylhiavgjzb2wa4ckivh7mu-story.html.  
4 Published April 7, 2020. Accessed April 20, 2020.  
5  
6  
7 13. Thebault R, Tran AB, Williams V. The coronavirus is infecting and killing black Americans  
8 at an alarmingly high rate. *Washington Post*.  
9 [https://www.washingtonpost.com/nation/2020/04/07/coronavirus-is-infecting-killing-black-](https://www.washingtonpost.com/nation/2020/04/07/coronavirus-is-infecting-killing-black-americans-an-alarmingly-high-rate-post-analysis-shows/)  
10 [americans-an-alarmingly-high-rate-post-analysis-shows/](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  
11 April 20, 2020.  
12  
13 14. United States Census Bureau. American Community Survey (ACS).  
14 <https://www.census.gov/programs-surveys/acs>. Published 2018. Accessed April 14, 2020.  
15  
16 15. United States Census Bureau. Gazetteer Files.  
17 <https://www.census.gov/geographies/reference-files/time-series/geo/gazetteer-files.html>.  
18 Accessed April 20, 2020.  
19  
20 16. Harris County Public Health. *Harris Cares: A 2020 Vision of Health in Harris County*.  
21 Houston, TX: Harris County Public Health; 2019.  
22 <http://publichealth.harriscountytexas.gov/Resources/Harris-Cares>. Accessed November 21,  
23 2019.  
24  
25 17. Quinn SC, Kumar S, Freimuth VS, Musa D, Casteneda-Angarita N, Kidwell K. Racial  
26 disparities in exposure, susceptibility, and access to health care in the US H1N1 influenza  
27 pandemic. *Am J Public Health*. 2011;101(2):285-293. doi:10.2105/AJPH.2009.188029  
28  
29 18. Rocklov J, Sjodin H. High population densities catalyze the spread of COVID-19. *J Travel*  
30 *Med*. 2020. doi: 10.1093/jtm/taaa038  
31  
32 19. Tarwater PM, Martin CF. Effects of population density on the spread of disease. *Complexity*.  
33 2001;6(6):29-36. doi:10.1002/cplx.10003  
34  
35 20. Yang K, LeJeune J, Alsdorf D, Lu B, Shum CK, Liang S. Global distribution of outbreaks of  
36 water-associated infectious diseases. *PLoS Negl Trop Dis*. 2012;6(2):e1483.  
37 doi:10.1371/journal.pntd.0001483  
38  
39 21. Wang R, Feng Z, Xue D, Liu Y, Wu R. Exploring the links between population density,  
40 lifestyle, and being overweight: secondary data analyses of middle-aged and older Chinese  
41 adults. *Health Qual Life Outcomes*. 2019;17(1):100. doi:10.1186/s12955-019-1172-3  
42  
43 22. Beenackers MA, Oude Groeniger J, Kamphuis CBM, Van Lenthe FJ. Urban population  
44 density and mortality in a compact Dutch city: 23-year follow-up of the Dutch GLOBE  
45 study. *Health & Place*. 2018;53:79-85. doi:10.1016/j.healthplace.2018.06.010  
46  
47 23. U.S. Bureau of Labor Statistics. Unemployment rates and earnings by educational  
48 attainment. <https://www.bls.gov/emp/chart-unemployment-earnings-education.htm>.  
49 Published September 4, 2019. Accessed April 15, 2020.  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

- 1  
2  
3 24. Centers for Disease Control and Prevention. Groups at higher risk for severe illness. cdc.gov.  
4 [https://www.cdc.gov/coronavirus/2019-ncov/need-extra-precautions/groups-at-higher-](https://www.cdc.gov/coronavirus/2019-ncov/need-extra-precautions/groups-at-higher-risk.html)  
5 [risk.html](https://www.cdc.gov/coronavirus/2019-ncov/need-extra-precautions/groups-at-higher-risk.html). Last updated April 17, 2020. Accessed April 20, 2020.  
6  
7  
8 25. Abuelgasim E, Saw LJ, Shirke M, Zeinah M, Harky A. COVID-19: Unique public health  
9 issues facing Black, Asian and minority ethnic communities. *Curr Probl Cardiol.*  
10 2020;100621. doi:10.1016/j.cpcardiol.2020.100621  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

## 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

For peer review only

**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) <sup>a</sup>
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 (%) <sup>b</sup>				
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 (%) <sup>c</sup>	17.8	16.8	29.2	2.02 (1.79 – 2.27)
Median Zip Household Income (IQR) <sup>d</sup>	70,658 (53,313–99,276)	70,758 (53,633–100,107)	66,523 (50,485–94,226)	-4,168 <sup>e</sup> (-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 <sup>f</sup>	2797.2 (1439.1 – 4260.9)	2797.2 (1439.1 – 4211.4)	3320.3 (1904.4 – 4439.7)	523.1 <sup>e</sup> (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) <sup>f</sup>
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) <sup>a</sup>
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)

<sup>a</sup> Unadjusted Odds Ratios and 95% Confidence Intervals for association between individual co-variates in SARS-CoV-2 positivity.

<sup>b</sup> Race: Missing, Unknown, Declined, n = 1157 (5.7%).

<sup>c</sup> Ethnicity: Missing, Unknown, Declined, n = 613 (3.0%).

<sup>d</sup> 2018 inflation adjusted USD. Missing n = 1,883 (9.3%). Pentiles were defined by categorizing the ordered distribution of median income into five categories.

<sup>e</sup> Difference in median and 95% CI of difference obtained via quantile regression.

<sup>f</sup> 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 <sup>a</sup> / 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
<b>Age Category (%)</b>				
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
<b>Females</b>	68.1	58.1	1.54 (1.42 – 1.66)	< 0.001
<b>Median (IQR) Zip Income</b>	64,022 (47,303 – 79,658)	76,163 (60,130 – 102,019)	-12,141 (-14,018, -10,263)	< 0.001
<b>Median Zip Income Pentiles (%) – Pentiles of increasing Income<sup>b</sup></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
<b>Population Density for Zip: Median (IQR)</b>	3217.6 (2040.7 – 4439.7)	2488.5 (812.2 – 4084.3)	729.1 (603.0 – 855.2)	< 0.001
<b>Population Density for Zip Pentiles (%) – Pentiles of increasing population density<sup>c</sup></b>				
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
<b>Charlson Comorbidity Index, median (IQR)</b>	2 (0 – 6)	3 (1 – 7)	-1 (-1.22, -0.78)	< 0.001
<b>Charlson Comorbidity Index (CCI) Categories (%)</b>				
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

<sup>a</sup> 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.

<sup>b,c</sup> 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 <sup>a</sup> / Median Difference <sup>a</sup> (95% CI)	P value
Age: Mean (SD)	45.1 (18.3)	52.8 (18.9)	0.98 (0.98 – 0.98)	< 0.001
<b>Age Category (%)</b>				
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
<b>Females</b>	64.5	61.4	1.14 (1.06 – 1.23)	0.001
<b>Median (IQR) Zip Income</b>	65,742 (48,345 – 82,708)	73,742 (56,288 – 102,008)	-8,000 (-9,450.5, -6,549.5)	< 0.001
<b>Median Zip Income Pentiles (%) – Pentiles of increasing Income<sup>b</sup></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
<b>Population Density for Zip: Median (IQR)</b>	3256.8 (1504.0 – 4299.7)	2741.6 (1408.1 – 4110.7)	515.2 (469.2 – 561.2)	< 0.001
<b>Population Density for Zip Pentiles (%) – Pentiles of increasing population density<sup>c</sup></b>				
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
<b>Charlson Comorbidity Index, median (IQR)</b>	1 (0 – 4)	2 (0 – 6)	-1 (-1.21, -0.79)	< 0.001
<b>Charlson Comorbidity Index (CCI) Categories (%)</b>				
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

<sup>a</sup> 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.

<sup>b,c</sup> 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) <sup>a</sup>	Hispanic vs. Non-Hispanic, aOR (95% CI) <sup>b</sup>
<b>Non-Hispanic Black (vs. Non-Hispanic White)</b>	<b>2.23 (1.90 – 2.60)<sup>c</sup></b>	
<b>Hispanic (vs. Non-Hispanic)</b>		<b>1.95 (1.72 – 2.20)<sup>c</sup></b>
<b>Age Categories</b>		
Up to 35 years	<i>Reference Category</i>	
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)
<b>Male (vs. Female)</b>	1.20 (1.04 – 1.39)	1.17 (1.05 – 1.32)
<b>Median Zip Household Income Categories (Pentiles of Increasing Income)</b>		
Category I	<i>Reference Category</i>	
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)	
<b>Primary Insurance Type</b>		
Medicare	<i>Reference Category</i>	
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)
<b>Charlson Comorbidity Index Categories</b>		
CCI: 0	<i>Reference Category</i>	
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)

<sup>a,b</sup> Hosmer and Lemeshow goodness of fit p-value: 0.58 and 0.73 ( $H_0$ : Model fit is correct).

<sup>c</sup> 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.

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

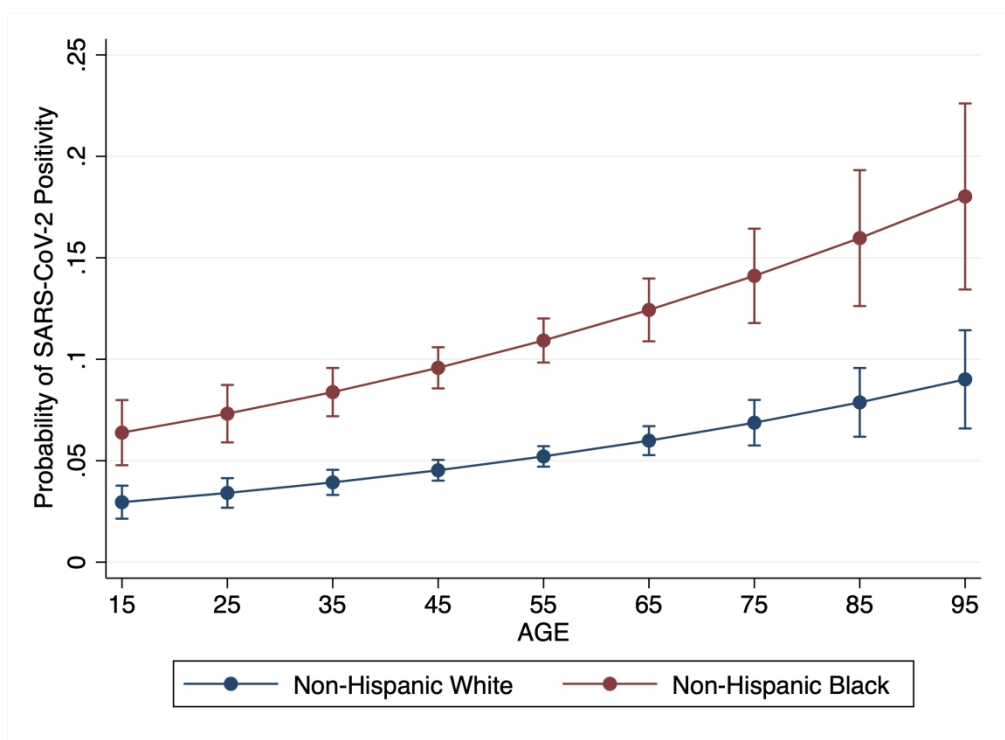


Figure 1: Adjusted Probability and 95% Confidence Interval of Positive SARS-CoV-2 PCR in Non-Hispanic Black vs. Non-Hispanic White by increasing age

467x339mm (144 x 144 DPI)

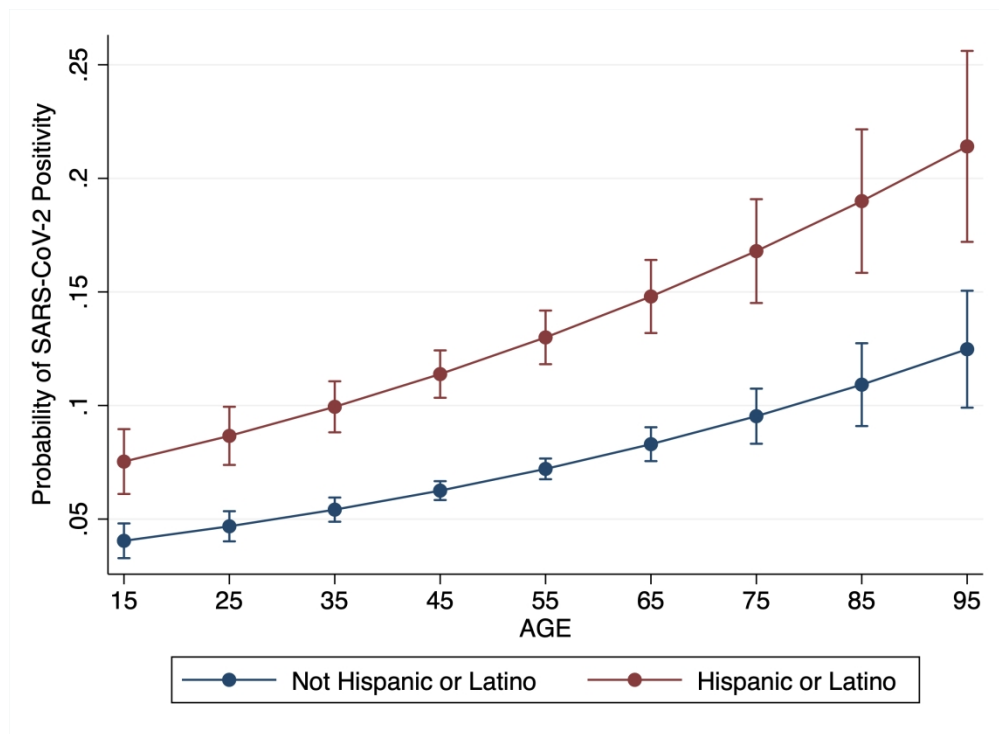


Figure 2: Adjusted Probability and 95% Confidence Interval of Positive SARS-CoV-2 PCR in Hispanic vs. Non-Hispanic by increasing age

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
<b>Title and abstract</b>	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
<b>Introduction</b>			
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
<b>Methods</b>			
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
<b>Results</b>			
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
<b>Discussion</b>			
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
<b>Other information</b>			
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](http://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:	<i>BMJ Open</i>
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, Bit A; Houston Methodist Research Institute; Texas A&M University School of Rural Public Health Nasir, Khurram; Houston Methodist Research Institute; Houston Methodist Hospital
<b>Primary Subject Heading</b>:	Epidemiology
Secondary Subject Heading:	Epidemiology, Infectious diseases, Public health
Keywords:	EPIDEMIOLOGY, Public health < INFECTIOUS DISEASES, INFECTIOUS DISEASES

SCHOLARONE™  
Manuscripts



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.

