

## PEER REVIEW HISTORY

BMJ Open publishes all reviews undertaken for accepted manuscripts. Reviewers are asked to complete a checklist review form (<http://bmjopen.bmj.com/site/about/resources/checklist.pdf>) and are provided with free text boxes to elaborate on their assessment. These free text comments are reproduced below.

### ARTICLE DETAILS

<b>TITLE (PROVISIONAL)</b>	Racial and Ethnic Disparities in SARS-CoV-2 Pandemic: Analysis of a COVID-19 Observational Registry for a Diverse U.S. Metropolitan Population
<b>AUTHORS</b>	Vahidy, Farhaan S; Nicolas, Juan Carlos; Meeks, Jennifer R; Khan, Osman; Pan, Alan; Jones, Stephen L.; Masud, Faisal; Sostman, H Dirk; Phillips, Robert; Andrieni, Julia D; Kash, Bitu A; Nasir, Khurram

### VERSION 1 - REVIEW

<b>REVIEWER</b>	Michael Horberg Kaiser Permanente Mid-Atlantic Permanente Research Institute, Rockville, Maryland, United States
<b>REVIEW RETURNED</b>	17-May-2020

<b>GENERAL COMMENTS</b>	<p>Overall: This manuscript attempts to analyze the racial/ethnic differences for COVID-19 testing and results in a large health system in Houston. They generally succeed in this analysis, but some specific issues prevent recommendation of acceptance "as is".</p> <ul style="list-style-type: none"><li>• The numbers are dependent on what the characteristics of the underlying population is. Has testing been biased toward or against minorities? A comparison to their underlying service population is needed.</li><li>• It is unclear why they created their own comorbidity burden score, rather than using a standard one? This needs to be better justified (page 9).</li><li>• There are a lot of grammar issues, as well as inconsistent style for presenting results, including changes of tense. These should be looked at prior to acceptance.</li><li>• The GSEM framework was done for African-Americans but not Hispanics. That should have been done for both, or explicitly stated why not.</li></ul> <p>Other specific critiques:</p> <p>Abstract: Results: should note the percentage 4,513 individuals is out of (i.e., the underlying population). Introduction: Page 6, line 31-32; "one of the nation's most diverse regions." Needs reference and proof. Methods: Page 8, line 7-8: Given all of the testing platforms were approved by EUA and many are now in disrepute, the brand of PCR testing should be noted. Results:</p>
-------------------------	-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------

	<p>Page 10, line 53-54: “Although African Americans had higher proportion of younger individuals and greater proportion of females. A significantly higher proportion...” I think these sentences are meant to be merged. Otherwise, the sentences make no sense.</p> <p>Discussion:</p> <p>Page 14, line23-36: Need to compare to the underlying population, in order to put the limitation into a more complete context.</p> <p>Table 1: Define all abbreviations in legend, or better not use any in tables. What is the OR comparing? Highly Unclear. Need to define the Zip Household Income and Pentiles. Define the median population density and pentiles.</p> <p>Table 2: Same comments as Table 1.</p> <p>Table 3: Same.</p> <p>Figure 1: Y-axis: Are these days since what date? Days since onset of a patient symptoms? Very unclear. And the table doesn't print well. Very hard to read (tiny font).</p>
--	----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------

<b>REVIEWER</b>	Amer Harky Liverpool
<b>REVIEW RETURNED</b>	20-May-2020

<b>GENERAL COMMENTS</b>	<p>I think your paper will add a great value to the scientific community. There is also a recent paper on the same topic which I suggest the authors to read and cite as well:</p> <p>Abuelgasim E, Saw LJ, Shirke M, Zeinah M, Harky A. COVID-19: Unique public health issues facing Black, Asian and minority ethnic communities [published online ahead of print, 2020 May 8]. <i>Curr Probl Cardiol.</i> 2020;100621. doi:10.1016/j.cpcardiol.2020.100621</p>
-------------------------	-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------

<b>REVIEWER</b>	Rebecca Stebbins UNC - Chapel Hill
<b>REVIEW RETURNED</b>	05-Jun-2020

