# BMJ Open

# Racial and ethnic disparities in SARS-CoV-2 pandemic: analysis of a COVID-19 observational registry for a diverse US metropolitan population

Farhaan S Vahidy [1,2] Juan Carlos Nicolas,[1] Jennifer R Meeks,[1] Osman Khan,[1] Alan Pan,[1] Stephen L Jones,[1,3,4] Faisal Masud,[3,5] H Dirk Sostman,[3,6,7] Robert Phillips,[1,3,8] Julia D Andrieni,[3,9] Bita A Kash,[1,3,10] Khurram Nasir[1,8]

For numbered affiliations see end of article.

**Correspondence to**
Dr Farhaan S Vahidy;
fvahidy@houstonmethodist.org

## ABSTRACT

**Introduction** Data on race and ethnic disparities for severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) infection are limited. We analysed sociodemographic 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** This is a cross-sectional analysis of the COVID-19 Surveillance and Outcomes Registry, which captures data for a large healthcare system, comprising 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 analysed sociodemographic (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 OR (aOR) and 95% CI for likelihood of a positive SARS-CoV-2 test. Structural equation modelling (SEM) framework was used to explore three mediation pathways (low income, high population density, high comorbidity burden) for the association between non-Hispanic black (NHB) race, Hispanic ethnicity and SARS-CoV-2 infection.

**Results** Among 20 228 tested individuals, 1551 (7.7%) tested positive. The overall mean (SD) age was 51.1 (19.0) years, 62% were females, 22% were black and 18% were Hispanic. NHB and Hispanic ethnicity were associated with lower socioeconomic status and higher population density residence. In the fully adjusted model, NHB (vs non-Hispanic white; aOR, 2.23, CI 1.90 to 2.60) and Hispanic ethnicity (vs non-Hispanic; aOR, 1.95, CI 1.72 to 2.20) had a higher likelihood of infection. Older individuals and males were also at higher risk of infection. The SEM framework demonstrated a significant indirect effect of NHB and Hispanic ethnicity on SARS-CoV-2 infection mediated via a pathway including residence in densely populated zip code.

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

## Strengths and limitations of this study

► This is one of the first studies to systematically evaluate race and ethnic disparities in susceptibility to severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) infection, while accounting for multiple sociodemographic characteristics and comorbidities.

► Study population represents a large and diverse metropolitan of the USA with data from one of the largest healthcare providers across the greater metropolitan area.

► The study evaluates potential mediation pathways for race disparities and demonstrates that residence in areas with high population density may mediate race and ethnic disparities in susceptibility to SARS-CoV-2 infection.

► This is a single-centre study with limited information on burden of comorbidity and lifestyle factors.

## INTRODUCTION

COVID-19, caused by infection with 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 USA has approximately 25% of the total global cases and has surpassed all countries in terms of absolute number of cases, cases per one million population and fatalities.[1 2] Experts project these numbers to continue rising as widespread testing is instituted and newer patterns of infectivity emerge. The geographical distribution of cases across the USA demonstrates that the predominant pandemic burden hits 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 US Virgin Islands.[3] As of 31 May 2020, the state of Texas has 64 287 reported cases of COVID-19,

with about one-third in the Greater Houston area.[4] The Greater Houston area is home to approximately seven million individuals, is the fourth largest metropolitan area by population in the USA and is considered one of the nation's most diverse regions.[5 6]

Initial reports indicate that specific individuals such as the elderly, males and people with comorbidities, including hypertension, diabetes, obesity, coronary artery disease and heart failure, have poor COVID-19 outcomes.[7–10] As the pandemic spread over the continental USA during the last 4 months, patterns of high-risk phenotypes started to emerge and reports of poor outcomes (particularly high case fatality) among racial minorities surfaced.[11–13] Although it is important to understand the determinants of poor outcomes among patients with COVID-19, 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 USA. Data on higher likelihood of SARS-CoV-2 infection among racial and ethnic minorities across diverse US metropolitan areas are limited. Furthermore, the mediators of SARS-CoV-2 infection among racial and ethnic minorities have not been described.

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

## METHODS

We analysed data between 5 March and 31 May 2020 collected as part of the COVID-19 Surveillance and Outcomes Registry (CURATOR) at Houston Methodist (HM). The CURATOR, designed and managed by the big data team at the Center for Outcomes Research at HM, is populated from multiple data sources across the HM system, such as electronic medical records, electronic databanks for laboratory and pharmacy, and electronic interactive patient interface tools. The HM system comprises a flagship tertiary care hospital in the Texas Medical Center, seven large community hospitals, a continuing care hospital, and multiple emergency centres and clinics throughout the Greater Houston area. Data from various sources are curated into a harmonised format, assessed for quality and integrity, and stored on a secure institutional Health Insurance Portability and Accountability Act (HIPAA)-compliant server.