1  
2  
3 **Racial and Ethnic Disparities in SARS-CoV-2 Pandemic: Analysis of a COVID-19**  
4 **Observational Registry for a Diverse U.S. Metropolitan Population**  
5  
6  
7

8 *Farhaan S. Vahidy PhD MBBS MPH,<sup>1,2</sup> Juan Carlos Nicolas BS,<sup>1</sup> Jennifer R. Meeks MS,<sup>1</sup>*  
9 *Osman Khan BS,<sup>1</sup> Alan Pan MS,<sup>1</sup> Stephen L. Jones MD MSHI,<sup>1,3,4</sup> Faisal N. Masud MD,<sup>4,5</sup>*  
10 *H. Dirk Sostman MD,<sup>4,6</sup> Robert A. Phillips MD PhD,<sup>1,4,7</sup> Julia D. Andrieni MD,<sup>4,8</sup>*  
11 *Bitu Kash PhD MBA,<sup>1,4,9</sup> Khurram Nasir MD MPH<sup>1,7</sup>*  
12  
13  
14  
15

- 16 1. Center for Outcomes Research, The Houston Methodist Research Institute, Houston TX
- 17 2. Houston Methodist Neurological Institute, Houston Methodist Hospital, Houston TX
- 18 3. Department of Surgery, Houston Methodist Hospital, Houston TX
- 19 4. Weill Cornell Medical College, New York NY
- 20 5. Department of Anesthesiology and Critical Care, Houston Methodist Hospital, Houston
- 21 TX
- 22 6. The Houston Methodist Research Institute and The Houston Methodist Academic
- 23 Institute, Houston TX
- 24 7. Department of Cardiology, DeBakey Heart and Vascular Center, Houston Methodist
- 25 Hospital, Houston TX
- 26 8. Department of Medicine, Houston Methodist Hospital, Houston TX
- 27 9. Department of Health Policy and Management, Texas A&M University School of Public
- 28 Health, College Station TX
- 29
- 30
- 31
- 32
- 33
- 34
- 35
- 36
- 37
- 38
- 39

40 **Corresponding Author**

41  
42 **Farhaan S. Vahidy, PhD MBBS MPH**

43 Josie Roberts Building, Suite 4.123

44 Houston Methodist

45 7550 Greenbriar Drive

46 Houston TX 77030

47 [fvahidy@houstonmethodist.org](mailto:fvahidy@houstonmethodist.org)  
48  
49  
50

51  
52  
53  
54  
55 Word Count: 3,063  
56  
57  
58  
59  
60



## 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

1  
2  
3 demonstrated a significant indirect effect of NHB and Hispanic ethnicity on SARS-CoV-2  
4 infection mediated via a pathway including residence in densely populated zip code.  
5  
6

7 **Conclusions:** There is strong evidence of race and ethnic disparities in the SARS-CoV-2  
8 pandemic, that is potentially mediated through unique social determinants of health.  
9  
10  
11  
12

### 13 **Strengths and limitations of this study**

- 14  
15  
16 • One of the first studies to systematically evaluate race and ethnic disparities in  
17 susceptibility to SARS-CoV-2 infection, while accounting for multiple socio-  
18 demographic characteristics and comorbidities  
19  
20  
21
- 22  
23 • Study population represents a large and diverse metropolitan of the U.S. with data from  
24 one of the largest healthcare providers across the greater metropolitan area  
25  
26
- 27  
28 • Study evaluates potential mediation pathways for race disparities and demonstrates that  
29 residence in areas with high population density may mediate race and ethnic disparities in  
30 susceptibility to SARS-CoV-2 infection  
31  
32  
33
- 34  
35 • Single center study with limited information about burden of comorbidity and lifestyle  
36 factors  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

## 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.<sup>1,2</sup> 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.<sup>3</sup> 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.<sup>4</sup> 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.<sup>5-6</sup>

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.<sup>7-10</sup> 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.<sup>11-13</sup> 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

1  
2  
3 areas are limited. Furthermore, the mediators of SARS-CoV-2 infection among racial and ethnic  
4  
5 minorities have not been described.  
6

7  
8 We explored socio-demographic characteristics such as age, sex, race, ethnicity, median  
9  
10 household income by zip codes, population density of residents' zip codes, and health insurance  
11  
12 status associated with positive SARS-CoV-2 testing in an urban and diverse population served  
13  
14 by one of the leading healthcare systems of the Greater Houston area. We further examined the  
15  
16 association between pre-existing comorbidities and higher likelihood of SARS-CoV-2 infection  
17  
18 in our study population. We hypothesized that older age, and racial and ethnic minorities will be  
19  
20 associated with significantly higher likelihood of SARS-CoV-2 infection, and factors such as  
21  
22 low socio-economic status, residence in high population density areas (proxy for potential  
23  
24 difficulties in social distancing) and higher comorbidity burden will mediate the effect of race  
25  
26 and ethnicity on SARS-CoV-2 infection.  
27  
28  
29  
30  
31  
32

### 33 **METHODS**

34  
35 We analyzed data between March 5 and May 31, 2020 collected as a part of the COVID-  
36  
37 19 Surveillance and Outcomes Registry (CURATOR) at Houston Methodist (HM). The HM  
38  
39 CURATOR has been approved by the HM Institutional Review Board (IRB) as an observational  
40  
41 quality of care registry for all suspected and confirmed COVID-19 patients. HM IRB granted  
42  
43 CURATOR a waiver of informed consent and HIPAA (Health Insurance Portability and  
44  
45 Accountability Act) authorization in accordance with current federal regulations. The  
46  
47 CURATOR, designed and managed by the big data team at the Center for Outcomes Research  
48  
49 (COR) at HM, is populated from multiple data sources across the HM system such as electronic  
50  
51 medical records, electronic databanks for laboratory and pharmacy, and electronic interactive  
52  
53  
54  
55  
56  
57  
58  
59  
60

1  
2  
3 patient interface tools. The HM system comprises a flagship tertiary care hospital in the Texas  
4  
5 Medical Center, seven large community hospitals, a continuing care hospital, and multiple  
6  
7 emergency centers and clinics throughout the Greater Houston area. Data from various sources  
8  
9 are curated into a harmonized format, assessed for quality and integrity, and stored on a secure  
10  
11 institutional HIPAA-compliant server.  
12  
13

14  
15 We flagged all individuals who were tested for the SARS-CoV-2 using the real time  
16  
17 Reverse Transcriptase (RT) Polymerized Chain Reaction (PCR) diagnostic panels. The three  
18  
19 cross-validated PCR tests utilized were the World Health Organization (WHO) nucleic acid  
20  
21 amplification test, Panther Fusion<sup>®</sup> SARS-CoV-2 Assay, and Cepheid Xpert<sup>®</sup> Xpress SARS-  
22  
23 CoV-2 Assay. These assays were verified for quantitative detection of novel SARS-CoV-2  
24  
25 isolated and purified from nasopharyngeal swab specimens obtained from individuals and  
26  
27 immersed in universal transport medium. Testing was carried out for symptomatic individuals or  
28  
29 for individuals who had a self-reported history of exposure to a COVID-19 case including recent  
30  
31 travel to other countries with high infection rates or hotspots within the U.S.  
32  
33

34  
35 Socio-demographic characteristics including age, sex, race, ethnicity, and payer-status  
36  
37 (insurance type) were obtained from the HM CURATOR for analyses. We also extracted  
38  
39 information on presence of comorbidities comprising the Charlson Comorbidity Index (CCI)  
40  
41 which include past history of myocardial infarction, congestive heart failure, peripheral vascular  
42  
43 disease, cerebrovascular disease, dementia, chronic pulmonary disease, rheumatic disease, peptic  
44  
45 ulcer disease, liver disease, diabetes with or without complications, hemiplegia, renal disease,  
46  
47 any malignancy (excluding skin neoplasms), metastatic solid tumors, and AIDS/HIV. Data on  
48  
49 hypertension and obesity were additionally obtained. We utilized the U.S. Census Bureau's  
50  
51 American Community Survey (ACS) 5-year data (2014–2018) to determine median household  
52  
53  
54  
55  
56  
57  
58  
59

1  
2  
3 income by individual zip code tabulation areas (ZCTA).<sup>14</sup> The median ZCTA household income  
4 was inflation-adjusted to 2018 USD. We also utilized the same data source to obtain population  
5 estimates by ZCTA, and calculated ZCTA level population density (population per mile square)  
6 by standardizing it for area measurements of ZCTA. For the purpose of population density  
7 determination, land area estimates were obtained from the Census Bureau's U.S. Gazetteer Files  
8 2010.<sup>15</sup> In the absence of granular and precise social distancing data, we have utilized population  
9 density as a proxy for potential difficulties in social distancing among crowded communities.

19 We provide descriptive summary data as means (standard deviations) and proportions.  
21 We fit univariable and multivariable logistic regression models to assess unadjusted and adjusted  
22 association between socio-demographic characteristics and likelihood of being tested positive for  
23 SARS-CoV-2. We additionally provide univariable comparison of various socio-demographic  
24 and comorbidity variables between non-Hispanic Black (NHB) and non-Hispanic White (NHW)  
25 race categories, as well as between Hispanic and Non-Hispanic ethnic groups. Age, income,  
26 population density and CCI were categorized for certain analyses. We determined *a priori* to  
27 include all variables (age, sex, race, ethnicity, zip code household income, insurance type, zip  
28 population density and CCI) in our initial multivariable model. Subsequently, factors  
29 demonstrating mediation were excluded from the final model. We assessed the model fit utilizing  
30 the Hosmer-Lemeshow goodness of fit test, and crude and adjusted odds ratios (OR and aOR)  
31 and 95% confidence intervals (CI) are reported. Post-estimation marginal probabilities of SARS-  
32 CoV-2 infection were determined from the final adjusted model for major covariates (race,  
33 ethnicity and age). We explored the mediation influence of comorbidity burden (CCI), socio-  
34 economic status (median income), and lack of social distancing (population density) on the  
35 relationship of Black race and Hispanic ethnicity with high likelihood of SARS-CoV-2 infection  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

1  
2  
3 using the Generalized Structural Equation Modeling (GSEM) framework. The GSEM framework  
4 was set up to provide estimates of direct and indirect effect of Black race and Hispanic ethnicity  
5 on SARS-CoV-2 infectivity. Statistically significant ( $p < 0.05$ ) indirect effects represent full or  
6 partial mediation by a tested covariate. We included all individuals tested for SARS-CoV-2  
7 across our healthcare system and did not perform formal sample size calculations.  
8  
9  
10  
11  
12  
13

### 14 **Patient and Public Involvement**

15  
16  
17 There was no direct patient or public involvement in the design and conduct of this study.  
18  
19  
20

## 21 **RESULTS**

### 22 **Socio-demographic and comorbidity characteristics of the study population**

23  
24  
25  
26 Across the time period of analysis, we identified a total of 20,228 presumed cases tested  
27 for SARS-CoV-2, among whom 1,551 (7.7%, CI: 7.3-8.0) tested positive. Overall, the mean  
28 (SD) age of the study population was 51.1 (19.0) years; 61.9% were female and 62.3% were  
29 White (including Hispanic ethnicity). The study sample was comparable to the overall  
30 population of patients treated across HM, who have a mean (SD) age of 49.0 (22) years, are 56%  
31 female, and 53% White. The HM system metrics was derived from a sample of 3,216,290  
32 patients managed across the system since May 22, 2016.  
33  
34  
35  
36  
37  
38  
39  
40  
41

42 The overall median (IQR) household income was USD \$70,658 (\$53,313–\$99,276), and  
43 42.6% of the study population had private or employer-based insurance. In our univariate  
44 analysis, Black race (vs. White; OR, CI: 1.55, 1.37–1.75), Hispanic ethnicity (vs. non-Hispanic;  
45 OR, CI: 2.02, 1.79–2.27), and males (vs. females; OR, CI: 1.17, 1.06–1.31) were associated with  
46 significantly higher likelihood of testing positive for SARS-CoV-2. Among the SARS-CoV-2  
47 positive patients, 40.8% were in the age category of 51–75 years, and 11.4% were greater than  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

1  
2  
3 75 years. These proportions were significantly higher than the reference group (up to 35 years;  
4  
5 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:  
6  
7 1.23, 1.02–1.49). Furthermore, individuals in higher percentiles of socio-economic status had  
8  
9 significantly lower likelihood, whereas those residing in higher population density ZCTAs had  
10  
11 higher likelihood of SARS-CoV-2 infection. We observed a significantly higher proportion of  
12  
13 SARS-CoV-2 positive individuals in the CCI 1-2 category compared to CCI of 0 (OR, CI: 1.35,  
14  
15 1.18–1.54). However, similar differences for higher CCI categories were not observed. For  
16  
17 specific comorbidities, a significantly greater proportion of diabetic individuals had SARS-CoV-  
18  
19 2 positive results (OR, CI: 1.40, 0.17–1.68). The socio-demographic characteristics and  
20  
21 comorbidity profiles for the overall and SARS-CoV-2 positive and negative patients are  
22  
23 summarized in Table 1.  
24  
25  
26  
27

### 28 **Socio-demographic and comorbidity characteristics associated with minority race and** 29 30 **ethnicity** 31 32

33 In our study sample comprising of 13,754 Non-Hispanic Black and White individuals, we  
34  
35 compared the association between race and various socio-demographic and comorbidity  
36  
37 characteristics (Table 2). Similarly, we also evaluated univariable differences for socio-  
38  
39 demographic variables and co-morbidities between Hispanic and non-Hispanic individuals  
40  
41 (Table 3). Minority race (NHB) and ethnicity (Hispanic) were both associated with younger age,  
42  
43 higher proportion of females, and residence in low income and higher population density  
44  
45 ZCTAs. However, NHB and Hispanic groups were both associated with an overall lower burden  
46  
47 of comorbidities (as demonstrated by significantly lower median CCI) compared respectively to  
48  
49 NHW and non-Hispanic categories. A higher proportion of individuals among minority race and  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60



1  
2  
3 ethnicity were diabetic, and a higher proportion of NHB were also hypertensive compared to  
4  
5 NHW.  
6

### 7 **Multivariable model and marginal probabilities for likelihood of SARS-CoV-2 infection** 8 9 **and racial and ethnic minorities**

10  
11  
12 The significantly higher likelihood of SARS-CoV-2 infection among minority race and  
13  
14 ethnic groups persisted after controlling for other demographics, insurance type, median  
15  
16 household income, population density, and comorbidities. Adjusted odds ratios (CI) for NHB vs.  
17  
18 NHW was 2.23 (1.90 – 2.60) and for Hispanic vs. Non-Hispanic was 1.95 (1.72 – 2.20). Higher  
19  
20 risk of infection among males (compared to females) and higher likelihood of SARS-CoV-2  
21  
22 infection among elderly also remained statistically significant. Detailed outputs of the fully  
23  
24 adjusted logistic regression models for minority race and ethnic groups are presented in Table 4.  
25  
26 Based on the marginal probabilities obtained from our fully adjusted model, the probability of  
27  
28 SARS-CoV-2 infection in a 45-year-old NHB is 9.6% whereas it is 4.5% in a 45-year-old NHW  
29  
30 individual, all other adjusted variables being constant. At the age of 75, this probability is 14.0%  
31  
32 for an NHB and 6.9% for a NHW. A similar relationship differential was observed for Hispanic  
33  
34 vs. non-Hispanic individuals. Multivariable model derived probabilities of SARS-CoV-2  
35  
36 infection for NHB vs. NHW and for Hispanic vs. Non-Hispanic across age spectrum are  
37  
38 presented in Figure 1 and Figure 2.  
39  
40  
41  
42  
43

### 44 **Generalized Structural Equation Modeling for mediation by income, population density** 45 46 **and Comorbidity Index**

47  
48  
49 Utilizing the GSEM framework, we determined the direct and indirect effects of NHB  
50  
51 and Hispanic ethnicity on SARS-CoV-2 infection with median income, population density and  
52  
53 CCI modeled as mediators in six separate equations adjusted for age and sex. The indirect effect  
54  
55  
56  
57  
58  
59  
60

1  
2  
3 of NHB mediated through population density was statistically significant (OR, CI: 1.03, 1.01 –  
4 1.05,  $p = 0.001$ ); however, the indirect effects mediated via median income and comorbidity  
5 scores were not statistically significant ( $p = 0.14$  and  $p = 0.64$  respectively). Among individuals  
6 identifying as Hispanic or Latino, both population density and income partially mediated the  
7 effect of ethnicity on SARS-CoV-2 positivity (OR, CI for population density: 1.02, 1.01 – 1.02,  
8  $p < 0.001$  and OR, CI for income: 1.04, 1.02 – 1.06,  $p < 0.001$ ). Evaluation of comorbidities did  
9 not suggest a mediation influence for either NHB or Hispanic categories.

## 20 **DISCUSSION**

21  
22  
23 The underlying race and ethnic healthcare disparities have been painfully highlighted in  
24 the wake of the COVID-19 pandemic. Most reports indicate higher mortality or case fatality  
25 among minority racial groups (Black / African American) across major U.S. metropolitan  
26 areas.<sup>11–13</sup> However, robust insights on the racial differences for SARS-CoV-2 infection are  
27 limited. Furthermore, comprehensive data evaluating higher susceptibility to SARS-CoV-2  
28 infection among Hispanic communities are also scarce. This is perhaps because of comparatively  
29 homogenous populations in non-U.S. regions of the world. Houston, as an exceptionally  
30 ethnically diverse population center,<sup>16</sup> is well suited for an investigation of racial, ethnic, and  
31 socioeconomic gradients in COVID-19 test positivity. We focus on highlighting the mechanisms  
32 of racial and ethnic disparities in susceptibility to SARS-CoV-2 infection and provide evidence  
33 of mediation of such disparities by novel social determinants of health (SDoH).