<b>GENERAL COMMENTS</b>	<p>The article Racial and Ethnic Disparities in SARS-CoV-2 Pandemic: Analysis of a COVID-19 Observational Registry for a Diverse U.S. Metropolitan Population aims to understand the racial and ethnic disparities in SARS-CoV-2 infection within the United States. This is a critically important and timely research question. However, there are a few major issues with the analysis that should be addressed prior to publication. Please see below for specific comments and recommendations.</p> <p>Major Issues:</p> <ul style="list-style-type: none"> <li>o The authors should remove the ORs from Table 1 as they are meaningless.</li> <li>o The adjusted odds ratio results are very biased by the inclusion of population density and socioeconomic status – both of which are mediators of the race/ethnicity – SARS-CoV-2 relationship – in the adjustment set for the model. You cannot adjust for mediators of</li> </ul>
-------------------------	----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------

	<p>your relationship of interest, because you are blocking several of the causal pathways through which the exposure may be associated with the outcome. The authors should redo the analysis with appropriate consideration for the adjustment variable set. These mediators should only be included as part of the GSEM analysis.</p> <ul style="list-style-type: none"> <li>o The presentation of the adjusted odds ratios (and 95% CIs) for each independent variable included in the models is inappropriate unless interpreted correctly as the “direct association adjusted for all other covariates in the model”. Reference this paper: Westreich D, Greenland S. The table 2 fallacy: presenting and interpreting confounder and modifier coefficients. <i>Am J Epidemiol.</i> 2013;177(4):292-298. doi:10.1093/aje/kws412</li> <li>o The first sentence of the abstract (line 6) should be rephrased. I am concerned that the way it is phrased could imply a genetic difference/susceptibility by race/ethnicity to the virus, and that is not what is being asked in this paper.</li> <li>o The comparison of the disparities by race/ethnicity and age or sex should not be made. I’m referring especially to the first paragraph on page 14. While the differences in susceptibility by age and sex are likely due to legitimate differences in biological susceptibility to the virus, this is not the case for racial and ethnic minorities. The authors should take care to appropriately discuss the social, structural, and environmental basis for the difference in infection prevalence by racial/ethnic groups.</li> <li>o Finally, a more thorough discussion of the limitations including inflation of odds ratios, bias due to confounding and lack of sensitivity of the SARS-CoV-2 diagnostic test, etc should be added.</li> </ul> <p>Minor Issues:</p> <ul style="list-style-type: none"> <li>• It would be more correct to use the term black to describe that racial category, rather than African American, as ___ American is generally used to describe immigrants or first-generation Americans.</li> <li>• Income-to-poverty ratio would be a much better indicator of socioeconomic status than simply household income, as it would account for the size of the household.</li> </ul>
--	-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------

**VERSION 1 – AUTHOR RESPONSE**

Reviewer: 1

Reviewer Name

Michael Horberg

Institution and Country

Kaiser Permanente Mid-Atlantic Permanente Research Institute, Rockville, Maryland, United States

Please state any competing interests or state ‘None declared’:

None declared.

Please leave your comments for the authors below

Overall: This manuscript attempts to analyze the racial/ethnic differences for COVID-19 testing and results in a large health system in Houston. They generally succeed in this analysis, but some specific issues prevent recommendation of acceptance “as is”.

- The numbers are dependent on what the characteristics of the underlying population is. Has testing been biased toward or against minorities? A comparison to their underlying service population is needed.

Response: We thank the reviewer for these comments. These numbers are now provided in the first paragraph of the results section. These statistics were derived from a > 3 million patient sample treated at our hospital system approximately during the last 5-year period.

- It is unclear why they created their own comorbidity burden score, rather than using a standard one? This needs to be better justified (page 9).

Response: Thank you for this suggestion. We have updated all our analyses by inclusion of the standard Charlson Comorbidity Index.

- There are a lot of grammar issues, as well as inconsistent style for presenting results, including changes of tense. These should be looked at prior to acceptance.

Response: Thank you for pointing this out. The revised version has been proofread and corrected for grammatical inconsistencies. These changes are highlighted throughout the manuscript text in the redlined version.

- The GSEM framework was done for African-Americans but not Hispanics. That should have been done for both, or explicitly stated why not.

Response: Thank you for this suggestion. As per the reviewer's recommendation, additional analyses for mediation pathways among ethnic minorities (Hispanic race) were undertaken and results have been included in the revised version of the manuscript.

Other specific critiques:

Abstract:

Results: should note the percentage 4,513 individuals is out of (i.e., the underlying population).