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

Sociodemographic characteristics including age, sex, race, ethnicity and payer status (insurance type) were obtained from the HM CURATOR for analyses. We also extracted information on the presence of comorbidities comprising the Charlson Comorbidity Index (CCI), which include history of myocardial infarction, congestive heart failure, peripheral vascular disease, cerebrovascular disease, dementia, chronic pulmonary disease, rheumatic disease, peptic ulcer disease, liver disease, diabetes with or without complications, hemiplegia, renal disease, any malignancy (excluding skin neoplasms), metastatic solid tumours and AIDS/HIV. Data on hypertension and obesity were additionally obtained. We used the US Census Bureau's American Community Survey 5-year data (2014–2018) to determine the median household income by individual zip code tabulation area (ZCTA).[14] The median ZCTA household income was inflation adjusted to 2018 US dollars. We also used the same data source to obtain population estimates by ZCTA, and calculated ZCTA-level population density (population per mile square) by standardising it for area measurements of ZCTA. For the purpose of population density determination, land area estimates were obtained from the Census Bureau's US Gazetteer Files 2010.[15] In the absence of granular and precise social distancing data, we have used population density as a proxy for potential difficulties in social distancing among crowded communities.

We provide descriptive summary data as mean (SD) and proportion. We fit univariable and multivariable logistic regression models to assess unadjusted and adjusted associations between sociodemographic characteristics and likelihood of being tested positive for SARS-CoV-2. We additionally provide univariable comparison of various sociodemographic and comorbidity variables between non-Hispanic black (NHB) and non-Hispanic white (NHW) race categories, as well as between Hispanic and non-Hispanic ethnic groups. Age, income, population density and CCI were categorised for certain analyses. We included age, sex, race, ethnicity, zip code household income, insurance type, zip population density and CCI in our initial multivariable model. Zip code household income, zip population density and CCI were evaluated

as mediators. Factors demonstrating mediation were excluded from the final models. However, the factors that did not demonstrate mediation were included in the final models, as we believe that they continue to importantly inform the variance of estimates for direct effects.[16] We assessed the model fit using the Hosmer-Lemeshow goodness of fit test, and crude OR and adjusted OR (aOR) and 95% CI are reported. Postestimation marginal probabilities of SARS-CoV-2 infection were determined from the final adjusted model for major covariates (race, ethnicity and age). We explored the mediation influence of comorbidity burden (CCI), socioeconomic 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 generalised structural equation modelling (GSEM) framework. The GSEM framework was set up to provide estimates of direct and indirect effects 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
### Sociodemographic 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 1551 (7.7%, CI 7.3 to 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 with 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 were derived from a sample of 3 216 290 patients managed across the system since 22 May 2016.

The overall median (IQR) household income was US$70 658 (US$53 313–US$99 276), and 42.6% of the study population had private or employer-based insurance. In our univariate analysis, black race (vs white; OR, 1.55, CI 1.37 to 1.75), Hispanic ethnicity (vs non-Hispanic; OR, 2.02, CI 1.79 to 2.27) and male (vs female; OR, 1.17, CI 1.06 to 1.31) were associated with significantly higher likelihood of testing positive for SARS-CoV-2. Among the SARS-CoV-2-positive patients, 40.8% were in the 51–75 years age category and 11.4% were older than 75 years. These proportions were significantly higher than the reference group (up to 35 years; for 51–75 years vs up to 35 years: OR 1.29, CI 1.12 to 1.48; for >75 years vs up to 35 years: OR 1.23, CI 1.02 to 1.49). Furthermore,

individuals in higher pentiles of socioeconomic status had significantly lower likelihood, whereas those residing in higher population density ZCTA had higher likelihood of SARS-CoV-2 infection. We observed a significantly higher proportion of SARS-CoV-2-positive individuals in the CCI 1–2 category compared with CCI of 0 (OR, 1.35, CI 1.18 to 1.54). However, similar differences for higher CCI categories were not observed. For specific comorbidities, a significantly greater proportion of individuals with diabetes had SARS-CoV-2-positive results (OR, 1.40, CI 1.24 to 1.57). The sociodemographic characteristics and comorbidity profiles for the overall and SARS-CoV-2 positive and negative patients are summarised in table 1.