34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48 Our study adds to the current literature by analyzing emerging data for individuals being  
49 tested across one of the largest healthcare systems in the Greater Houston area. We report that  
50 racial and ethnic minorities (non-Hispanic Black and Hispanic individuals) are almost twice as  
51 likely to test positive for SARS-CoV-2 than the non-Hispanic White and non-Hispanic  
52  
53  
54  
55  
56  
57  
58  
59  
60

1  
2  
3 population. These findings illuminate systematic racial / ethnic disparities in testing positive for  
4 SARS-CoV-2 infection. Though there are limited prior SARS-CoV-2 data, such racial and ethnic  
5  
6 disparities have previously been described for the U.S. H1N1 influenza pandemic.<sup>17</sup> These data  
7  
8 indicated that Spanish-speaking Hispanic and Black individuals were at a greater risk of H1N1  
9  
10 infection, primarily attributable to lack of healthcare access.  
11  
12  
13

14  
15 We explored three possible mechanisms of race disparities in our data. These included  
16  
17 lower socio-economic status, residence in higher population dense areas, and higher level of  
18  
19 comorbidities. We demonstrate that NHB race is significantly associated with all three potential  
20  
21 disparity pathways, and in the traditional multivariable analyses, racial and ethnic disparities  
22  
23 persisted even after controlling for these pathways. However, our mediation analyses highlighted  
24  
25 the potential influence of residence in high population density areas as a viable pathway that at  
26  
27 least partially explains the observed racial and ethnic disparity. Furthermore, residence in low  
28  
29 income areas emerged as a significant mediation pathway for ethnic differences in SARS-CoV-2  
30  
31 positivity. Pathways mediating the influence of comorbidity status did not demonstrate a  
32  
33 significant effect. We utilized population density as a marker for potential inability to maintain  
34  
35 adequate social distancing as it has been indicated that maintaining the WHO recommended safe  
36  
37 distance between people becomes challenging with high population densities.<sup>18</sup> Furthermore,  
38  
39 overall effects of population density and disease spread has been previously described in  
40  
41 literature.<sup>19,20</sup> In addition to lack of social distancing, higher population density may also be  
42  
43 associated with several other behavioral and socio-demographic attributes that may predispose  
44  
45 populations to both viral spread and increased susceptibility. For example, there are reports  
46  
47 linking obesity, lack of physical activity, and higher mortality with residence in densely  
48  
49 populated neighborhoods.<sup>21,22</sup>  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

1  
2  
3 As reported, our data also corroborate that older populations may be more susceptible to  
4 SARS-CoV-2 infection.<sup>10</sup> However, younger populations still have cause for concern as nearly 1  
5  
6 in 4 of the infected cases in our sample were between 36–50 years of age. Finally, our data  
7  
8 demonstrate that males may be approximately 20% more likely to test positive for the SARS-  
9  
10 CoV-2 infection. Potential sex differences in infectivity to SARS-CoV-2 and intersectionality  
11  
12 with racial and ethnic socioeconomic factors need to be explored further in future analyses.  
13  
14 Additional policy-oriented research should prioritize studying the intersectionality of these  
15  
16 vulnerable economic statuses and racial disparities in COVID infection indicated by the present  
17  
18 study.  
19  
20  
21  
22  
23

24 Findings of our study need to be interpreted in the light of certain limitations. Our data  
25  
26 are from a single center and may not be generalizable to the wider U.S. population. These  
27  
28 findings need to be replicated in larger data sets across other large heterogenous U.S.  
29  
30 metropolitans. However, the Houston metropolitan area is one of the most diverse and  
31  
32 representative in the U.S.,<sup>16</sup> and our healthcare system is one of the largest systems providing  
33  
34 care to COVID-19 patients in the Greater Houston area. Our sample was composed of 22%  
35  
36 Black, 18% Hispanic, and 62% female population. We did not have information on certain  
37  
38 demographic covariates such as education or household size. Educational status has been linked  
39  
40 to healthcare awareness and may be important to adjust for in analyses of potential disparities,  
41  
42 and household size may be used to provide more precise estimates of socio-economic status.  
43  
44 However, we obtained and adjusted for zip code income data from the U.S. Census, as income  
45  
46 has previously been shown to have strong correlation with educational attainment and socio-  
47  
48 economic status.<sup>23</sup> Since testing was based on suspicion of infection and may have been  
49  
50 influenced by factors such as access to care, the potential for selection bias cannot be ruled out.  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

1  
2  
3 Furthermore, lack of sensitivity of SARS-CoV-2 diagnostics tests have been reported; however,  
4 the three assays utilized for testing were cross validated for internal consistency. Finally, we did  
5 not have detailed information on comorbidities and their management in the study population.  
6  
7 However, we did control for major comorbidities which are being reported as associated with  
8 COVID-19 outcomes.<sup>24</sup>  
9  
10  
11  
12  
13  
14  
15  
16

## 17 **Conclusions**

18  
19 The strong association between racial and ethnic minorities and SARS-CoV-2 infection  
20 demonstrated in our data, even after adjustment for other important socio-demographic and  
21 comorbidity factors, highlight a potential catastrophe of inequality within the existential crisis of  
22 a global pandemic. Our data, representing a large heterogeneous U.S. metropolitan area, also  
23 provide preliminary evidence into the potential pathways for this disparity. It is highly likely that  
24 higher comorbidity burden and detrimental effects of adverse social determinants, including  
25 those that may not adequately permit safe practices of social distancing, mediate higher SARS-  
26 CoV-2 infectivity among racial and ethnic minorities.  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36

37  
38 As the pandemic continues to spread and evolve across the continental U.S., emerging  
39 data on association between SARS-CoV-2 infection and various socio-demographic factors will  
40 continue to enhance our understanding of targeted risks related to SARS-CoV-2 infection, and  
41 such data would enable us to comprehend healthcare services and access factors related to  
42 development and outcomes of COVID-19 among minority populations. Our findings substantiate  
43 prior calls for collection of robust data on race and ethnicity as a part of international  
44 collaborations,<sup>25</sup> and further drive home the critical importance of quantifying novel SDoH.  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

1  
2  
3 **Acknowledgment:** The authors thank Jacob M. Kolman, Senior Scientific Writer of the Houston  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

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

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

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

For peer review only

## 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](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->



- 1  
2  
3 deaths-demographics-lightfoot-20200406-77nlylhiavgjzb2wa4ckivh7mu-story.html.  
4 Published April 7, 2020. Accessed April 20, 2020.  
5  
6  
7 13. Thebault R, Tran AB, Williams V. The coronavirus is infecting and killing black Americans  
8 at an alarmingly high rate. *Washington Post*.  
9 [https://www.washingtonpost.com/nation/2020/04/07/coronavirus-is-infecting-killing-black-](https://www.washingtonpost.com/nation/2020/04/07/coronavirus-is-infecting-killing-black-americans-an-alarmingly-high-rate-post-analysis-shows/)  
10 [americans-an-alarmingly-high-rate-post-analysis-shows/](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  
11 April 20, 2020.  
12  
13 14. United States Census Bureau. American Community Survey (ACS).  
14 <https://www.census.gov/programs-surveys/acs>. Published 2018. Accessed April 14, 2020.  
15  
16 15. United States Census Bureau. Gazetteer Files.  
17 <https://www.census.gov/geographies/reference-files/time-series/geo/gazetteer-files.html>.  
18 Accessed April 20, 2020.  
19  
20 16. Harris County Public Health. *Harris Cares: A 2020 Vision of Health in Harris County*.  
21 Houston, TX: Harris County Public Health; 2019.  
22 <http://publichealth.harriscountytexas.gov/Resources/Harris-Cares>. Accessed November 21,  
23 2019.  
24  
25 17. Quinn SC, Kumar S, Freimuth VS, Musa D, Casteneda-Angarita N, Kidwell K. Racial  
26 disparities in exposure, susceptibility, and access to health care in the US H1N1 influenza  
27 pandemic. *Am J Public Health*. 2011;101(2):285-293. doi:10.2105/AJPH.2009.188029  
28  
29 18. Rocklov J, Sjodin H. High population densities catalyze the spread of COVID-19. *J Travel*  
30 *Med*. 2020. doi: 10.1093/jtm/taaa038  
31  
32 19. Tarwater PM, Martin CF. Effects of population density on the spread of disease. *Complexity*.  
33 2001;6(6):29-36. doi:10.1002/cplx.10003  
34  
35 20. Yang K, LeJeune J, Alsdorf D, Lu B, Shum CK, Liang S. Global distribution of outbreaks of  
36 water-associated infectious diseases. *PLoS Negl Trop Dis*. 2012;6(2):e1483.  
37 doi:10.1371/journal.pntd.0001483  
38  
39 21. Wang R, Feng Z, Xue D, Liu Y, Wu R. Exploring the links between population density,  
40 lifestyle, and being overweight: secondary data analyses of middle-aged and older Chinese  
41 adults. *Health Qual Life Outcomes*. 2019;17(1):100. doi:10.1186/s12955-019-1172-3  
42  
43 22. Beenackers MA, Oude Groeniger J, Kamphuis CBM, Van Lenthe FJ. Urban population  
44 density and mortality in a compact Dutch city: 23-year follow-up of the Dutch GLOBE  
45 study. *Health & Place*. 2018;53:79-85. doi:10.1016/j.healthplace.2018.06.010  
46  
47 23. U.S. Bureau of Labor Statistics. Unemployment rates and earnings by educational  
48 attainment. <https://www.bls.gov/emp/chart-unemployment-earnings-education.htm>.  
49 Published September 4, 2019. Accessed April 15, 2020.  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

- 1  
2  
3 24. Centers for Disease Control and Prevention. Groups at higher risk for severe illness. cdc.gov.  
4 [https://www.cdc.gov/coronavirus/2019-ncov/need-extra-precautions/groups-at-higher-](https://www.cdc.gov/coronavirus/2019-ncov/need-extra-precautions/groups-at-higher-risk.html)  
5 [risk.html](https://www.cdc.gov/coronavirus/2019-ncov/need-extra-precautions/groups-at-higher-risk.html). Last updated April 17, 2020. Accessed April 20, 2020.  
6  
7  
8 25. Abuelgasim E, Saw LJ, Shirke M, Zeinah M, Harky A. COVID-19: Unique public health  
9 issues facing Black, Asian and minority ethnic communities. *Curr Probl Cardiol.*  
10 2020;100621. doi:10.1016/j.cpcardiol.2020.100621  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

**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

For peer review only

**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) <sup>a</sup>
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 (%) <sup>b</sup>				
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 (%) <sup>c</sup>	17.8	16.8	29.2	2.02 (1.79 – 2.27)
Median Zip Household Income (IQR) <sup>d</sup>	70,658 (53,313–99,276)	70,758 (53,633–100,107)	66,523 (50,485–94,226)	-4,168 <sup>e</sup> (-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 <sup>f</sup>	2797.2 (1439.1 – 4260.9)	2797.2 (1439.1 – 4211.4)	3320.3 (1904.4 – 4439.7)	523.1 <sup>e</sup> (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) <sup>f</sup>
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) <sup>a</sup>
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)

<sup>a</sup> Unadjusted Odds Ratios and 95% Confidence Intervals for association between individual co-variates in SARS-CoV-2 positivity.

<sup>b</sup> Race: Missing, Unknown, Declined, n = 1157 (5.7%).

<sup>c</sup> Ethnicity: Missing, Unknown, Declined, n = 613 (3.0%).

<sup>d</sup> 2018 inflation adjusted USD. Missing n = 1,883 (9.3%). Pentiles were defined by categorizing the ordered distribution of median income into five categories.

<sup>e</sup> Difference in median and 95% CI of difference obtained via quantile regression.

<sup>f</sup> 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 <sup>a</sup> / 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
<b>Age Category (%)</b>				
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
<b>Females</b>	68.1	58.1	1.54 (1.42 – 1.66)	< 0.001
<b>Median (IQR) Zip Income</b>	64,022 (47,303 – 79,658)	76,163 (60,130 – 102,019)	-12,141 (-14,018, -10,263)	< 0.001
<b>Median Zip Income Pentiles (%) – Pentiles of increasing Income<sup>b</sup></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
<b>Population Density for Zip: Median (IQR)</b>	3217.6 (2040.7 – 4439.7)	2488.5 (812.2 – 4084.3)	729.1 (603.0 – 855.2)	< 0.001
<b>Population Density for Zip Pentiles (%) – Pentiles of increasing population density<sup>c</sup></b>				
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
<b>Charlson Comorbidity Index, median (IQR)</b>	2 (0 – 6)	3 (1 – 7)	-1 (-1.22, -0.78)	< 0.001
<b>Charlson Comorbidity Index (CCI) Categories (%)</b>				
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

<sup>a</sup> 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.

<sup>b,c</sup> 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 <sup>a</sup> / Median Difference <sup>a</sup> (95% CI)	P value
Age: Mean (SD)	45.1 (18.3)	52.8 (18.9)	0.98 (0.98 – 0.98)	< 0.001
<b>Age Category (%)</b>				
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
<b>Females</b>	64.5	61.4	1.14 (1.06 – 1.23)	0.001
<b>Median (IQR) Zip Income</b>	65,742 (48,345 – 82,708)	73,742 (56,288 – 102,008)	-8,000 (-9,450.5, -6,549.5)	< 0.001
<b>Median Zip Income Pentiles (%) – Pentiles of increasing Income<sup>b</sup></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
<b>Population Density for Zip: Median (IQR)</b>	3256.8 (1504.0 – 4299.7)	2741.6 (1408.1 – 4110.7)	515.2 (469.2 – 561.2)	< 0.001
<b>Population Density for Zip Pentiles (%) – Pentiles of increasing population density<sup>c</sup></b>				
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
<b>Charlson Comorbidity Index, median (IQR)</b>	1 (0 – 4)	2 (0 – 6)	-1 (-1.21, -0.79)	< 0.001
<b>Charlson Comorbidity Index (CCI) Categories (%)</b>				
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

<sup>a</sup> 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.

<sup>b,c</sup> 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) <sup>a</sup>	Hispanic vs. Non-Hispanic, aOR (95% CI) <sup>b</sup>
<b>Non-Hispanic Black (vs. Non-Hispanic White)</b>	<b>2.23 (1.90 – 2.60)<sup>c</sup></b>	
<b>Hispanic (vs. Non-Hispanic)</b>		<b>1.95 (1.72 – 2.20)<sup>c</sup></b>
<b>Age Categories</b>		
Up to 35 years	<i>Reference Category</i>	
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)
<b>Male (vs. Female)</b>	1.20 (1.04 – 1.39)	1.17 (1.05 – 1.32)
<b>Median Zip Household Income Categories (Pentiles of Increasing Income)</b>		
Category I	<i>Reference Category</i>	
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)	
<b>Primary Insurance Type</b>		
Medicare	<i>Reference Category</i>	
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)
<b>Charlson Comorbidity Index Categories</b>		
CCI: 0	<i>Reference Category</i>	
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)

<sup>a,b</sup> Hosmer and Lemeshow goodness of fit p-value: 0.58 and 0.73 ( $H_0$ : Model fit is correct).

<sup>c</sup> 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.