Response: Thank you for this comment. Our data has now been updated to a more recent time period, and the total number of unique individuals tested for SARS-CoV-2 across our system between March 5 and May 31, 2020 are being reported. The total number (20,228) is not a subset. It represents the total number of tested individuals. We have included the date range in our methods section to clarify this point.

Introduction:

Page 6, line 31-32; "one of the nation's most diverse regions." Needs reference and proof.

Response: We have now provided the requested references.

Methods:

Page 8, line 7-8: Given all of the testing platforms were approved by EUA and many are now in dispute, the brand of PCR testing should be noted.

Response: Thank you for seeking this clarification. The nomenclature (brand) of the tests used are now provided in the revised version of the manuscript.

Results:

Page 10, line 53-54: "Although African Americans had higher proportion of younger individuals and greater proportion of females. A significantly higher proportion..." I think these sentences are meant to be merged. Otherwise, the sentences make no sense.

Response: Thank you for highlighting this. The language has been revised.

Discussion:

Page 14, line 23-36: Need to compare to the underlying population, in order to put the limitation into a more complete context.

Response: Thank you for requesting this. We have now provided additional data (results section) which highlight the basic demographic break down of our patient population across an approximate five-year period. We state that this demographic composition is largely similar to the characteristics of the patients tested for SARS-CoV-2. However, we have also acknowledged in our limitations that since individuals were not systematically tested, a potential for selection bias cannot be ruled out.

Table 1:

Define all abbreviations in legend, or better not use any in tables.

What is the OR comparing? Highly Unclear.

Need to define the Zip Household Income and Pentiles.

Define the median population density and pentiles.

Response: All abbreviations have been defined in table footnotes, interpretation of odds ratios has been provided and pentiles for zip household income and population density have been defined.

Table 2:

Same comments as Table 1.

Response: All abbreviations have been defined in table footnotes, interpretation of odds ratios has been provided and pentiles for zip household income and population density have been defined.

Table 3:

Same.

Response: All abbreviations have been defined in table footnotes, interpretation of odds ratios has been provided and pentiles for zip household income and population density have been defined.

Figure 1:

Y-axis: Are these days since what date? Days since onset of a patient symptoms? Very unclear. And the table doesn't print well. Very hard to read (tiny font).

Response: Thank you for highlighting these issues. We have decided to exclude / remove figure 1. The dates represent when PCR test for SARS-CoV-2 were performed. These are now mentioned in the methods section of our revised manuscript.

Reviewer: 2

Reviewer Name

Amer Harky

Institution and Country

Liverpool

Please state any competing interests or state 'None declared':

None

Please leave your comments for the authors below

I think your paper will add a great value to the scientific community. There is also a recent paper on the same topic which I suggest the authors to read and cite as well:

Abuelgasim E, Saw LJ, Shirke M, Zeinah M, Harky A. COVID-19: Unique public health issues facing Black, Asian and minority ethnic communities [published online ahead of print, 2020 May 8]. *Curr Probl Cardiol.* 2020;100621. doi:10.1016/j.cpcardiol.2020.100621

Response: We thank the reviewer for their comments and for appreciating the importance of our work. We reviewed the suggested reference and agree with this authors' emphasis on reporting of race and ethnicity data. We have incorporated this conclusion and the reference has been added in the discussion / conclusion section of our revised manuscript.

Reviewer: 3

Reviewer Name

Rebecca Stebbins

Institution and Country

UNC - Chapel Hill

Please state any competing interests or state 'None declared':

None declared.

Please leave your comments for the authors below

The article Racial and Ethnic Disparities in SARS-CoV-2 Pandemic: Analysis of a COVID-19 Observational Registry for a Diverse U.S. Metropolitan Population aims to understand the racial and ethnic disparities in SARS-CoV-2 infection within the United States. This is a critically important and timely research question. However, there are a few major issues with the analysis that should be addressed prior to publication. Please see below for specific comments and recommendations.

Major Issues:

- The authors should remove the ORs from Table 1 as they are meaningless.

Response: We appreciate the reviewer's comment. The odds ratios and their associated 95% Confidence Intervals in Table 1 provide an association between SARS-CoV-2 positivity (our primary outcome) and individual socio-demographic and comorbidity variables. These estimates may be regarded as un-adjusted estimates for the individual co-variables. Reporting of such estimates are preferred over reporting p values by several experts in the field and are also now preferred by several epidemiology and clinical journals. We choose to include these estimates in table 1, such that reasonable comparisons can be made between crude (un-adjusted) and adjusted estimates (table 4)

by the readers of our work. We have added footnotes for table 1 to provide interpretation of these odds ratios.