### Sociodemographic and comorbidity characteristics associated with minority race and ethnicity
In our study sample comprising 13 754 non-Hispanic black and white individuals, we compared the association between race and various sociodemographic and comorbidity characteristics (table 2). Similarly, we also evaluated univariable differences for sociodemographic variables and comorbidities between Hispanic and non-Hispanic individuals (table 3). Minority race (NHB) and ethnicity (Hispanic) were both associated with younger age, higher proportion of females, and residence in low income and higher population density ZCTA. However, NHB and Hispanic groups were both associated with an overall lower burden of comorbidities (as demonstrated by significantly lower median CCI) compared, respectively, with NHW and non-Hispanic categories. A higher proportion of individuals among minority race and ethnicity were diabetic and a higher proportion of NHB were also hypertensive compared with NHW.

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

Table 1 Summary measures and univariable association of sociodemographic 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=1551) | OR (95% CI)* |
|---|---|---|---|---|
| Age, mean (SD) | 51.1 (19.0) | 51.0 (19.1) | 52.1 (18.1) | 1.00 (1.00 to 1.01) |
| Age categories (%), years | | | | |
| Up to 35 | 25.1 | 25.5 | 20.9 | Reference category |
| 36–50 | 25 | 24.9 | 27 | 1.32 (1.14 to 1.54) |
| 51–75 | 38.7 | 38.5 | 40.8 | 1.29 (1.12 to 1.48) |
| >75 | 11.2 | 11.2 | 11.4 | 1.23 (1.02 to 1.49) |
| Female (%) | 61.9 | 62.3 | 58.3 | 0.85 (0.76 to 0.94) |
| Race (%)† | | | | |
| White | 62.3 | 62.9 | 55.7 | Reference category |
| Black | 21.6 | 21 | 28.8 | 1.55 (1.37 to 1.75) |
| Asian | 9.2 | 9.1 | 9.7 | 1.20 (1.00 to 1.44) |
| Mixed/other | 1.3 | 1.2 | 1.5 | 1.36 (0.88 to 2.11) |
| Hispanic (%)‡ | 17.8 | 16.8 | 29.2 | 2.02 (1.79 to 2.27) |
| Median zip household income (IQR)§ | 70 658 (53 313–99 276) | 70 758 (53 633–100 107) | 66 523 (50 485–94 226) | –4168¶ (–6463.2 to –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 to 1.04) |
| III: 65 897–79 869 | 18.3 | 18.2 | 19.6 | 0.82 (0.70 to 0.96) |
| IV: 80 039–106 067 | 18.6 | 19 | 13.7 | 0.55 (0.46 to 0.65) |
| V: 106 415–240 417 | 17.2 | 17.3 | 16.1 | 0.71 (0.60 to 0.83) |
| Median (IQR) population density** | 2797.2 (1439.1–4260.9) | 2797.2 (1439.1–4211.4) | 3320.3 (1904.4–4439.7) | 523.1¶ (454.9 to 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 | 19.1 | 18.1 | 1.31 (1.09 to 1.58) |
| III: 2330.8–3328.9 | 17.4 | 17.4 | 18.3 | 1.45 (1.21 to 1.75) |
| IV: 3360.1–4665.6 | 18.1 | 17.8 | 22.1 | 1.71 (1.43 to 2.05) |
| V: 4742.6–98 025.9 | 18.1 | 17.7 | 22.6 | 1.76 (1.47 to 2.10) |
| Insurance status (%) | | | | |
| Medicare | 29 | 29.1 | 27.2 | Reference category |
| Medicaid | 4.7 | 4.8 | 4.3 | 0.96 (0.73 to 1.26) |
| Private/employer-based | 42.6 | 43.2 | 36.4 | 0.90 (0.79 to 1.03) |
| Healthcare exchange | 1.7 | 1.6 | 3 | 2.02 (1.46 to 2.80) |
| Self-pay | 20.6 | 20 | 28.4 | 1.52 (1.33 to 1.75) |
| Veterans affairs | 1.3 | 1.4 | 0.7 | 0.54 (0.29 to 1.00) |
| CCI, median (IQR) | 2 (0–6) | 2 (0–6) | 2 (0–5) | 0 (–0.36 to 0.36)** |
| CCI categories (%) | | | | |
| 0 | 33.1 | 33.4 | 30.4 | Reference category |
| 1–2 | 23.7 | 23.1 | 28.6 | 1.35 (1.18 to 1.54) |
| 3–6 | 20.3 | 20.2 | 20.4 | 0.10 (0.65 to 1.28) |
| >6 | 22.9 | 23.1 | 20.6 | 0.98 (0.84 to 1.13) |
| Hypertension | 47.2 | 47.1 | 48.4 | 1.06 (0.95 to 1.17) |

**Table 1** Continued

| Characteristics | Overall (N=20 228) | SARS-CoV-2-negative (n=18 677) | SARS-CoV-2-positive (n=1551) | OR (95% CI)* |
|---|---|---|---|---|
| Diabetes (without complications) | 24.2 | 23.7 | 30.3 | 1.40 (1.24 to 1.57) |
| Obesity | 28 | 28.2 | 25.2 | 0.86 (0.76 to 0.96) |