1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

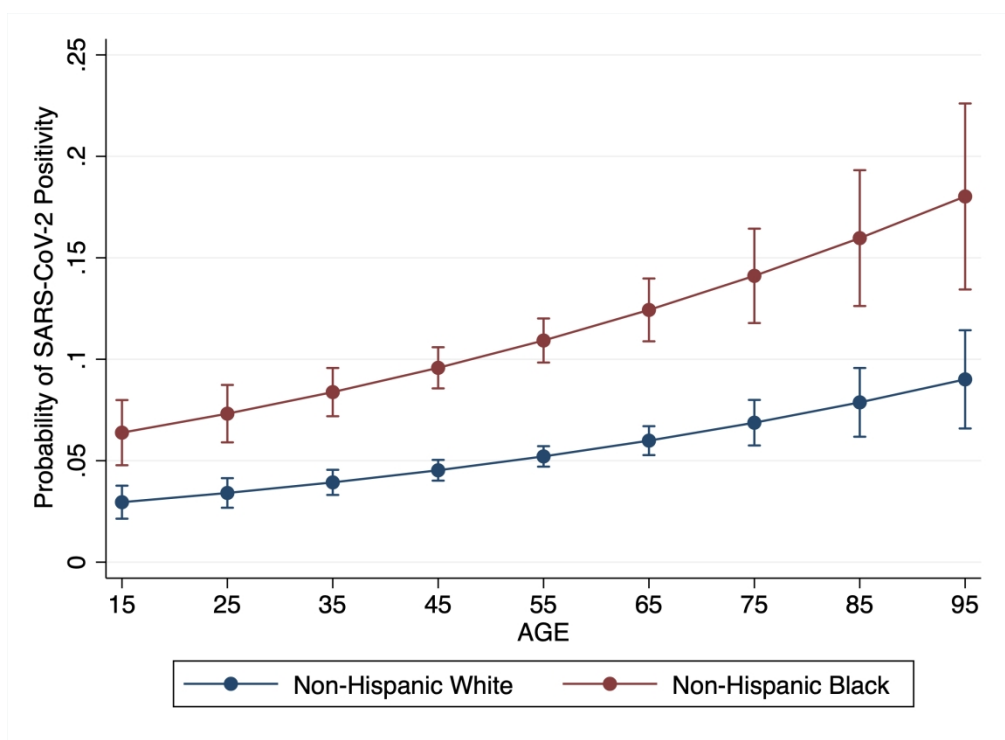


Figure 1: Adjusted Probability and 95% Confidence Interval of Positive SARS-CoV-2 PCR in Non-Hispanic Black vs. Non-Hispanic White by increasing age

467x339mm (144 x 144 DPI)

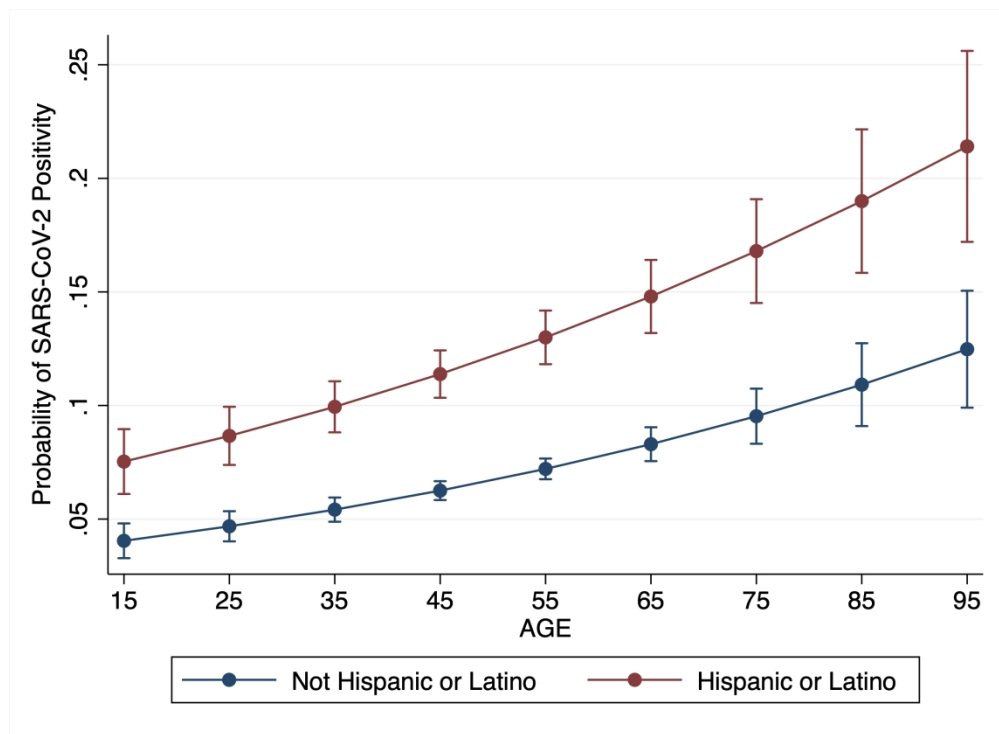


Figure 2: Adjusted Probability and 95% Confidence Interval of Positive SARS-CoV-2 PCR in Hispanic vs. Non-Hispanic by increasing age

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
<b>Title and abstract</b>	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
<b>Introduction</b>			
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
<b>Methods</b>			
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
<b>Results</b>			
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
<b>Discussion</b>			
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
<b>Other information</b>			
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](http://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:	<i>BMJ Open</i>
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, Bit A; Houston Methodist Research Institute; Texas A&M University School of Rural Public Health Nasir, Khurram; Houston Methodist Research Institute; Houston Methodist Hospital
<b>Primary Subject Heading</b>:	Epidemiology
Secondary Subject Heading:	Epidemiology, Infectious diseases, Public health
Keywords:	EPIDEMIOLOGY, Public health < INFECTIOUS DISEASES, INFECTIOUS DISEASES

SCHOLARONE™  
Manuscripts



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.

1  
2  
3 **Racial and Ethnic Disparities in SARS-CoV-2 Pandemic: Analysis of a COVID-19**  
4 **Observational Registry for a Diverse U.S. Metropolitan Population**  
5  
6  
7

8 *Farhaan S. Vahidy PhD MBBS MPH,<sup>1,2</sup> Juan Carlos Nicolas BS,<sup>1</sup> Jennifer R. Meeks MS,<sup>1</sup>*  
9 *Osman Khan BS,<sup>1</sup> Alan Pan MS,<sup>1</sup> Stephen L. Jones MD MSHI,<sup>1,3,4</sup> Faisal N. Masud MD,<sup>4,5</sup>*  
10 *H. Dirk Sostman MD,<sup>4,6</sup> Robert A. Phillips MD PhD,<sup>1,4,7</sup> Julia D. Andrieni MD,<sup>4,8</sup>*  
11 *Bitu Kash PhD MBA,<sup>1,4,9</sup> Khurram Nasir MD MPH<sup>1,7</sup>*  
12  
13  
14  
15

- 16 1. Center for Outcomes Research, The Houston Methodist Research Institute, Houston TX  
17 2. Houston Methodist Neurological Institute, Houston Methodist Hospital, Houston TX  
18 3. Department of Surgery, Houston Methodist Hospital, Houston TX  
19 4. Weill Cornell Medical College, New York NY  
20 5. Department of Anesthesiology and Critical Care, Houston Methodist Hospital, Houston  
21 TX  
22 6. The Houston Methodist Research Institute and The Houston Methodist Academic  
23 Institute, Houston TX  
24 7. Department of Cardiology, DeBakey Heart and Vascular Center, Houston Methodist  
25 Hospital, Houston TX  
26 8. Department of Medicine, Houston Methodist Hospital, Houston TX  
27 9. Department of Health Policy and Management, Texas A&M University School of Public  
28 Health, College Station TX  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39

40 **Corresponding Author**

41  
42 **Farhaan S. Vahidy, PhD MBBS MPH**

43 Josie Roberts Building, Suite 4.123

44 Houston Methodist

45 7550 Greenbriar Drive

46 Houston TX 77030

47 [fvahidy@houstonmethodist.org](mailto:fvahidy@houstonmethodist.org)  
48  
49  
50

51  
52  
53  
54  
55 Word Count: 3,093  
56  
57  
58  
59  
60

## 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



1  
2  
3 demonstrated a significant indirect effect of NHB and Hispanic ethnicity on SARS-CoV-2  
4  
5 infection mediated via a pathway including residence in densely populated zip code.  
6

7  
8 **Conclusions:** There is strong evidence of race and ethnic disparities in the SARS-CoV-2  
9  
10 pandemic, that is potentially mediated through unique social determinants of health.  
11

### 12 13 **Strengths and limitations of this study**

- 14  
15  
16 • One of the first studies to systematically evaluate race and ethnic disparities in  
17  
18 susceptibility to SARS-CoV-2 infection, while accounting for multiple socio-  
19  
20 demographic characteristics and comorbidities  
21
- 22  
23 • Study population represents a large and diverse metropolitan of the U.S. with data from  
24  
25 one of the largest healthcare providers across the greater metropolitan area  
26
- 27  
28 • Study evaluates potential mediation pathways for race disparities and demonstrates that  
29  
30 residence in areas with high population density may mediate race and ethnic disparities in  
31  
32 susceptibility to SARS-CoV-2 infection  
33
- 34  
35 • Single center study with limited information about burden of comorbidity and lifestyle  
36  
37 factors  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

## 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.<sup>1,2</sup> 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.<sup>3</sup> 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.<sup>4</sup> 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.<sup>5-6</sup>

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.<sup>7-10</sup> 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.<sup>11-13</sup> 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

1  
2  
3 areas are limited. Furthermore, the mediators of SARS-CoV-2 infection among racial and ethnic  
4  
5 minorities have not been described.  
6

7  
8 We explored socio-demographic characteristics such as age, sex, race, ethnicity, median  
9  
10 household income by zip codes, population density of residents' zip codes, and health insurance  
11  
12 status associated with positive SARS-CoV-2 testing in an urban and diverse population served  
13  
14 by one of the leading healthcare systems of the Greater Houston area. We further examined the  
15  
16 association between pre-existing comorbidities and higher likelihood of SARS-CoV-2 infection  
17  
18 in our study population. We hypothesized that older age, and racial and ethnic minorities will be  
19  
20 associated with significantly higher likelihood of SARS-CoV-2 infection, and factors such as  
21  
22 low socio-economic status, residence in high population density areas (proxy for potential  
23  
24 difficulties in social distancing) and higher comorbidity burden will mediate the effect of race  
25  
26 and ethnicity on SARS-CoV-2 infection.  
27  
28  
29  
30  
31  
32

### 33 **METHODS**

34  
35 We analyzed data between March 5 and May 31, 2020 collected as a part of the COVID-  
36  
37 19 Surveillance and Outcomes Registry (CURATOR) at Houston Methodist (HM). The HM  
38  
39 CURATOR has been approved by the HM Institutional Review Board (IRB) as an observational  
40  
41 quality of care registry for all suspected and confirmed COVID-19 patients. HM IRB granted  
42  
43 CURATOR a waiver of informed consent and HIPAA (Health Insurance Portability and  
44  
45 Accountability Act) authorization in accordance with current federal regulations. The  
46  
47 CURATOR, designed and managed by the big data team at the Center for Outcomes Research  
48  
49 (COR) at HM, is populated from multiple data sources across the HM system such as electronic  
50  
51 medical records, electronic databanks for laboratory and pharmacy, and electronic interactive  
52  
53  
54  
55  
56  
57  
58  
59  
60

1  
2  
3 patient interface tools. The HM system comprises a flagship tertiary care hospital in the Texas  
4  
5 Medical Center, seven large community hospitals, a continuing care hospital, and multiple  
6  
7 emergency centers and clinics throughout the Greater Houston area. Data from various sources  
8  
9 are curated into a harmonized format, assessed for quality and integrity, and stored on a secure  
10  
11 institutional HIPAA-compliant server.  
12  
13

14  
15 We flagged all individuals who were tested for the SARS-CoV-2 using the real time  
16  
17 Reverse Transcriptase (RT) Polymerized Chain Reaction (PCR) diagnostic panels. The three  
18  
19 cross-validated PCR tests utilized were the World Health Organization (WHO) nucleic acid  
20  
21 amplification test, Panther Fusion<sup>®</sup> SARS-CoV-2 Assay, and Cepheid Xpert<sup>®</sup> Xpress SARS-  
22  
23 CoV-2 Assay. These assays were verified for quantitative detection of novel SARS-CoV-2  
24  
25 isolated and purified from nasopharyngeal swab specimens obtained from individuals and  
26  
27 immersed in universal transport medium. Testing was carried out for symptomatic individuals or  
28  
29 for individuals who had a self-reported history of exposure to a COVID-19 case including recent  
30  
31 travel to other countries with high infection rates or hotspots within the U.S.  
32  
33

34  
35 Socio-demographic characteristics including age, sex, race, ethnicity, and payer-status  
36  
37 (insurance type) were obtained from the HM CURATOR for analyses. We also extracted  
38  
39 information on presence of comorbidities comprising the Charlson Comorbidity Index (CCI)  
40  
41 which include past history of myocardial infarction, congestive heart failure, peripheral vascular  
42  
43 disease, cerebrovascular disease, dementia, chronic pulmonary disease, rheumatic disease, peptic  
44  
45 ulcer disease, liver disease, diabetes with or without complications, hemiplegia, renal disease,  
46  
47 any malignancy (excluding skin neoplasms), metastatic solid tumors, and AIDS/HIV. Data on  
48  
49 hypertension and obesity were additionally obtained. We utilized the U.S. Census Bureau's  
50  
51 American Community Survey (ACS) 5-year data (2014–2018) to determine median household  
52  
53  
54  
55  
56  
57  
58  
59

1  
2  
3 income by individual zip code tabulation areas (ZCTA).<sup>14</sup> The median ZCTA household income  
4 was inflation-adjusted to 2018 USD. We also utilized the same data source to obtain population  
5 estimates by ZCTA, and calculated ZCTA level population density (population per mile square)  
6 by standardizing it for area measurements of ZCTA. For the purpose of population density  
7 determination, land area estimates were obtained from the Census Bureau's U.S. Gazetteer Files  
8 2010.<sup>15</sup> In the absence of granular and precise social distancing data, we have utilized population  
9 density as a proxy for potential difficulties in social distancing among crowded communities.

19 We provide descriptive summary data as means (standard deviations) and proportions.  
21 We fit univariable and multivariable logistic regression models to assess unadjusted and adjusted  
22 association between socio-demographic characteristics and likelihood of being tested positive for  
23 SARS-CoV-2. We additionally provide univariable comparison of various socio-demographic  
24 and comorbidity variables between non-Hispanic Black (NHB) and non-Hispanic White (NHW)  
25 race categories, as well as between Hispanic and Non-Hispanic ethnic groups. Age, income,  
26 population density and CCI were categorized for certain analyses. We determined *a priori* to  
27 include all variables (age, sex, race, ethnicity, zip code household income, insurance type, zip  
28 population density and CCI) in our initial multivariable model. Subsequently, factors  
29 demonstrating mediation were excluded from the final models. However, the factors that did not  
30 demonstrate mediation were included in the final models, as we believe that they continue to  
31 importantly inform the variance of estimates for direct effects.<sup>16</sup> We assessed the model fit  
32 utilizing the Hosmer-Lemeshow goodness of fit test, and crude and adjusted odds ratios (OR and  
33 aOR) and 95% confidence intervals (CI) are reported. Post-estimation marginal probabilities of  
34 SARS-CoV-2 infection were determined from the final adjusted model for major covariates  
35 (race, ethnicity and age). We explored the mediation influence of comorbidity burden (CCI),  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

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

1  
2  
3 significantly higher likelihood of testing positive for SARS-CoV-2. Among the SARS-CoV-2  
4 positive patients, 40.8% were in the age category of 51–75 years, and 11.4% were greater than  
5 75 years. These proportions were significantly higher than the reference group (up to 35 years;  
6 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:  
7 1.23, 1.02–1.49). Furthermore, individuals in higher percentiles of socio-economic status had  
8 significantly lower likelihood, whereas those residing in higher population density ZCTAs had  
9 higher likelihood of SARS-CoV-2 infection. We observed a significantly higher proportion of  
10 SARS-CoV-2 positive individuals in the CCI 1-2 category compared to CCI of 0 (OR, CI: 1.35,  
11 1.18–1.54). However, similar differences for higher CCI categories were not observed. For  
12 specific comorbidities, a significantly greater proportion of diabetic individuals had SARS-CoV-  
13 2 positive results (OR, CI: 1.40, 0.17–1.68). The socio-demographic characteristics and  
14 comorbidity profiles for the overall and SARS-CoV-2 positive and negative patients are  
15 summarized in Table 1.