- The adjusted odds ratio results are very biased by the inclusion of population density and socioeconomic status – both of which are mediators of the race/ethnicity – SARS-CoV-2 relationship – in the adjustment set for the model. You cannot adjust for mediators of your relationship of interest, because you are blocking several of the causal pathways through which the exposure may be associated with the outcome. The authors should redo the analysis with appropriate consideration for the adjustment variable set. These mediators should only be included as part of the GSEM analysis.

Response: We appreciate the reviewer's observation about potential causal pathways. Our updated multivariable analyses do not include mediators for race and ethnic associations with SARS-CoV-2 positivity. Please see table 4 of the revised manuscript.

- The presentation of the adjusted odds ratios (and 95% CIs) for each independent variable included in the models is inappropriate unless interpreted correctly as the "direct association adjusted for all other covariates in the model". Reference this paper: Westreich D, Greenland S. The table 2 fallacy: presenting and interpreting confounder and modifier coefficients. *Am J Epidemiol.* 2013;177(4):292-298. doi:10.1093/aje/kws412

Response: Thank you for this comment. This interpretation of adjusted ORs and 95% CI has now been added to the footnote of table 4 in the revised version of the manuscript.

- The first sentence of the abstract (line 6) should be rephrased. I am concerned that the way it is phrased could imply a genetic difference/susceptibility by race/ethnicity to the virus, and that is not what is being asked in this paper.

Response: Thank you for this comment. We have updated the language in our revised version.

- The comparison of the disparities by race/ethnicity and age or sex should not be made. I'm referring especially to the first paragraph on page 14. While the differences in susceptibility by age and sex are likely due to legitimate differences in biological susceptibility to the virus, this is not the case for racial and ethnic minorities. The authors should take care to appropriately discuss the social, structural, and environmental basis for the difference in infection prevalence by racial/ethnic groups.

Response: Thank you for this comment. We have updated the language in our revised version.

- Finally, a more thorough discussion of the limitations including inflation of odds ratios, bias due to confounding and lack of sensitivity of the SARS-CoV-2 diagnostic test, etc should be added.

Response: We thank the reviewer for these suggestions. Based on the reviewer's recommendations we have added these additional aspects to our limitations section in the revised version of the manuscript.

#### Minor Issues:

- It would be more correct to use the term black to describe that racial category, rather than African American, as \_\_\_\_ American is generally used to describe immigrants or first-generation Americans.

Response: Thank you. The suggested changes have been made.

- Income-to-poverty ratio would be a much better indicator of socioeconomic status than simply household income, as it would account for the size of the household.

Response: Thank you for highlighting this important point. We unfortunately did not have household size information. We have acknowledged this in the limitations of our manuscript.

### VERSION 2 – REVIEW

<b>REVIEWER</b>	Michael Horberg Kaiser Permanente Mid-Atlantic Permanente Research Institute, USA
<b>REVIEW RETURNED</b>	09-Jul-2020

<b>GENERAL COMMENTS</b>	The authors have successfully addressed my concerns. I have no further criticisms, or any issues that need further resolution.
-------------------------	--------------------------------------------------------------------------------------------------------------------------------

<b>REVIEWER</b>	Rebecca Stebbins UNC - Chapel Hill
<b>REVIEW RETURNED</b>	08-Jul-2020