*Unadjusted OR and 95% CI for the association between individual covariates in SARS-CoV-2 positivity.
†Race: missing, unknown, declined, n=1157 (5.7%).
‡Ethnicity: missing, unknown, declined, n=613 (3.0%).
§2018 inflation-adjusted US dollars. Missing, n=1883 (9.3%). Pentiles were defined by categorising the ordered distribution of median income into five categories.
¶Difference in median and 95% CI of the difference obtained via quantile regression.
**Population density: missing, n=1854 (9.2%). Pentiles were defined by categorising the ordered distribution of population density into five categories.
CCI, Charlson Comorbidity Index; CURATOR, COVID-19 Surveillance and Outcomes Registry; HM, Houston Methodist; SARS-CoV-2, severe acute respiratory syndrome coronavirus 2.

non-Hispanic across age spectrum are presented in figures 1 and 2.

### GSEM for mediation by income, population density and comorbidity index

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

### DISCUSSION

The underlying race and ethnic healthcare disparities have been painfully highlighted in the wake of the COVID-19 pandemic. Most reports indicate higher mortality or case fatality among minority racial groups (black/African–American) across major US metropolitan areas.[11–13] However, robust insights on the racial differences for SARS-CoV-2 infection are limited. Furthermore, comprehensive data evaluating higher susceptibility to SARS-CoV-2 infection among Hispanic communities are also scarce. This is perhaps because of comparatively homogeneous populations in non-US regions of the world. Houston, as an exceptionally ethnically diverse population centre,[17] is well suited for an investigation of racial, ethnic and socioeconomic gradients in COVID-19

test positivity. We focus on highlighting the mechanisms of racial and ethnic disparities in susceptibility to SARS-CoV-2 infection and provide evidence of mediation of such disparities by novel social determinants of health (SDoH).

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

We explored three possible mechanisms of race disparities in our data. These included lower socioeconomic status, residence in higher population dense areas and higher level of comorbidities. We demonstrate that NHB race is significantly associated with all three potential disparity pathways, and in the traditional multivariable analyses racial and ethnic disparities persisted even after controlling for these pathways. However, our mediation analyses highlighted the potential influence of residence in high population density areas as a viable pathway that at least partially explains the observed racial and ethnic disparity. Furthermore, residence in low-income areas emerged as a significant mediation pathway for ethnic differences in SARS-CoV-2 positivity. Pathways mediating the influence of comorbidity status did not demonstrate a significant effect. We used population density as a marker for potential inability to maintain adequate social distancing as it has been indicated that maintaining the WHO-recommended safe distance between people

**Table 2**  Univariable comparison of sociodemographic and comorbidity factors between NHB and NHW race categories

| | NHB (n=4285) | NHW (n=9469) | OR*/median difference (95% CI) | P value |
|---|---|---|---|---|
| Age, mean (SD) | 49.4 (17.8) | 56.0 (19.2) | 0.98 (0.98 to 0.98) | <0.001 |
| Age categories (%) | | | | |
| Up to 35 | 24.3 | 18.8 | Reference | |
| 36–50 | 28.8 | 19.8 | 1.12 (1.01 to 1.25) | 0.03 |
| 51–75 | 39.1 | 44.9 | 0.68 (0.61 to 0.74) | <0.001 |
| >75 | 7.8 | 16.5 | 0.37 (0.32 to 0.42) | <0.001 |
| Female | 68.1 | 58.1 | 1.54 (1.42 to 1.66) | <0.001 |
| Median (IQR) zip income | 64 022 (47 303–79 658) | 76 163 (60 130–102 019) | −12 141 (−14 018 to −10 263) | <0.001 |
| Median zip income pentiles (%), pentiles of increasing income† | | | | |
| Category I | 33.5 | 13.1 | Reference | |
| Category II | 21.3 | 19.7 | 0.42 (0.38 to 0.47) | <0.001 |
| Category III | 22.7 | 19.1 | 0.47 (0.42 to 0.52) | <0.001 |
| Category IV | 10.5 | 26.2 | 0.16 (0.14 to 0.18) | <0.001 |
| Category V | 11.9 | 22.0 | 0.21 (0.19 to 0.24) | <0.001 |
| Population density for zip, median (IQR) | 3217.6 (2040.7–4439.7) | 2488.5 (812.2–4084.3) | 729.1 (603.0 to 855.2) | <0.001 |
| Population density for zip pentiles (%), pentiles of increasing population density† | | | | |
| Category I | 11.5 | 28.3 | Reference | |
| Category II | 21.7 | 20.9 | 2.52 (2.21 to 2.86) | <0.001 |
| Category III | 21.8 | 16.9 | 3.20 (2.81 to 3.65) | <0.001 |
| Category IV | 23.0 | 15.5 | 3.66 (3.21 to 4.17) | <0.001 |
| Category V | 22.2 | 18.3 | 2.99 (2.63 to 3.41) | <0.001 |
| CCI, median (IQR) | 2 (0–6) | 3 (1–7) | −1 (−1.22 to −0.78) | <0.001 |
| CCI categories (%) | | | | |
| 0 | 30.8 | 24.9 | Reference | |
| 1–2 | 25.1 | 22.3 | 0.91 (0.82 to 1.01) | 0.07 |
| 3–6 | 19.7 | 24.0 | 0.66 (0.60 to 0.74) | <0.001 |
| >6 | 24.4 | 28.8 | 0.68 (0.62 to 0.75) | <0.001 |
| Hypertension | 56.0 | 52.6 | 1.14 (1.07 to 1.23) | <0.001 |
| Diabetes (without complications) | 29.6 | 21.9 | 1.50 (1.38 to 1.63) | <0.001 |
| Obesity | 33.1 | 30.8 | 1.11 (1.03 to 1.20) | 0.008 |