### 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 **Socio-demographic and comorbidity characteristics associated with minority race and** 34 **ethnicity** 35

36  
37 In our study sample comprising of 13,754 Non-Hispanic Black and White individuals, we  
38 compared the association between race and various socio-demographic and comorbidity  
39 characteristics (Table 2). Similarly, we also evaluated univariable differences for socio-  
40 demographic variables and co-morbidities between Hispanic and non-Hispanic individuals  
41 (Table 3). Minority race (NHB) and ethnicity (Hispanic) were both associated with younger age,  
42 higher proportion of females, and residence in low income and higher population density  
43 ZCTAs. However, NHB and Hispanic groups were both associated with an overall lower burden  
44 of comorbidities (as demonstrated by significantly lower median CCI) compared respectively to  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

1  
2  
3 NHW and non-Hispanic categories. A higher proportion of individuals among minority race and  
4  
5 ethnicity were diabetic, and a higher proportion of NHB were also hypertensive compared to  
6  
7 NHW.  
8  
9

### 10 **Multivariable model and marginal probabilities for likelihood of SARS-CoV-2 infection** 11 12 **and racial and ethnic minorities** 13

14  
15 The significantly higher likelihood of SARS-CoV-2 infection among minority race and  
16  
17 ethnic groups persisted after controlling for other demographics, insurance type, median  
18  
19 household income, population density, and comorbidities. Adjusted odds ratios (CI) for NHB vs.  
20  
21 NHW was 2.23 (1.90 – 2.60) and for Hispanic vs. Non-Hispanic was 1.95 (1.72 – 2.20). Higher  
22  
23 risk of infection among males (compared to females) and higher likelihood of SARS-CoV-2  
24  
25 infection among elderly also remained statistically significant. Detailed outputs of the fully  
26  
27 adjusted logistic regression models for minority race and ethnic groups are presented in Table 4.  
28  
29 Based on the marginal probabilities obtained from our fully adjusted model, the probability of  
30  
31 SARS-CoV-2 infection in a 45-year-old NHB is 9.6% whereas it is 4.5% in a 45-year-old NHW  
32  
33 individual, all other adjusted variables being constant. At the age of 75, this probability is 14.0%  
34  
35 for an NHB and 6.9% for a NHW. A similar relationship differential was observed for Hispanic  
36  
37 vs. non-Hispanic individuals. Multivariable model derived probabilities of SARS-CoV-2  
38  
39 infection for NHB vs. NHW and for Hispanic vs. Non-Hispanic across age spectrum are  
40  
41 presented in Figure 1 and Figure 2.  
42  
43  
44  
45

### 46 **Generalized Structural Equation Modeling for mediation by income, population density** 47 48 **and Comorbidity Index** 49

50  
51 Utilizing the GSEM framework, we determined the direct and indirect effects of NHB  
52  
53 and Hispanic ethnicity on SARS-CoV-2 infection with median income, population density and  
54  
55  
56  
57  
58  
59



1  
2  
3 CCI modeled as mediators in six separate equations adjusted for age and sex. The indirect effect  
4 of NHB mediated through population density was statistically significant (OR, CI: 1.03, 1.01 –  
5 1.05,  $p = 0.001$ ); however, the indirect effects mediated via median income and comorbidity  
6 scores were not statistically significant ( $p = 0.14$  and  $p = 0.64$  respectively). Among individuals  
7 identifying as Hispanic or Latino, both population density and income partially mediated the  
8 effect of ethnicity on SARS-CoV-2 positivity (OR, CI for population density: 1.02, 1.01 – 1.02,  
9  $p < 0.001$  and OR, CI for income: 1.04, 1.02 – 1.06,  $p < 0.001$ ). Evaluation of comorbidities did  
10 not suggest a mediation influence for either NHB or Hispanic categories.  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21

## 22 **DISCUSSION**

23  
24 The underlying race and ethnic healthcare disparities have been painfully highlighted in  
25 the wake of the COVID-19 pandemic. Most reports indicate higher mortality or case fatality  
26 among minority racial groups (Black / African American) across major U.S. metropolitan  
27 areas.<sup>11–13</sup> However, robust insights on the racial differences for SARS-CoV-2 infection are  
28 limited. Furthermore, comprehensive data evaluating higher susceptibility to SARS-CoV-2  
29 infection among Hispanic communities are also scarce. This is perhaps because of comparatively  
30 homogenous populations in non-U.S. regions of the world. Houston, as an exceptionally  
31 ethnically diverse population center,<sup>17</sup> is well suited for an investigation of racial, ethnic, and  
32 socioeconomic gradients in COVID-19 test positivity. We focus on highlighting the mechanisms  
33 of racial and ethnic disparities in susceptibility to SARS-CoV-2 infection and provide evidence  
34 of mediation of such disparities by novel social determinants of health (SDoH).  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49

50 Our study adds to the current literature by analyzing emerging data for individuals being  
51 tested across one of the largest healthcare systems in the Greater Houston area. We report that  
52 racial and ethnic minorities (non-Hispanic Black and Hispanic individuals) are almost twice as  
53  
54  
55  
56  
57  
58  
59

1  
2  
3 likely to test positive for SARS-CoV-2 than the non-Hispanic White and non-Hispanic  
4  
5 population. These findings illuminate systematic racial / ethnic disparities in testing positive for  
6  
7 SARS-CoV-2 infection. Though there are limited prior SARS-CoV-2 data, such racial and ethnic  
8  
9 disparities have previously been described for the U.S. H1N1 influenza pandemic.<sup>18</sup> These data  
10  
11 indicated that Spanish-speaking Hispanic and Black individuals were at a greater risk of H1N1  
12  
13 infection, primarily attributable to lack of healthcare access.  
14  
15

16  
17 We explored three possible mechanisms of race disparities in our data. These included  
18  
19 lower socio-economic status, residence in higher population dense areas, and higher level of  
20  
21 comorbidities. We demonstrate that NHB race is significantly associated with all three potential  
22  
23 disparity pathways, and in the traditional multivariable analyses, racial and ethnic disparities  
24  
25 persisted even after controlling for these pathways. However, our mediation analyses highlighted  
26  
27 the potential influence of residence in high population density areas as a viable pathway that at  
28  
29 least partially explains the observed racial and ethnic disparity. Furthermore, residence in low  
30  
31 income areas emerged as a significant mediation pathway for ethnic differences in SARS-CoV-2  
32  
33 positivity. Pathways mediating the influence of comorbidity status did not demonstrate a  
34  
35 significant effect. We utilized population density as a marker for potential inability to maintain  
36  
37 adequate social distancing as it has been indicated that maintaining the WHO recommended safe  
38  
39 distance between people becomes challenging with high population densities.<sup>19</sup> Furthermore,  
40  
41 overall effects of population density and disease spread has been previously described in  
42  
43 literature.<sup>20,21</sup> In addition to lack of social distancing, higher population density may also be  
44  
45 associated with several other behavioral and socio-demographic attributes that may predispose  
46  
47 populations to both viral spread and increased susceptibility. For example, there are reports  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

1  
2  
3 linking obesity, lack of physical activity, and higher mortality with residence in densely  
4  
5 populated neighborhoods.<sup>22,23</sup>  
6

7  
8 As reported, our data also corroborate that older populations may be more susceptible to  
9  
10 SARS-CoV-2 infection.<sup>10</sup> However, younger populations still have cause for concern as nearly 1  
11  
12 in 4 of the infected cases in our sample were between 36–50 years of age. Finally, our data  
13  
14 demonstrate that males may be approximately 20% more likely to test positive for the SARS-  
15  
16 CoV-2 infection. Potential sex differences in infectivity to SARS-CoV-2 and intersectionality  
17  
18 with racial and ethnic socioeconomic factors need to be explored further in future analyses.  
19  
20 Additional policy-oriented research should prioritize studying the intersectionality of these  
21  
22 vulnerable economic statuses and racial disparities in COVID infection indicated by the present  
23  
24 study.  
25  
26

27  
28 Findings of our study need to be interpreted in the light of certain limitations. Our data  
29  
30 are from a single center and may not be generalizable to the wider U.S. population. These  
31  
32 findings need to be replicated in larger data sets across other large heterogenous U.S.  
33  
34 metropolitans. However, the Houston metropolitan area is one of the most diverse and  
35  
36 representative in the U.S.<sup>17</sup> and our healthcare system is one of the largest systems providing  
37  
38 care to COVID-19 patients in the Greater Houston area. Our sample was composed of 22%  
39  
40 Black, 18% Hispanic, and 62% female population. We did not have information on certain  
41  
42 demographic covariates such as education or household size. Educational status has been linked  
43  
44 to healthcare awareness and may be important to adjust for in analyses of potential disparities,  
45  
46 and household size may be used to provide more precise estimates of socio-economic status.  
47  
48 However, we obtained and adjusted for zip code income data from the U.S. Census, as income  
49  
50 has previously been shown to have strong correlation with educational attainment and socio-  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

1  
2  
3 economic status.<sup>24</sup> Since testing was based on suspicion of infection and may have been  
4 influenced by factors such as access to care, the potential for selection bias cannot be ruled out.  
5  
6 Furthermore, lack of sensitivity of SARS-CoV-2 diagnostics tests have been reported; however,  
7  
8 the three assays utilized for testing were cross validated for internal consistency. Finally, we did  
9  
10 not have detailed information on comorbidities and their management in the study population.  
11  
12 However, we did control for major comorbidities which are being reported as associated with  
13  
14 COVID-19 outcomes.<sup>25</sup>  
15  
16  
17  
18  
19  
20

## 21 **Conclusions**

22  
23  
24 The strong association between racial and ethnic minorities and SARS-CoV-2 infection  
25 demonstrated in our data, even after adjustment for other important socio-demographic and  
26 comorbidity factors, highlight a potential catastrophe of inequality within the existential crisis of  
27 a global pandemic. Our data, representing a large heterogeneous U.S. metropolitan area, also  
28 provide preliminary evidence into the potential pathways for this disparity. It is highly likely that  
29 higher comorbidity burden and detrimental effects of adverse social determinants, including  
30 those that may not adequately permit safe practices of social distancing, mediate higher SARS-  
31 CoV-2 infectivity among racial and ethnic minorities.  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41

42 As the pandemic continues to spread and evolve across the continental U.S., emerging  
43 data on association between SARS-CoV-2 infection and various socio-demographic factors will  
44 continue to enhance our understanding of targeted risks related to SARS-CoV-2 infection, and  
45 such data would enable us to comprehend healthcare services and access factors related to  
46 development and outcomes of COVID-19 among minority populations. Our findings substantiate  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

1  
2  
3 prior calls for collection of robust data on race and ethnicity as a part of international  
4  
5 collaborations,<sup>26</sup> and further drive home the critical importance of quantifying novel SDoH.  
6  
7  
8  
9

10 **Acknowledgment:** The authors thank Jacob M. Kolman, Senior Scientific Writer of the Houston  
11  
12 Methodist Center for Outcomes Research, for reviewing the language and format of the  
13  
14 manuscript.  
15

16  
17 **IRB Approval:** This work was carried out under an approved protocol for the Houston  
18  
19 Methodist COVID-19 Surveillance and Outcomes Registry (HM CURATOR) by the Houston  
20  
21 Methodist Research Institute Institutional Review Board (HMRI IRB).  
22

23  
24 **Contribution Statement:**

25 FV: design, data analysis and interpretation, drafting the manuscript, critical revision for  
26  
27 important intellectual content, final approval  
28

29 JCN: data acquisition, data analysis, drafting the manuscript, final approval  
30

31 JRM: data acquisition, drafting the manuscript, final approval  
32

33 OK: data acquisition, data analysis, drafting the manuscript, final approval  
34

35 AP: data acquisition, data analysis, drafting the manuscript, final approval  
36

37 SLJ: data acquisition, data interpretation, critical revision for important intellectual content, final  
38 approval  
39

40 FNM: critical revision for important intellectual content, final approval  
41

42 HDS: critical revision for important intellectual content, final approval  
43

44 RAP: critical revision for important intellectual content, final approval  
45

46 JDA: critical revision for important intellectual content, final approval  
47

48 BK: critical revision for important intellectual content, final approval  
49

50 KN: design, interpretation of data, critical revision for important intellectual content, final  
51 approval  
52

53 **Competing Interests:** No authors declare a competing interest to this work  
54

55 **Funding:** There are no funders to report for this submission.  
56  
57  
58  
59  
60

1  
2  
3 **Date sharing statement:** All requests for de-identified data should be made to the corresponding  
4 author. All reasonable requests will be evaluated by the CURATOR Data Governance and  
5 Sharing Committee comprising of FV, SLJ, BK and KN in the light of institutional policies and  
6 guidelines.  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

## 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](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.harriscountytexas.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



- 1  
2  
3 24. U.S. Bureau of Labor Statistics. Unemployment rates and earnings by educational  
4 attainment. <https://www.bls.gov/emp/chart-unemployment-earnings-education.htm>.  
5 Published September 4, 2019. Accessed April 15, 2020.  
6  
7  
8 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).  
9 [https://www.cdc.gov/coronavirus/2019-ncov/need-extra-precautions/groups-at-higher-](https://www.cdc.gov/coronavirus/2019-ncov/need-extra-precautions/groups-at-higher-risk.html)  
10 [risk.html](https://www.cdc.gov/coronavirus/2019-ncov/need-extra-precautions/groups-at-higher-risk.html). Last updated April 17, 2020. Accessed April 20, 2020.  
11  
12 26. Abuelgasim E, Saw LJ, Shirke M, Zeinah M, Harky A. COVID-19: Unique public health  
13 issues facing Black, Asian and minority ethnic communities. *Curr Probl Cardiol*.  
14 2020;100621. doi:10.1016/j.cpcardiol.2020.100621  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

## 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

For peer review only

**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) <sup>a</sup>
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 (%) <sup>b</sup>				
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 (%) <sup>c</sup>	17.8	16.8	29.2	2.02 (1.79 – 2.27)
Median Zip Household Income (IQR) <sup>d</sup>	70,658 (53,313–99,276)	70,758 (53,633–100,107)	66,523 (50,485–94,226)	-4,168 <sup>e</sup> (-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 <sup>f</sup>	2797.2 (1439.1 – 4260.9)	2797.2 (1439.1 – 4211.4)	3320.3 (1904.4 – 4439.7)	523.1 <sup>e</sup> (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) <sup>f</sup>
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) <sup>a</sup>
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)

<sup>a</sup> Unadjusted Odds Ratios and 95% Confidence Intervals for association between individual co-variates in SRAS-CoV-2 positivity.

<sup>b</sup> Race: Missing, Unknown, Declined, n = 1157 (5.7%).

<sup>c</sup> Ethnicity: Missing, Unknown, Declined, n = 613 (3.0%).

<sup>d</sup> 2018 inflation adjusted USD. Missing n = 1,883 (9.3%). Pentiles were defined by categorizing the ordered distribution of median income into five categories.

<sup>e</sup> Difference in median and 95% CI of difference obtained via quantile regression.

<sup>f</sup> 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 <sup>a</sup> / 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
<b>Age Category (%)</b>				
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
<b>Females</b>	68.1	58.1	1.54 (1.42 – 1.66)	< 0.001
<b>Median (IQR) Zip Income</b>	64,022 (47,303 – 79,658)	76,163 (60,130 – 102,019)	-12,141 (-14,018, -10,263)	< 0.001
<b>Median Zip Income Pentiles (%) – Pentiles of increasing Income<sup>b</sup></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
<b>Population Density for Zip: Median (IQR)</b>	3217.6 (2040.7 – 4439.7)	2488.5 (812.2 – 4084.3)	729.1 (603.0 – 855.2)	< 0.001
<b>Population Density for Zip Pentiles (%) – Pentiles of increasing population density<sup>c</sup></b>				
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
<b>Charlson Comorbidity Index, median (IQR)</b>	2 (0 – 6)	3 (1 – 7)	-1 (-1.22, -0.78)	< 0.001
<b>Charlson Comorbidity Index (CCI) Categories (%)</b>				
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

<sup>a</sup> 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.