<b>GENERAL COMMENTS</b>	<p>The authors of the article Racial and Ethnic Disparities in SARS-CoV-2 Pandemic: Analysis of a COVID-19 Observational Registry for a Diverse U.S. Metropolitan Population have responded to reviewer comments and the updated submission is much improved. However, there are still a few small issues with the analysis that should be addressed prior to publication. Please see below for specific comments and recommendations.</p> <p>Issues:</p> <ul style="list-style-type: none"><li>• The authors write in their response that they no longer adjust for mediators. However, in the methods section of the paper, they state that they have determined a priori to adjust for a set of variables which includes the mediators they are testing. Then, they remove any covariate that is statistically significant as a mediator based on the GSEM. However, an adjustment set determined a priori should include only confounding variables and not include mediators. You cannot both determine your adjustment set a priori AND change it after statistical testing.</li><li>• The Charlson Comorbidity Index the authors use includes an indicator for diabetes. Therefore, diabetes should not also be included as its own covariate in the models – it is being adjusted for twice because of this.</li><li>• Page 3, lines 9-12: the sentence should all be in the same verb tense</li><li>• Page 3, lines 40-44: the abbreviations NHW and NHB are used but never spelled out</li><li>• Page 8, line 47: the authors refer to a “fully adjusted model” that has only been adjusted for “major covariate” (race, sex, and age). The model is not fully adjusted if it doesn’t adjust for all of the covariates you stated you would adjust for (authors had previously stated models would also adjust for CCI, etc.)</li></ul>
-------------------------	---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------



## VERSION 2 – AUTHOR RESPONSE

Reviewer 3 Comments:

The authors write in their response that they no longer adjust for mediators. However, in the methods section of the paper, they state that they have determined a priori to adjust for a set of variables which includes the mediators they are testing. Then, they remove any covariate that is statistically significant as a mediator based on the GSEM. However, an adjustment set determined a priori should include only confounding variables and not include mediators. You cannot both determine your adjustment set a priori AND change it after statistical testing.

Response: Thank you for seeking this clarification. We state that we a priori decided a starting point for analyses – not the final model that we would present. We have now modified language to clarify this further.

The Charlson Comorbidity Index the authors use includes an indicator for diabetes. Therefore, diabetes should not also be included as its own covariate in the models – it is being adjusted for twice because of this.

Response: In addition to a summary comorbidity score (Charlson Comorbidity Index), we were also interested in evaluating association of individual comorbidities that have been reported to be potential confounders, i.e. unequally distributed across various race groups and be independently associated with outcome. We therefore choose to present our data as such. We understand the reviewer’s concern about possible collinearity in the model. Upon formal testing; none of the variables in our final model demonstrate variance inflation beyond usually acceptable limits.

Page 3, lines 9-12: the sentence should all be in the same verb tense

Response: The verb tense has been updated.

Page 3, lines 40-44: the abbreviations NHW and NHB are used but never spelled out

Response: NHW and NHB are now defined at first use in the abstract and main manuscript text.

Page 8, line 47: the authors refer to a “fully adjusted model” that has only been adjusted for “major covariate” (race, sex, and age). The model is not fully adjusted if it doesn’t adjust for all of the covariates you stated you would adjust for (authors had previously stated models would also adjust for CCI, etc.)

Response: The wording in the manuscript has been updated to clarify.

## VERSION 3 – REVIEW

<b>REVIEWER</b>	Rebecca Stebbins UNC - Chapel Hill
<b>REVIEW RETURNED</b>	21-Jul-2020

<b>GENERAL COMMENTS</b>	Unfortunately, the authors have not sufficiently addressed my concerns about adjustment for mediators in their analysis. Adjustment for mediators is not appropriate for estimating total associations in epidemiologic research. However, if the authors are
-------------------------	---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------

	<p>intent on adjusting for mediators, they could refer to their effect measures more accurately by describing them as "direct" associations, adjusted for x mediators. This may still provide biased results, however. The authors might reference the following article for further clarification: Lorenzo Richiardi, Rino Bellocco, Daniela Zugna, Mediation analysis in epidemiology: methods, interpretation and bias, International Journal of Epidemiology, Volume 42, Issue 5, October 2013, Pages 1511–1519, <a href="https://doi.org/10.1093/ije/dyt127">https://doi.org/10.1093/ije/dyt127</a></p> <p>Additionally, if the authors believe the mediator may also have a confounding relationship with the exposure and outcome, they should state that as a reason for keeping the variable in their regression models.</p>
--	---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------

### VERSION 3 – AUTHOR RESPONSE

Please leave your comments for the authors below

Unfortunately, the authors have not sufficiently addressed my concerns about adjustment for mediators in their analysis. Adjustment for mediators is not appropriate for estimating total associations in epidemiologic research. However, if the authors are intent on adjusting for mediators, they could refer to their effect measures more accurately by describing them as "direct" associations, adjusted for x mediators. This may still provide biased results, however.