*Unadjusted OR and 95% CI for the association between individual covariates in SARS-CoV-2 positivity or difference in median and 95% CI of the difference obtained via quantile regression.
†Pentiles were defined by categorising the ordered distribution of median income and population density into five categories.
NHB, non-Hispanic black; NHW, non-Hispanic white; SARS-CoV-2, severe acute respiratory syndrome coronavirus 2.

becomes challenging with high population densities.[19] Furthermore, overall effects of population density and disease spread have been previously described in the literature.[20 21] In addition to lack of social distancing, higher population density may also be associated with several other behavioural and sociodemographic attributes that may predispose populations to both viral spread and increased susceptibility. For example, there are reports linking obesity, lack of physical activity and higher mortality with residence in densely populated neighbourhoods.[22 23]

As reported, our data also corroborate that older populations may be more susceptible to SARS-CoV-2 infection.[10] However, younger populations still are cause for concern as nearly one in four of the infected cases in our sample were between 36 and 50 years of age. Finally, our data demonstrate that males may be approximately 20% more likely to test positive for the SARS-CoV-2 infection.

**Table 3** Univariable comparison of sociodemographic and comorbidity factors between Hispanic and non-Hispanic ethnicities

| | Hispanic (n=3590) | Non-Hispanic (n=16 025) | OR*/median difference* (95% CI) | P value |
|---|---|---|---|---|
| Age, mean (SD) | 45.1 (18.3) | 52.8 (18.9) | 0.98 (0.98 to 0.98) | <0.001 |
| Age categories (%) | | | | |
| Up to 35 | 36.0 | 21.8 | Reference | |
| 36–50 | 26.1 | 24.6 | 0.64 (0.58 to 0.71) | <0.001 |
| 51–75 | 32.0 | 40.1 | 0.47 (0.43 to 0.52) | <0.001 |
| >75 | 5.9 | 12.7 | 0.28 (0.24 to 0.36) | <0.001 |
| Female | 64.5 | 61.4 | 1.14 (1.06 to 1.23) | 0.001 |
| Median (IQR) zip income | 65 742 (48 345–82 708) | 73 742 (56 288–102 008) | −8000 (−9450.5 to −6549.5) | <0.001 |
| Median zip income pentiles (%), pentiles of increasing income† | | | | |
| Category I | 29.2 | 18.2 | Reference | |
| Category II | 23.3 | 19.6 | 0.74 (0.66 to 0.83) | <0.001 |
| Category III | 21.8 | 19.8 | 0.68 (0.61 to 0.76) | <0.001 |
| Category IV | 15.0 | 21.6 | 0.43 (0.38 to 0.49) | <0.001 |
| Category V | 10.6 | 20.8 | 0.32 (0.28 to 0.36) | <0.001 |
| Population density for zip, median (IQR) | 3256.8 (1504.0–4299.7) | 2741.6 (1408.1–4110.7) | 515.2 (469.2 to 561.2) | <0.001 |
| Population density for zip pentiles (%), pentiles of increasing population density† | | | | |
| Category I | 15.5 | 21.2 | Reference | |
| Category II | 21.1 | 20.9 | 1.39 (1.22 to 1.57) | <0.001 |
| Category III | 15.3 | 20.0 | 1.05 (0.92 to 1.20) | 0.48 |
| Category IV | 27.0 | 18.3 | 2.02 (1.79 to 2.28) | <0.001 |
| Category V | 21.1 | 19.5 | 1.48 (1.31 to 1.68) | <0.001 |
| CCI, median (IQR) | 1 (0–4) | 2 (0–6) | −1 (−1.21 to −0.79) | <0.001 |
| CCI categories (%) | | | | |
| 0 | 41.8 | 29.7 | Reference | |
| 1–2 | 25.4 | 23.6 | 0.76 (0.70 to 0.84) | <0.001 |
| 3–6 | 15.9 | 21.7 | 0.52 (0.47 to 0.58) | <0.001 |
| >6 | 16.9 | 24.9 | 0.48 (0.43 to 0.54) | <0.001 |
| Hypertension | 38.0 | 50.5 | 0.60 (0.56 to 0.65) | <0.001 |
| Diabetes (without complications) | 27.4 | 23.6 | 1.22 (1.12 to 1.33) | <0.001 |
| Obesity | 27.3 | 28.9 | 0.92 (0.85 to 1.00) | 0.06 |