<sup>b,c</sup> 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 <sup>a</sup> / Median Difference <sup>a</sup> (95% CI)	P value
Age: Mean (SD)	45.1 (18.3)	52.8 (18.9)	0.98 (0.98 – 0.98)	< 0.001
<b>Age Category (%)</b>				
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
<b>Females</b>	64.5	61.4	1.14 (1.06 – 1.23)	0.001
<b>Median (IQR) Zip Income</b>	65,742 (48,345 – 82,708)	73,742 (56,288 – 102,008)	-8,000 (-9,450.5, -6,549.5)	< 0.001
<b>Median Zip Income Pentiles (%) – Pentiles of increasing Income<sup>b</sup></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
<b>Population Density for Zip: Median (IQR)</b>	3256.8 (1504.0 – 4299.7)	2741.6 (1408.1 – 4110.7)	515.2 (469.2 – 561.2)	< 0.001
<b>Population Density for Zip Pentiles (%) – Pentiles of increasing population density<sup>c</sup></b>				
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
<b>Charlson Comorbidity Index, median (IQR)</b>	1 (0 – 4)	2 (0 – 6)	-1 (-1.21, -0.79)	< 0.001
<b>Charlson Comorbidity Index (CCI) Categories (%)</b>				
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

<sup>a</sup> 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.

<sup>b,c</sup> 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) <sup>a</sup>	Hispanic vs. Non-Hispanic, aOR (95% CI) <sup>b</sup>
<b>Non-Hispanic Black (vs. Non-Hispanic White)</b>	<b>2.23 (1.90 – 2.60)<sup>c</sup></b>	
<b>Hispanic (vs. Non-Hispanic)</b>		<b>1.95 (1.72 – 2.20)<sup>c</sup></b>
<b>Age Categories</b>		
Up to 35 years	<i>Reference Category</i>	
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)
<b>Male (vs. Female)</b>	1.20 (1.04 – 1.39)	1.17 (1.05 – 1.32)
<b>Median Zip Household Income Categories (Pentiles of Increasing Income)</b>		
Category I	<i>Reference Category</i>	
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)	
<b>Primary Insurance Type</b>		
Medicare	<i>Reference Category</i>	
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)
<b>Charlson Comorbidity Index Categories</b>		
CCI: 0	<i>Reference Category</i>	
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)

<sup>a,b</sup> Hosmer and Lemeshow goodness of fit p-value: 0.58 and 0.73 ( $H_0$ : Model fit is correct).

<sup>c</sup> 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.

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

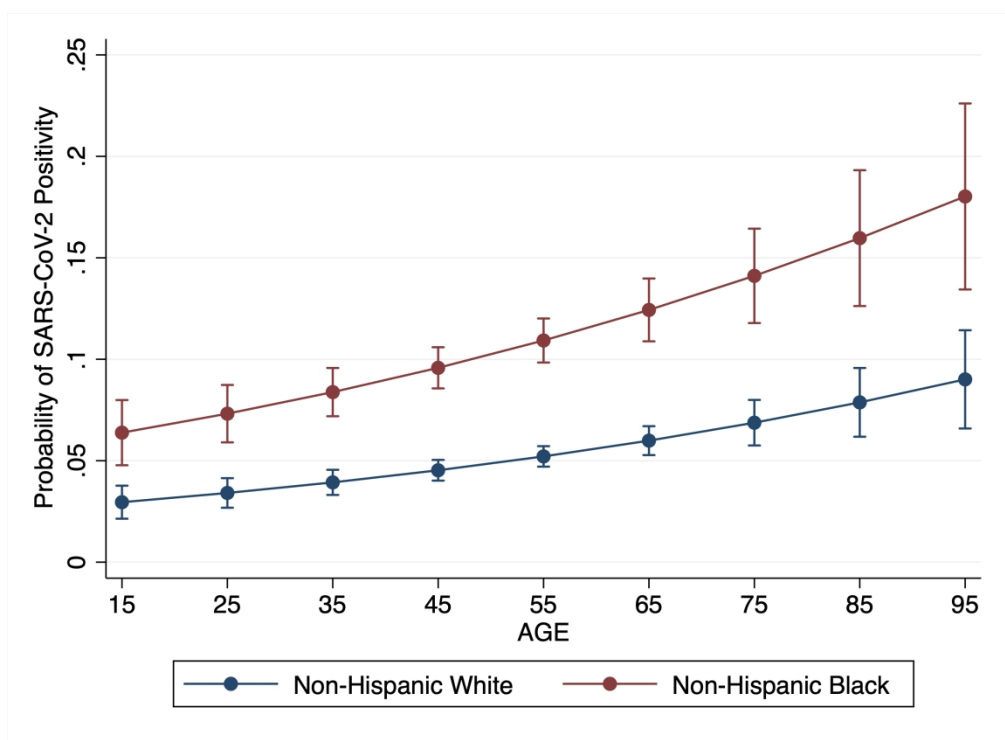


Figure 1: Adjusted Probability and 95% Confidence Interval of Positive SARS-CoV-2 PCR in Non-Hispanic Black vs. Non-Hispanic White by increasing age

467x339mm (144 x 144 DPI)



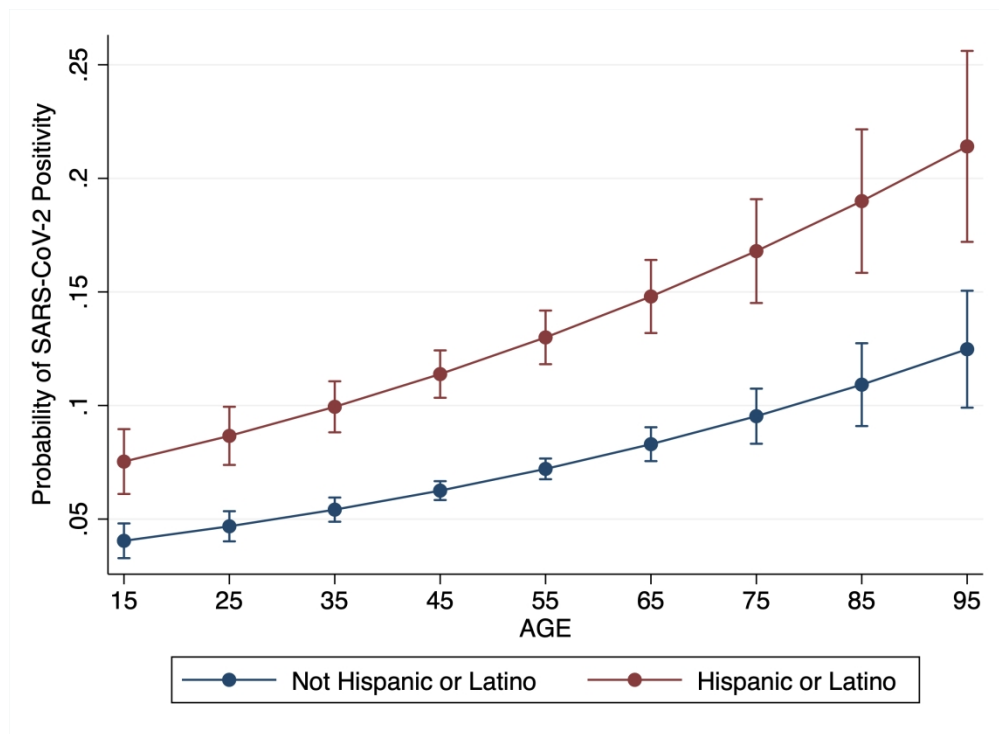


Figure 2: Adjusted Probability and 95% Confidence Interval of Positive SARS-CoV-2 PCR in Hispanic vs. Non-Hispanic by increasing age

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
<b>Title and abstract</b>	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
<b>Introduction</b>			
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
<b>Methods</b>			
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
<b>Results</b>			
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
<b>Discussion</b>			
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
<b>Other information</b>			
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](http://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:	<i>BMJ Open</i>
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, Bit A; Houston Methodist Research Institute; Texas A&M University School of Rural Public Health Nasir, Khurram; Houston Methodist Research Institute; Houston Methodist Hospital
<b>Primary Subject Heading</b>:	Epidemiology
Secondary Subject Heading:	Epidemiology, Infectious diseases, Public health
Keywords:	EPIDEMIOLOGY, Public health < INFECTIOUS DISEASES, INFECTIOUS DISEASES

SCHOLARONE™  
Manuscripts



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.

1  
2  
3 **Racial and Ethnic Disparities in SARS-CoV-2 Pandemic: Analysis of a COVID-19**  
4 **Observational Registry for a Diverse U.S. Metropolitan Population**  
5  
6  
7

8 *Farhaan S. Vahidy PhD MBBS MPH,<sup>1,2</sup> Juan Carlos Nicolas BS,<sup>1</sup> Jennifer R. Meeks MS,<sup>1</sup>*  
9 *Osman Khan BS,<sup>1</sup> Alan Pan MS,<sup>1</sup> Stephen L. Jones MD MSHI,<sup>1,3,4</sup> Faisal N. Masud MD,<sup>4,5</sup>*  
10 *H. Dirk Sostman MD,<sup>4,6</sup> Robert A. Phillips MD PhD,<sup>1,4,7</sup> Julia D. Andrieni MD,<sup>4,8</sup>*  
11 *Bitu Kash PhD MBA,<sup>1,4,9</sup> Khurram Nasir MD MPH<sup>1,7</sup>*  
12  
13  
14  
15

- 16 1. Center for Outcomes Research, The Houston Methodist Research Institute, Houston TX
- 17 2. Houston Methodist Neurological Institute, Houston Methodist Hospital, Houston TX
- 18 3. Department of Surgery, Houston Methodist Hospital, Houston TX
- 19 4. Weill Cornell Medical College, New York NY
- 20 5. Department of Anesthesiology and Critical Care, Houston Methodist Hospital, Houston
- 21 TX
- 22 6. The Houston Methodist Research Institute and The Houston Methodist Academic
- 23 Institute, Houston TX
- 24 7. Department of Cardiology, DeBakey Heart and Vascular Center, Houston Methodist
- 25 Hospital, Houston TX
- 26 8. Department of Medicine, Houston Methodist Hospital, Houston TX
- 27 9. Department of Health Policy and Management, Texas A&M University School of Public
- 28 Health, College Station TX
- 29
- 30
- 31
- 32
- 33
- 34
- 35
- 36
- 37
- 38
- 39

40 **Corresponding Author**

41  
42 **Farhaan S. Vahidy, PhD MBBS MPH**

43 Josie Roberts Building, Suite 4.123

44 Houston Methodist

45 7550 Greenbriar Drive

46 Houston TX 77030

47 [fvahidy@houstonmethodist.org](mailto:fvahidy@houstonmethodist.org)  
48  
49  
50

51  
52  
53  
54  
55 Word Count: 3,126  
56  
57  
58  
59  
60

## 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

1  
2  
3 demonstrated a significant indirect effect of NHB and Hispanic ethnicity on SARS-CoV-2  
4  
5 infection mediated via a pathway including residence in densely populated zip code.  
6

7  
8 **Conclusions:** There is strong evidence of race and ethnic disparities in the SARS-CoV-2  
9  
10 pandemic, that is potentially mediated through unique social determinants of health.  
11

### 12 13 **Strengths and limitations of this study**

- 14  
15  
16 • One of the first studies to systematically evaluate race and ethnic disparities in  
17  
18 susceptibility to SARS-CoV-2 infection, while accounting for multiple socio-  
19  
20 demographic characteristics and comorbidities  
21
- 22  
23 • Study population represents a large and diverse metropolitan of the U.S. with data from  
24  
25 one of the largest healthcare providers across the greater metropolitan area  
26
- 27  
28 • Study evaluates potential mediation pathways for race disparities and demonstrates that  
29  
30 residence in areas with high population density may mediate race and ethnic disparities in  
31  
32 susceptibility to SARS-CoV-2 infection  
33
- 34  
35 • Single center study with limited information about burden of comorbidity and lifestyle  
36  
37 factors  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60



## 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.<sup>1,2</sup> 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.<sup>3</sup> 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.<sup>4</sup> 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.<sup>5-6</sup>

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.<sup>7-10</sup> 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.<sup>11-13</sup> 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

1  
2  
3 areas are limited. Furthermore, the mediators of SARS-CoV-2 infection among racial and ethnic  
4  
5 minorities have not been described.  
6

7  
8 We explored socio-demographic characteristics such as age, sex, race, ethnicity, median  
9  
10 household income by zip codes, population density of residents' zip codes, and health insurance  
11  
12 status associated with positive SARS-CoV-2 testing in an urban and diverse population served  
13  
14 by one of the leading healthcare systems of the Greater Houston area. We further examined the  
15  
16 association between pre-existing comorbidities and higher likelihood of SARS-CoV-2 infection  
17  
18 in our study population. We hypothesized that older age, and racial and ethnic minorities will be  
19  
20 associated with significantly higher likelihood of SARS-CoV-2 infection, and factors such as  
21  
22 low socio-economic status, residence in high population density areas (proxy for potential  
23  
24 difficulties in social distancing) and higher comorbidity burden will mediate the effect of race  
25  
26 and ethnicity on SARS-CoV-2 infection.  
27  
28  
29  
30  
31  
32

### 33 **METHODS**

34  
35 We analyzed data between March 5 and May 31, 2020 collected as a part of the COVID-  
36  
37 19 Surveillance and Outcomes Registry (CURATOR) at Houston Methodist (HM). The HM  
38  
39 CURATOR has been approved by the HM Institutional Review Board (IRB) as an observational  
40  
41 quality of care registry for all suspected and confirmed COVID-19 patients. HM IRB granted  
42  
43 CURATOR a waiver of informed consent and HIPAA (Health Insurance Portability and  
44  
45 Accountability Act) authorization in accordance with current federal regulations. The  
46  
47 CURATOR, designed and managed by the big data team at the Center for Outcomes Research  
48  
49 (COR) at HM, is populated from multiple data sources across the HM system such as electronic  
50  
51 medical records, electronic databanks for laboratory and pharmacy, and electronic interactive  
52  
53  
54  
55  
56  
57  
58  
59  
60

1  
2  
3 patient interface tools. The HM system comprises a flagship tertiary care hospital in the Texas  
4  
5 Medical Center, seven large community hospitals, a continuing care hospital, and multiple  
6  
7 emergency centers and clinics throughout the Greater Houston area. Data from various sources  
8  
9 are curated into a harmonized format, assessed for quality and integrity, and stored on a secure  
10  
11 institutional HIPAA-compliant server.  
12  
13

14  
15 We flagged all individuals who were tested for the SARS-CoV-2 using the real time  
16  
17 Reverse Transcriptase (RT) Polymerized Chain Reaction (PCR) diagnostic panels. The three  
18  
19 cross-validated PCR tests utilized were the World Health Organization (WHO) nucleic acid  
20  
21 amplification test, Panther Fusion<sup>®</sup> SARS-CoV-2 Assay, and Cepheid Xpert<sup>®</sup> Xpress SARS-  
22  
23 CoV-2 Assay. These assays were verified for quantitative detection of novel SARS-CoV-2  
24  
25 isolated and purified from nasopharyngeal swab specimens obtained from individuals and  
26  
27 immersed in universal transport medium. Testing was carried out for symptomatic individuals or  
28  
29 for individuals who had a self-reported history of exposure to a COVID-19 case including recent  
30  
31 travel to other countries with high infection rates or hotspots within the U.S.  
32  
33

34  
35 Socio-demographic characteristics including age, sex, race, ethnicity, and payer-status  
36  
37 (insurance type) were obtained from the HM CURATOR for analyses. We also extracted  
38  
39 information on presence of comorbidities comprising the Charlson Comorbidity Index (CCI)  
40  
41 which include past history of myocardial infarction, congestive heart failure, peripheral vascular  
42  
43 disease, cerebrovascular disease, dementia, chronic pulmonary disease, rheumatic disease, peptic  
44  
45 ulcer disease, liver disease, diabetes with or without complications, hemiplegia, renal disease,  
46  
47 any malignancy (excluding skin neoplasms), metastatic solid tumors, and AIDS/HIV. Data on  
48  
49 hypertension and obesity were additionally obtained. We utilized the U.S. Census Bureau's  
50  
51 American Community Survey (ACS) 5-year data (2014–2018) to determine median household  
52  
53  
54  
55  
56  
57  
58  
59

1  
2  
3 income by individual zip code tabulation areas (ZCTA).<sup>14</sup> The median ZCTA household income  
4 was inflation-adjusted to 2018 USD. We also utilized the same data source to obtain population  
5 estimates by ZCTA, and calculated ZCTA level population density (population per mile square)  
6 by standardizing it for area measurements of ZCTA. For the purpose of population density  
7 determination, land area estimates were obtained from the Census Bureau's U.S. Gazetteer Files  
8 2010.<sup>15</sup> In the absence of granular and precise social distancing data, we have utilized population  
9 density as a proxy for potential difficulties in social distancing among crowded communities.