The authors might reference the following article for further clarification: Lorenzo Richiardi, Rino Bellocco, Daniela Zugna, Mediation analysis in epidemiology: methods, interpretation and bias, International Journal of Epidemiology, Volume 42, Issue 5, October 2013, Pages 1511–1519

Additionally, if the authors believe the mediator may also have a confounding relationship with the exposure and outcome, they should state that as a reason for keeping the variable in their regression models.

Response: We thank the reviewer for their comments. We would like to point out that the factors that demonstrated mediation have been removed from the final models (Table 4). As stated in our methods, our intent was to explore three potential pathways of mediation (i.e. population density, residence in lower income areas, and comorbidity burden). The pathways (factors) that did not demonstrate mediation (indirect effect) continue to inform the variance and degree of influence on the direct effect of minority race / ethnicity on SARS-CoV-2 susceptibility. Therefore, our final model includes these factors. In Table 4 (foot note 'C') we state that the ORs and 95% CIs represent the direct association adjusted for all other covariates in the model.

We have now provided additional language in the methods section to clarify the choice of our final model and have included the reference recommended by the reviewer.

### VERSION 4 – REVIEW

<b>REVIEWER</b>	Rebecca Stebbins UNC - Chapel Hill, USA
<b>REVIEW RETURNED</b>	23-Jul-2020

<b>GENERAL COMMENTS</b>	The authors state on page 8 lines 33-38 "We determined a priori to
-------------------------	--------------------------------------------------------------------

	<p>include all variables (age, sex, race, ethnicity, zip code household income, insurance type, zip population density and CCI) in our initial multivariable model.” It seems necessary to note that “a priori” means “based on theoretical deduction rather than empirical observation.” The authors explicitly state that they are determining their adjustment set a priori. However, a priori determined adjustment sets should only include the minimally sufficient set of confounders required to block all backdoor pathways to the outcome. Including mediators will bias the results (as referenced in my previous comments). The authors do not do this and instead use an approach closer to backward selection of variables, by determining whether covariates they think are mediators are statistically significant, and then removing those that are from the final model. If the authors are unwilling to remove the mediators from their initial adjustment set (as suggested in my prior comments), they could adjust their language to be more accurate and remove the statement that they determine their adjustment set a priori. If they do this and maintain the final model with the theoretical mediator, they should include a statement in the limitations section of the discussion that notes that the analysis may be biased due to the inclusion of a potential mediator in the regression models, though they have taken care to see that it is not a “statistically significant” mediator.</p>
--	----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------

**VERSION 4 – AUTHOR RESPONSE**

Please leave your comments for the authors below

The authors state on page 8 lines 33-38 “We determined a priori to include all variables (age, sex, race, ethnicity, zip code household income, insurance type, zip population density and CCI) in our initial multivariable model.” It seems necessary to note that “a priori” means “based on theoretical deduction rather than empirical observation.” The authors explicitly state that they are determining their adjustment set a priori. However, a priori determined adjustment sets should only include the minimally sufficient set of confounders required to block all backdoor pathways to the outcome. Including mediators will bias the results (as referenced in my previous comments). The authors do not do this and instead use an approach closer to backward selection of variables, by determining whether covariates they think are mediators are statistically significant, and then removing those that are from the final model.

If the authors are unwilling to remove the mediators from their initial adjustment set (as suggested in my prior comments), they could adjust their language to be more accurate and remove the statement that they determine their adjustment set a priori. If they do this and maintain the final model with the theoretical mediator, they should include a statement in the limitations section of the discussion that notes that the analysis may be biased due to the inclusion of a potential mediator in the regression models, though they have taken care to see that it is not a “statistically significant” mediator.

Response: We thank the reviewer for clarifying their viewpoint and for their continued time and effort in reviewing our work. As per their suggestion we have made two changes:

1. We have removed the language of a priori determination of adjustment set. Instead, we merely report the variables that were included in the initial model.

2. We have added a limitation stating that inclusion of potential mediators in the final model may have produced biased estimates. However, these potential mediators did not demonstrate a statistically significant indirect effect in our analyses.

**VERSION 5 – REVIEW**

<b>REVIEWER</b>	Rebecca Stebbins UNC - Chapel Hill, USA
<b>REVIEW RETURNED</b>	25-Jul-2020

<b>GENERAL COMMENTS</b>	The authors have sufficiently addressed my methodologic concerns.
-------------------------	-------------------------------------------------------------------