*Unadjusted OR and 95% CI for the association between individual covariates in SARS-CoV-2 positivity or difference in median and 95% CI of the difference obtained via quantile regression.
†Pentiles were defined by categorising the ordered distribution of median income and population density into five categories.
SARS-CoV-2, severe acute respiratory syndrome coronavirus 2.

Potential sex differences in infectivity to SARS-CoV-2 and intersectionality with racial and ethnic socioeconomic factors need to be explored further in future analyses. Additional policy-oriented research should prioritise studying the intersectionality of these vulnerable economic statuses and racial disparities in COVID-19 infection indicated by the present study.

Findings of our study need to be interpreted in the light of certain limitations. Our data are from a single centre and may not be generalisable to the wider US population. These findings need to be replicated in larger data sets across other large heterogeneous US metropolitans. However, the Houston metropolitan area is one of the most diverse and representative in the USA[17] and our healthcare system is one of the largest systems providing care to patients with COVID-19 in the Greater Houston area. Our sample was composed of 22% black, 18% Hispanic and 62% female population. Our final multivariable models included

**Table 4** Adjusted OR and 95% CI for likelihood of SARS-CoV-2 positivity among minority race and ethnic groups

| Covariate | NHB vs NHW, aOR (95% CI)* | Hispanic vs non-Hispanic, aOR (95% CI)* |
|---|---|---|
| NHB (vs NHW) | **2.23 (1.90 to 2.60)†** | |
| Hispanic (vs non-Hispanic) | | **1.95 (1.72 to 2.20)†** |
| Age categories (years) | | |
| Up to 35 | Reference category | |
| 36–50 | 1.36 (1.07 to 1.72) | 1.42 (1.20 to 1.68) |
| 51–75 | 1.60 (1.21 to 2.11) | 1.71 (1.39 to 2.11) |
| >75 | 2.20 (1.52 to 3.19) | 2.08 (1.56 to 2.77) |
| Male (vs female) | 1.20 (1.04 to 1.39) | 1.17 (1.05 to 1.32) |
| Median zip household income categories (pentiles of increasing income) | | |
| Category I | Reference category | |
| Category II | 0.95 (0.76 to 1.18) | |
| Category III | 1.02 (0.84 to 1.26) | |
| Category IV | 0.74 (0.58 to 0.94) | |
| Category V | 0.97 (0.77 to 1.22) | |
| Primary insurance type | | |
| Medicare | Reference category | |
| Medicaid | 0.86 (0.55 to 1.33) | 1.01 (0.74 to 1.37) |
| Private/employer-based | 1.11 (0.88 to 1.40) | 0.93 (0.80 to 1.12) |
| Healthcare exchange | 2.06 (1.25 to 3.40) | 1.80 (1.27 to 2.54) |
| Self-pay | 1.75 (1.33 to 2.30) | 1.80 (1.27 to 2.54) |
| Veterans affairs | 0.72 (0.36 to 1.43) | 0.58 (0.31 to 1.07) |
| CCI categories | | |
| 0 | Reference category | |
| 1–2 | 1.26 (0.99 to 1.60) | 1.03 (0.87 to 1.23) |
| 3–6 | 0.94 (0.68 to 1.28) | 0.71 (0.56 to 0.91) |
| >6 | 0.91 (0.65 to 1.30) | 0.58 (0.44 to 0.76) |
| Hypertension | 0.87 (0.73 to 1.05) | 1.00 (0.87 to 1.15) |
| Diabetes (without complications) | 1.42 (1.17 to 1.71) | 1.62 (1.40 to 1.87) |
| Obesity | 0.82 (0.70 to 0.97) | 0.83 (0.73 to 0.94) |