19 We provide descriptive summary data as means (standard deviations) and proportions.  
21 We fit univariable and multivariable logistic regression models to assess unadjusted and adjusted  
22 association between socio-demographic characteristics and likelihood of being tested positive for  
23 SARS-CoV-2. We additionally provide univariable comparison of various socio-demographic  
24 and comorbidity variables between non-Hispanic Black (NHB) and non-Hispanic White (NHW)  
25 race categories, as well as between Hispanic and Non-Hispanic ethnic groups. Age, income,  
26 population density and CCI were categorized for certain analyses. We included age, sex, race,  
27 ethnicity, zip code household income, insurance type, zip population density and CCI in our  
28 initial multivariable model. Zip code household income, zip population density and CCI were  
29 evaluated as mediators. Factors demonstrating mediation were excluded from the final models.  
30 However, the factors that did not demonstrate mediation were included in the final models, as we  
31 believe that they continue to importantly inform the variance of estimates for direct effects.<sup>16</sup> We  
32 assessed the model fit utilizing the Hosmer-Lemeshow goodness of fit test, and crude and  
33 adjusted odds ratios (OR and aOR) and 95% confidence intervals (CI) are reported. Post-  
34 estimation marginal probabilities of SARS-CoV-2 infection were determined from the final  
35 adjusted model for major covariates (race, ethnicity and age). We explored the mediation  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

1  
2  
3 influence of comorbidity burden (CCI), socio-economic status (median income), and lack of  
4 social distancing (population density) on the relationship of Black race and Hispanic ethnicity  
5 with high likelihood of SARS-CoV-2 infection using the Generalized Structural Equation  
6 Modeling (GSEM) framework. The GSEM framework was set up to provide estimates of direct  
7 and indirect effect of Black race and Hispanic ethnicity on SARS-CoV-2 infectivity. Statistically  
8 significant ( $p < 0.05$ ) indirect effects represent full or partial mediation by a tested covariate. We  
9 included all individuals tested for SARS-CoV-2 across our healthcare system and did not  
10 perform formal sample size calculations.  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20

### 21 **Patient and Public Involvement**

22  
23  
24 There was no direct patient or public involvement in the design and conduct of this study.  
25  
26  
27

## 28 **RESULTS**

### 29 **Socio-demographic and comorbidity characteristics of the study population**

30  
31  
32  
33 Across the time period of analysis, we identified a total of 20,228 presumed cases tested  
34 for SARS-CoV-2, among whom 1,551 (7.7%, CI: 7.3-8.0) tested positive. Overall, the mean  
35 (SD) age of the study population was 51.1 (19.0) years; 61.9% were female and 62.3% were  
36 White (including Hispanic ethnicity). The study sample was comparable to the overall  
37 population of patients treated across HM, who have a mean (SD) age of 49.0 (22) years, are 56%  
38 female, and 53% White. The HM system metrics was derived from a sample of 3,216,290  
39 patients managed across the system since May 22, 2016.  
40  
41  
42  
43  
44  
45  
46  
47  
48

49 The overall median (IQR) household income was USD \$70,658 (\$53,313–\$99,276), and  
50 42.6% of the study population had private or employer-based insurance. In our univariate  
51 analysis, Black race (vs. White; OR, CI: 1.55, 1.37–1.75), Hispanic ethnicity (vs. non-Hispanic;  
52  
53  
54  
55  
56  
57  
58  
59  
60

1  
2  
3 OR, CI: 2.02, 1.79–2.27), and males (vs. females; OR, CI: 1.17, 1.06–1.31) were associated with  
4 significantly higher likelihood of testing positive for SARS-CoV-2. Among the SARS-CoV-2  
5 positive patients, 40.8% were in the age category of 51–75 years, and 11.4% were greater than  
6 75 years. These proportions were significantly higher than the reference group (up to 35 years;  
7 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:  
8 1.23, 1.02–1.49). Furthermore, individuals in higher percentiles of socio-economic status had  
9 significantly lower likelihood, whereas those residing in higher population density ZCTAs had  
10 higher likelihood of SARS-CoV-2 infection. We observed a significantly higher proportion of  
11 SARS-CoV-2 positive individuals in the CCI 1-2 category compared to CCI of 0 (OR, CI: 1.35,  
12 1.18–1.54). However, similar differences for higher CCI categories were not observed. For  
13 specific comorbidities, a significantly greater proportion of diabetic individuals had SARS-CoV-  
14 2 positive results (OR, CI: 1.40, 0.17–1.68). The socio-demographic characteristics and  
15 comorbidity profiles for the overall and SARS-CoV-2 positive and negative patients are  
16 summarized in Table 1.

### 17 **Socio-demographic and comorbidity characteristics associated with minority race and** 18 **ethnicity**

19  
20 In our study sample comprising of 13,754 Non-Hispanic Black and White individuals, we  
21 compared the association between race and various socio-demographic and comorbidity  
22 characteristics (Table 2). Similarly, we also evaluated univariable differences for socio-  
23 demographic variables and co-morbidities between Hispanic and non-Hispanic individuals  
24 (Table 3). Minority race (NHB) and ethnicity (Hispanic) were both associated with younger age,  
25 higher proportion of females, and residence in low income and higher population density  
26 ZCTAs. However, NHB and Hispanic groups were both associated with an overall lower burden

1  
2  
3 of comorbidities (as demonstrated by significantly lower median CCI) compared respectively to  
4  
5 NHW and non-Hispanic categories. A higher proportion of individuals among minority race and  
6  
7 ethnicity were diabetic, and a higher proportion of NHB were also hypertensive compared to  
8  
9  
10 NHW.

### 11 12 **Multivariable model and marginal probabilities for likelihood of SARS-CoV-2 infection** 13 14 **and racial and ethnic minorities**

15  
16  
17 The significantly higher likelihood of SARS-CoV-2 infection among minority race and  
18  
19 ethnic groups persisted after controlling for other demographics, insurance type, median  
20  
21 household income, population density, and comorbidities. Adjusted odds ratios (CI) for NHB vs.  
22  
23 NHW was 2.23 (1.90 – 2.60) and for Hispanic vs. Non-Hispanic was 1.95 (1.72 – 2.20). Higher  
24  
25 risk of infection among males (compared to females) and higher likelihood of SARS-CoV-2  
26  
27 infection among elderly also remained statistically significant. Detailed outputs of the fully  
28  
29 adjusted logistic regression models for minority race and ethnic groups are presented in Table 4.  
30  
31  
32 Based on the marginal probabilities obtained from our fully adjusted model, the probability of  
33  
34 SARS-CoV-2 infection in a 45-year-old NHB is 9.6% whereas it is 4.5% in a 45-year-old NHW  
35  
36 individual, all other adjusted variables being constant. At the age of 75, this probability is 14.0%  
37  
38 for an NHB and 6.9% for a NHW. A similar relationship differential was observed for Hispanic  
39  
40 vs. non-Hispanic individuals. Multivariable model derived probabilities of SARS-CoV-2  
41  
42 infection for NHB vs. NHW and for Hispanic vs. Non-Hispanic across age spectrum are  
43  
44 presented in Figure 1 and Figure 2.  
45  
46  
47  
48

### 49 **Generalized Structural Equation Modeling for mediation by income, population density** 50 51 **and Comorbidity Index**

1  
2  
3 Utilizing the GSEM framework, we determined the direct and indirect effects of NHB  
4 and Hispanic ethnicity on SARS-CoV-2 infection with median income, population density and  
5 CCI modeled as mediators in six separate equations adjusted for age and sex. The indirect effect  
6 of NHB mediated through population density was statistically significant (OR, CI: 1.03, 1.01 –  
7 1.05,  $p = 0.001$ ); however, the indirect effects mediated via median income and comorbidity  
8 scores were not statistically significant ( $p = 0.14$  and  $p = 0.64$  respectively). Among individuals  
9 identifying as Hispanic or Latino, both population density and income partially mediated the  
10 effect of ethnicity on SARS-CoV-2 positivity (OR, CI for population density: 1.02, 1.01 – 1.02,  
11  $p < 0.001$  and OR, CI for income: 1.04, 1.02 – 1.06,  $p < 0.001$ ). Evaluation of comorbidities did  
12 not suggest a mediation influence for either NHB or Hispanic categories.  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26

## 27 **DISCUSSION**

28  
29 The underlying race and ethnic healthcare disparities have been painfully highlighted in  
30 the wake of the COVID-19 pandemic. Most reports indicate higher mortality or case fatality  
31 among minority racial groups (Black / African American) across major U.S. metropolitan  
32 areas.<sup>11-13</sup> However, robust insights on the racial differences for SARS-CoV-2 infection are  
33 limited. Furthermore, comprehensive data evaluating higher susceptibility to SARS-CoV-2  
34 infection among Hispanic communities are also scarce. This is perhaps because of comparatively  
35 homogenous populations in non-U.S. regions of the world. Houston, as an exceptionally  
36 ethnically diverse population center,<sup>17</sup> is well suited for an investigation of racial, ethnic, and  
37 socioeconomic gradients in COVID-19 test positivity. We focus on highlighting the mechanisms  
38 of racial and ethnic disparities in susceptibility to SARS-CoV-2 infection and provide evidence  
39 of mediation of such disparities by novel social determinants of health (SDoH).  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60



1  
2  
3 Our study adds to the current literature by analyzing emerging data for individuals being  
4 tested across one of the largest healthcare systems in the Greater Houston area. We report that  
5 racial and ethnic minorities (non-Hispanic Black and Hispanic individuals) are almost twice as  
6 likely to test positive for SARS-CoV-2 than the non-Hispanic White and non-Hispanic  
7 population. These findings illuminate systematic racial / ethnic disparities in testing positive for  
8 SARS-CoV-2 infection. Though there are limited prior SARS-CoV-2 data, such racial and ethnic  
9 disparities have previously been described for the U.S. H1N1 influenza pandemic.<sup>18</sup> These data  
10 indicated that Spanish-speaking Hispanic and Black individuals were at a greater risk of H1N1  
11 infection, primarily attributable to lack of healthcare access.  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22

23  
24 We explored three possible mechanisms of race disparities in our data. These included  
25 lower socio-economic status, residence in higher population dense areas, and higher level of  
26 comorbidities. We demonstrate that NHB race is significantly associated with all three potential  
27 disparity pathways, and in the traditional multivariable analyses, racial and ethnic disparities  
28 persisted even after controlling for these pathways. However, our mediation analyses highlighted  
29 the potential influence of residence in high population density areas as a viable pathway that at  
30 least partially explains the observed racial and ethnic disparity. Furthermore, residence in low  
31 income areas emerged as a significant mediation pathway for ethnic differences in SARS-CoV-2  
32 positivity. Pathways mediating the influence of comorbidity status did not demonstrate a  
33 significant effect. We utilized population density as a marker for potential inability to maintain  
34 adequate social distancing as it has been indicated that maintaining the WHO recommended safe  
35 distance between people becomes challenging with high population densities.<sup>19</sup> Furthermore,  
36 overall effects of population density and disease spread has been previously described in  
37 literature.<sup>20,21</sup> In addition to lack of social distancing, higher population density may also be  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

1  
2  
3 associated with several other behavioral and socio-demographic attributes that may predispose  
4 populations to both viral spread and increased susceptibility. For example, there are reports  
5 linking obesity, lack of physical activity, and higher mortality with residence in densely  
6 populated neighborhoods.<sup>22,23</sup>  
7  
8  
9  
10

11  
12 As reported, our data also corroborate that older populations may be more susceptible to  
13 SARS-CoV-2 infection.<sup>10</sup> However, younger populations still have cause for concern as nearly 1  
14 in 4 of the infected cases in our sample were between 36–50 years of age. Finally, our data  
15 demonstrate that males may be approximately 20% more likely to test positive for the SARS-  
16 CoV-2 infection. Potential sex differences in infectivity to SARS-CoV-2 and intersectionality  
17 with racial and ethnic socioeconomic factors need to be explored further in future analyses.  
18  
19 Additional policy-oriented research should prioritize studying the intersectionality of these  
20 vulnerable economic statuses and racial disparities in COVID infection indicated by the present  
21 study.  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32

33 Findings of our study need to be interpreted in the light of certain limitations. Our data  
34 are from a single center and may not be generalizable to the wider U.S. population. These  
35 findings need to be replicated in larger data sets across other large heterogenous U.S.  
36 metropolitans. However, the Houston metropolitan area is one of the most diverse and  
37 representative in the U.S.<sup>17</sup> and our healthcare system is one of the largest systems providing  
38 care to COVID-19 patients in the Greater Houston area. Our sample was composed of 22%  
39 Black, 18% Hispanic, and 62% female population. Our final multivariable models included  
40 potential mediators which may produce biased estimates.<sup>16</sup> However, these potential mediators  
41 did not demonstrate a statistically significant indirect effect in our analyses. We did not have  
42 information on certain demographic covariates such as education or household size. Educational  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

1  
2  
3 status has been linked to healthcare awareness and may be important to adjust for in analyses of  
4 potential disparities, and household size may be used to provide more precise estimates of socio-  
5 economic status. However, we obtained and adjusted for zip code income data from the U.S.  
6 Census, as income has previously been shown to have strong correlation with educational  
7 attainment and socio-economic status.<sup>24</sup> Since testing was based on suspicion of infection and  
8 may have been influenced by factors such as access to care, the potential for selection bias  
9 cannot be ruled out. Furthermore, lack of sensitivity of SARS-CoV-2 diagnostics tests have been  
10 reported; however, the three assays utilized for testing were cross validated for internal  
11 consistency. Finally, we did not have detailed information on comorbidities and their  
12 management in the study population. However, we did control for major comorbidities which are  
13 being reported as associated with COVID-19 outcomes.<sup>25</sup>  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30

## 31 **Conclusions**

32  
33 The strong association between racial and ethnic minorities and SARS-CoV-2 infection  
34 demonstrated in our data, even after adjustment for other important socio-demographic and  
35 comorbidity factors, highlight a potential catastrophe of inequality within the existential crisis of  
36 a global pandemic. Our data, representing a large heterogeneous U.S. metropolitan area, also  
37 provide preliminary evidence into the potential pathways for this disparity. It is highly likely that  
38 higher comorbidity burden and detrimental effects of adverse social determinants, including  
39 those that may not adequately permit safe practices of social distancing, mediate higher SARS-  
40 CoV-2 infectivity among racial and ethnic minorities.  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50

51 As the pandemic continues to spread and evolve across the continental U.S., emerging  
52 data on association between SARS-CoV-2 infection and various socio-demographic factors will  
53  
54  
55  
56  
57  
58  
59  
60

1  
2  
3 continue to enhance our understanding of targeted risks related to SARS-CoV-2 infection, and  
4 such data would enable us to comprehend healthcare services and access factors related to  
5 development and outcomes of COVID-19 among minority populations. Our findings substantiate  
6 prior calls for collection of robust data on race and ethnicity as a part of international  
7 collaborations,<sup>26</sup> and further drive home the critical importance of quantifying novel SDoH.  
8  
9  
10  
11  
12  
13  
14  
15  
16

17 **Acknowledgment:** The authors thank Jacob M. Kolman, Senior Scientific Writer of the Houston  
18 Methodist Center for Outcomes Research, for reviewing the language and format of the  
19 manuscript.  
20  
21  
22  
23