Greyed 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.
Bold text in table represent the fully adjusted estimates for association between race/ethnicity and SARS-CoV-2 positivity
*Hosmer-Lemeshow goodness of fit p value: 0.58 and 0.73 (H₀: model fit is correct).
†OR (95% CI) represents the direct association adjusted for all other covariates in the model.
aOR, adjusted OR; NHB, non-Hispanic black; NHW, non-Hispanic white; SARS-CoV-2, severe acute respiratory syndrome coronavirus 2.
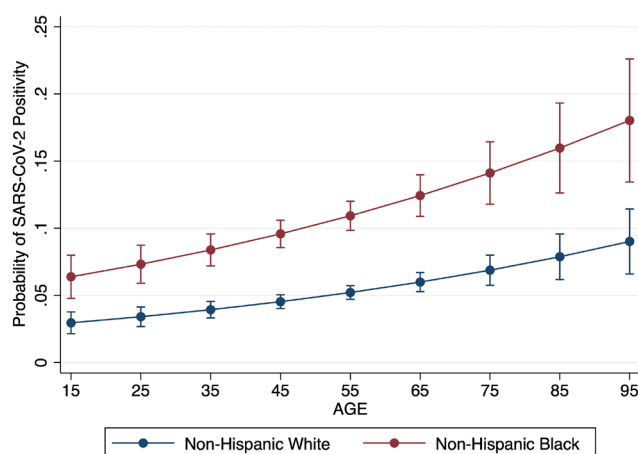


**Figure 1** Adjusted probability and 95% CI of positive SARS-CoV-2 PCR in non-Hispanic black versus non-Hispanic white by increasing age. SARS-CoV-2, severe acute respiratory syndrome coronavirus 2.

potential mediators which may produce biased estimates.[16] However, these potential mediators did not demonstrate a statistically significant indirect effect in our analyses. We did not have information on certain demographic covariates such as education or household size. Educational status has been linked to healthcare awareness and may be important to adjust for in analyses of potential disparities, and household size may be used to provide more precise estimates of socioeconomic status. However, we obtained and adjusted for zip code income data from the US Census, as income has previously been shown to have strong correlation with educational attainment and socioeconomic status.[24] Since testing was based on suspicion of infection and may have been influenced by factors such as access to care, the potential for selection bias cannot be ruled out. Furthermore, lack of sensitivity of SARS-CoV-2 diagnostic tests have been reported; however, the three assays used for testing
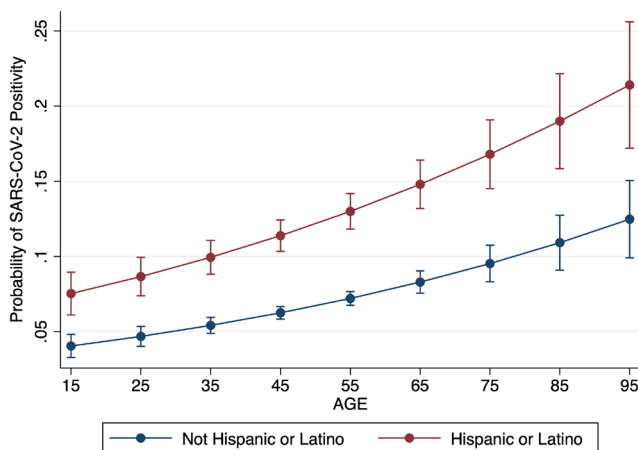


**Figure 2** Adjusted probability and 95% CI of positive SARS-CoV-2 PCR in Hispanic versus non-Hispanic by increasing age. SARS-CoV-2, severe acute respiratory syndrome coronavirus 2.

were cross-validated for internal consistency. Finally, we did not have detailed information on comorbidities and their management in the study population. However, we did control for major comorbidities which are being reported as associated with COVID-19 outcomes.[25]

## CONCLUSIONS

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

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

**Author affiliations**
[1]Center for Outcomes Research, Houston Methodist Research Institute, Houston, Texas, USA
[2]Houston Methodist Neurological Institute, Houston, Texas, USA
[3]Weill Cornell Medicine, New York, New York, USA
[4]Department of Surgery, Houston Methodist Hospital, Houston, TX, United States
[5]Department of Anesthesiology and Critical Care, Houston Methodist Hospital, Houston, TX, United States
[6]Houston Methodist Research Institute, Houston, Texas, USA
[7]Houston Methodist Academic Institute, Houston, TX, United States
[8]Department of Cardiology, DeBakey Heart and Vascular Center, Houston Methodist Hospital, Houston, TX, United States
[9]Department of Medicine, Houston Methodist Hospital, Houston, TX, United States
[10]Texas A&M University School of Rural Public Health, College Station, Texas, USA