24 **IRB Approval:** This work was carried out under an approved protocol for the Houston  
25 Methodist COVID-19 Surveillance and Outcomes Registry (HM CURATOR) by the Houston  
26 Methodist Research Institute Institutional Review Board (HMRI IRB).  
27  
28  
29  
30

31 **Contribution Statement:**

32 FV: design, data analysis and interpretation, drafting the manuscript, critical revision for  
33 important intellectual content, final approval  
34  
35

36 JCN: data acquisition, data analysis, drafting the manuscript, final approval  
37

38 JRM: data acquisition, drafting the manuscript, final approval  
39

40 OK: data acquisition, data analysis, drafting the manuscript, final approval  
41

42 AP: data acquisition, data analysis, drafting the manuscript, final approval  
43

44 SLJ: data acquisition, data interpretation, critical revision for important intellectual content, final  
45 approval  
46

47 FNM: critical revision for important intellectual content, final approval  
48

49 HDS: critical revision for important intellectual content, final approval  
50

51 RAP: critical revision for important intellectual content, final approval  
52

53 JDA: critical revision for important intellectual content, final approval  
54

55 BK: critical revision for important intellectual content, final approval  
56  
57  
58  
59  
60

1  
2  
3 KN: design, interpretation of data, critical revision for important intellectual content, final  
4 approval  
5

6 **Competing Interests:** No authors declare a competing interest to this work  
7

8  
9 **Funding:** There are no funders to report for this submission.  
10

11 **Date sharing statement:** All requests for de-identified data should be made to the corresponding  
12 author. All reasonable requests will be evaluated by the CURATOR Data Governance and  
13 Sharing Committee comprising of FV, SLJ, BK and KN in the light of institutional policies and  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

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](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->

- 1  
2  
3 deaths-demographics-lightfoot-20200406-77nlylhiavgjzb2wa4ckivh7mu-story.html.  
4 Published April 7, 2020. Accessed April 20, 2020.  
5  
6  
7 13. Thebault R, Tran AB, Williams V. The coronavirus is infecting and killing black Americans  
8 at an alarmingly high rate. *Washington Post*.  
9 [https://www.washingtonpost.com/nation/2020/04/07/coronavirus-is-infecting-killing-black-](https://www.washingtonpost.com/nation/2020/04/07/coronavirus-is-infecting-killing-black-americans-an-alarmingly-high-rate-post-analysis-shows/)  
10 [americans-an-alarmingly-high-rate-post-analysis-shows/](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  
11 April 20, 2020.  
12  
13 14. United States Census Bureau. American Community Survey (ACS).  
14 <https://www.census.gov/programs-surveys/acs>. Published 2018. Accessed April 14, 2020.  
15  
16 15. United States Census Bureau. Gazetteer Files.  
17 <https://www.census.gov/geographies/reference-files/time-series/geo/gazetteer-files.html>.  
18 Accessed April 20, 2020.  
19  
20 16. Richiardi L, Bellocco R, Zugna D. Mediation analysis in epidemiology: methods,  
21 interpretation and bias. *Int J Epidemiol*. 2013;42(5):1511-1519. doi:10.1093/ije/dyt127  
22  
23 17. Harris County Public Health. *Harris Cares: A 2020 Vision of Health in Harris County*.  
24 Houston, TX: Harris County Public Health; 2019.  
25 <http://publichealth.harriscountytexas.gov/Resources/Harris-Cares>. Accessed November 21,  
26 2019.  
27  
28 18. Quinn SC, Kumar S, Freimuth VS, Musa D, Casteneda-Angarita N, Kidwell K. Racial  
29 disparities in exposure, susceptibility, and access to health care in the US H1N1 influenza  
30 pandemic. *Am J Public Health*. 2011;101(2):285-293. doi:10.2105/AJPH.2009.188029  
31  
32 19. Rocklov J, Sjodin H. High population densities catalyze the spread of COVID-19. *J Travel*  
33 *Med*. 2020. doi: 10.1093/jtm/taaa038  
34  
35 20. Tarwater PM, Martin CF. Effects of population density on the spread of disease. *Complexity*.  
36 2001;6(6):29-36. doi:10.1002/cplx.10003  
37  
38 21. Yang K, LeJeune J, Alsdorf D, Lu B, Shum CK, Liang S. Global distribution of outbreaks of  
39 water-associated infectious diseases. *PLoS Negl Trop Dis*. 2012;6(2):e1483.  
40 doi:10.1371/journal.pntd.0001483  
41  
42 22. Wang R, Feng Z, Xue D, Liu Y, Wu R. Exploring the links between population density,  
43 lifestyle, and being overweight: secondary data analyses of middle-aged and older Chinese  
44 adults. *Health Qual Life Outcomes*. 2019;17(1):100. doi:10.1186/s12955-019-1172-3  
45  
46 23. Beenackers MA, Oude Groeniger J, Kamphuis CBM, Van Lenthe FJ. Urban population  
47 density and mortality in a compact Dutch city: 23-year follow-up of the Dutch GLOBE  
48 study. *Health & Place*. 2018;53:79-85. doi:10.1016/j.healthplace.2018.06.010  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

- 1  
2  
3 24. U.S. Bureau of Labor Statistics. Unemployment rates and earnings by educational  
4 attainment. <https://www.bls.gov/emp/chart-unemployment-earnings-education.htm>.  
5 Published September 4, 2019. Accessed April 15, 2020.  
6  
7  
8 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).  
9 [https://www.cdc.gov/coronavirus/2019-ncov/need-extra-precautions/groups-at-higher-](https://www.cdc.gov/coronavirus/2019-ncov/need-extra-precautions/groups-at-higher-risk.html)  
10 [risk.html](https://www.cdc.gov/coronavirus/2019-ncov/need-extra-precautions/groups-at-higher-risk.html). Last updated April 17, 2020. Accessed April 20, 2020.  
11  
12 26. Abuelgasim E, Saw LJ, Shirke M, Zeinah M, Harky A. COVID-19: Unique public health  
13 issues facing Black, Asian and minority ethnic communities. *Curr Probl Cardiol*.  
14 2020;100621. doi:10.1016/j.cpcardiol.2020.100621  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60



## 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

For peer review only

**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) <sup>a</sup>
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 (%) <sup>b</sup>				
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 (%) <sup>c</sup>	17.8	16.8	29.2	2.02 (1.79 – 2.27)
Median Zip Household Income (IQR) <sup>d</sup>	70,658 (53,313–99,276)	70,758 (53,633–100,107)	66,523 (50,485–94,226)	-4,168 <sup>e</sup> (-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 <sup>f</sup>	2797.2 (1439.1 – 4260.9)	2797.2 (1439.1 – 4211.4)	3320.3 (1904.4 – 4439.7)	523.1 <sup>e</sup> (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) <sup>f</sup>
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) <sup>a</sup>
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)

<sup>a</sup> Unadjusted Odds Ratios and 95% Confidence Intervals for association between individual co-variates in SARS-CoV-2 positivity.

<sup>b</sup> Race: Missing, Unknown, Declined, n = 1157 (5.7%).

<sup>c</sup> Ethnicity: Missing, Unknown, Declined, n = 613 (3.0%).

<sup>d</sup> 2018 inflation adjusted USD. Missing n = 1,883 (9.3%). Pentiles were defined by categorizing the ordered distribution of median income into five categories.

<sup>e</sup> Difference in median and 95% CI of difference obtained via quantile regression.

<sup>f</sup> 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 <sup>a</sup> / 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
<b>Age Category (%)</b>				
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
<b>Females</b>	68.1	58.1	1.54 (1.42 – 1.66)	< 0.001
<b>Median (IQR) Zip Income</b>	64,022 (47,303 – 79,658)	76,163 (60,130 – 102,019)	-12,141 (-14,018, -10,263)	< 0.001
<b>Median Zip Income Pentiles (%) – Pentiles of increasing Income<sup>b</sup></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
<b>Population Density for Zip: Median (IQR)</b>	3217.6 (2040.7 – 4439.7)	2488.5 (812.2 – 4084.3)	729.1 (603.0 – 855.2)	< 0.001
<b>Population Density for Zip Pentiles (%) – Pentiles of increasing population density<sup>c</sup></b>				
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
<b>Charlson Comorbidity Index, median (IQR)</b>	2 (0 – 6)	3 (1 – 7)	-1 (-1.22, -0.78)	< 0.001
<b>Charlson Comorbidity Index (CCI) Categories (%)</b>				
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

<sup>a</sup> 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.

<sup>b,c</sup> 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 <sup>a</sup> / Median Difference <sup>a</sup> (95% CI)	P value
Age: Mean (SD)	45.1 (18.3)	52.8 (18.9)	0.98 (0.98 – 0.98)	< 0.001
<b>Age Category (%)</b>				
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
<b>Females</b>	64.5	61.4	1.14 (1.06 – 1.23)	0.001
<b>Median (IQR) Zip Income</b>	65,742 (48,345 – 82,708)	73,742 (56,288 – 102,008)	-8,000 (-9,450.5, -6,549.5)	< 0.001
<b>Median Zip Income Pentiles (%) – Pentiles of increasing Income<sup>b</sup></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
<b>Population Density for Zip: Median (IQR)</b>	3256.8 (1504.0 – 4299.7)	2741.6 (1408.1 – 4110.7)	515.2 (469.2 – 561.2)	< 0.001
<b>Population Density for Zip Pentiles (%) – Pentiles of increasing population density<sup>c</sup></b>				
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
<b>Charlson Comorbidity Index, median (IQR)</b>	1 (0 – 4)	2 (0 – 6)	-1 (-1.21, -0.79)	< 0.001
<b>Charlson Comorbidity Index (CCI) Categories (%)</b>				
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

<sup>a</sup> 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.

<sup>b,c</sup> 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) <sup>a</sup>	Hispanic vs. Non-Hispanic, aOR (95% CI) <sup>b</sup>
<b>Non-Hispanic Black (vs. Non-Hispanic White)</b>	<b>2.23 (1.90 – 2.60)<sup>c</sup></b>	
<b>Hispanic (vs. Non-Hispanic)</b>		<b>1.95 (1.72 – 2.20)<sup>c</sup></b>
<b>Age Categories</b>		
Up to 35 years	<i>Reference Category</i>	
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)
<b>Male (vs. Female)</b>	1.20 (1.04 – 1.39)	1.17 (1.05 – 1.32)
<b>Median Zip Household Income Categories (Pentiles of Increasing Income)</b>		
Category I	<i>Reference Category</i>	
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)	
<b>Primary Insurance Type</b>		
Medicare	<i>Reference Category</i>	
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)
<b>Charlson Comorbidity Index Categories</b>		
CCI: 0	<i>Reference Category</i>	
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)

<sup>a,b</sup> Hosmer and Lemeshow goodness of fit p-value: 0.58 and 0.73 ( $H_0$ : Model fit is correct).

<sup>c</sup> 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.

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

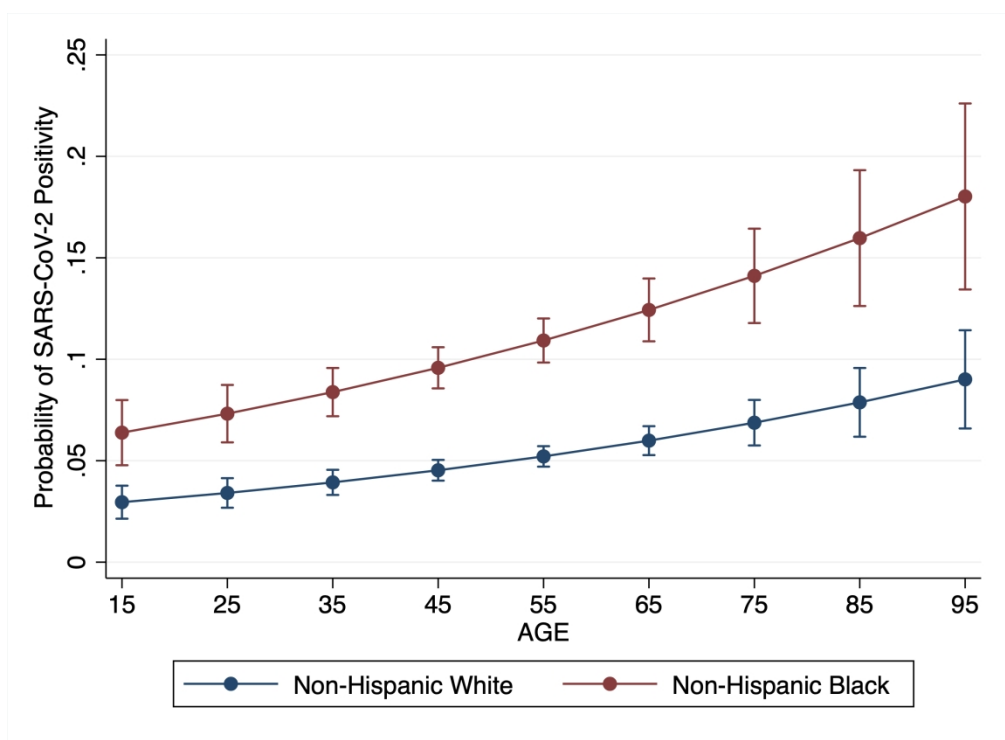


Figure 1: Adjusted Probability and 95% Confidence Interval of Positive SARS-CoV-2 PCR in Non-Hispanic Black vs. Non-Hispanic White by increasing age

467x339mm (144 x 144 DPI)

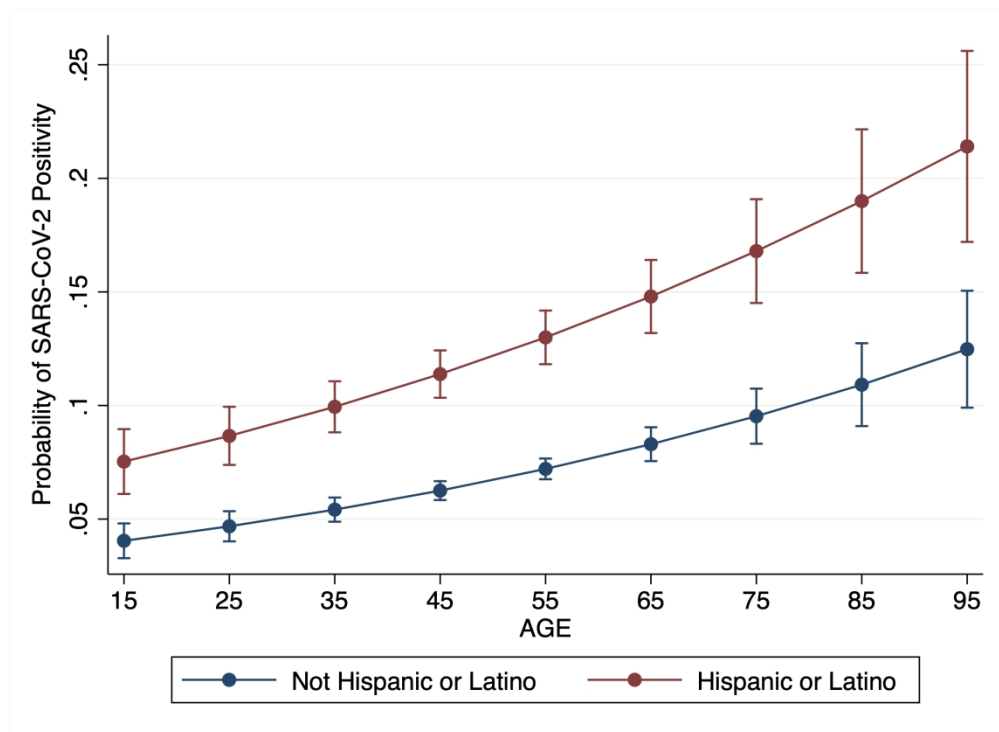


Figure 2: Adjusted Probability and 95% Confidence Interval of Positive SARS-CoV-2 PCR in Hispanic vs. Non-Hispanic by increasing age

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
<b>Title and abstract</b>	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
<b>Introduction</b>			
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
<b>Methods</b>			
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
<b>Results</b>			
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
<b>Discussion</b>			
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
<b>Other information</b>			
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](http://www.strobe-statement.org).