**ORCID iD**
Farhaan S Vahidy http://orcid.org/0000-0002-3464-2111

## REFERENCES

1. World Health Organization. Who COVID-19 dashboard., 2020. Available: https://covid19.who.int/. [Accessed 6 Apr 2020].
2. Center for Systems Science and Engineering (CSSE) at Johns Hopkins University. COVID-19 map. Johns Hopkins coronavirus resource center, 2020. Available: https://coronavirus.jhu.edu/map.html. [Accessed 6 Apr 2020].
3. Centers for Disease Control and Prevention. Coronavirus disease 2019 (COVID-19) in the U.S. cdc.gov, 2020. Available: https://www.cdc.gov/coronavirus/2019-ncov/cases-updates/cases-in-us.html. [Accessed 14 Apr 2020].
4. Texas Health and Human Services. COVID-19 in Texas (dashboard), 2020. Available: https://txdshs.maps.arcgis.com/apps/opsdashboard/index.html#/ed483ecd702b4298ab01e8b9cafc8b83 [Accessed 18 Apr 2020].
5. McCann A. Most diverse cities in the U.S. Wallethub, 2019. Available: https://wallethub.com/edu/most-diverse-cities/12690/ [Accessed 29 Jun 2020].
6. Mejia B. How Houston has become the most diverse place in America. *Los Angeles Times*, 2017. Available: https://www.latimes.com/nation/la-na-houston-diversity-2017-htmlstory.html [Accessed 29 Jun 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* 2020;5:802.
8. Bonow RO, Fonarow GC, O'Gara PT, et al. Association of coronavirus disease 2019 (COVID-19) with myocardial injury and mortality. *JAMA Cardiol* 2020:751.
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* 2020:1574.
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;395:1054–62.
11. Deslatte M. Louisiana data: virus hits blacks, people with hypertension. *AP News*, 2020. Available: https://apnews.com/cde897293f88f788d3f3bfdb3d46c466. [Accessed 20 Apr 2020].
12. Reyes C, Husain N, Gutowski C, et al. Chicago's coronavirus disparity: Black Chicagoans are dying at nearly six times the rate of white residents, data show. *Chicago Tribune*, 2020. Available: https://www.chicagotribune.com/coronavirus/ct-coronavirus-chicago-coronavirus-deaths-demographics-lightfoot-20200406-77nlylhiavgjzb2wa4ckivh7mu-story.html.[Accessed 20 Apr 2020].
13. Thebault R, Tran AB, Williams V. The coronavirus is infecting and killing black Americans at an alarmingly high rate. *Washington Post*, 2020. Available: https://www.washingtonpost.com/nation/2020/04/07/coronavirus-is-infecting-killing-black-americans-an-alarmingly-high-rate-post-analysis-shows/. [Accessed 20 Apr 2020].

14 United States Census Bureau. American community survey (ACS), 2018. Available: https://www.census.gov/programs-surveys/acs [Accessed 14 Apr 2020].

15 United States Census Bureau. Gazetteer files. Available: https://www.census.gov/geographies/reference-files/time-series/geo/gazetteer-files.html [Accessed 20 Apr 2020].

16 Richiardi L, Bellocco R, Zugna D. Mediation analysis in epidemiology: methods, interpretation and bias. *Int J Epidemiol* 2013;42:1511–9.

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

18 Quinn SC, Kumar S, Freimuth VS, *et al*. Racial disparities in exposure, susceptibility, and access to health care in the US H1N1 influenza pandemic. *Am J Public Health* 2011;101:285–93.

19 Rocklöv J, Sjödin H. High population densities catalyse the spread of COVID-19. *J Travel Med* 2020;27:jtm/taaa038.

20 Tarwater PM, Martin CF. Effects of population density on the spread of disease. *Complexity* 2001;6:29–36.

21 Yang K, LeJeune J, Alsdorf D, *et al*. Global distribution of outbreaks of water-associated infectious diseases. *PLoS Negl Trop Dis* 2012;6:e1483.

22 Wang R, Feng Z, Xue D, *et al*. 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;17:100.

23 Beenackers MA, Oude Groeniger J, Kamphuis CBM, *et al*. 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.

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

25 Centers for Disease Control and Prevention. Groups at higher risk for severe illness. cdc.gov. Available: https://www.cdc.gov/coronavirus/2019-ncov/need-extra-precautions/groups-at-higher-risk.html [Accessed 17 Apr 2020].

26 Abuelgasim E, Saw LJ, Shirke M, *et al*. COVID-19: unique public health issues facing black, Asian and minority ethnic communities. *Curr Probl Cardiol* 2020;45:100621.