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

TITLE (PROVISIONAL)	The risk of COVID hospital admission and COVID mortality during the first COVID 19 wave with a special emphasis on Ethnic Minorities: an observational study of a single, deprived, multi ethnic UK health economy
AUTHORS	Singh, Baldev; Bateman, James; Viswanath, Ananth; Klaire, Vijay; Mahmud, Sultan; Nevill, Alan; Dunmore, Simon

VERSION 1 – REVIEW

REVIEWER	Sung-mok Jung Kyoto University School of Public Health
REVIEW RETURNED	24-Nov-2020

GENERAL COMMENTS	The authors describe the characteristics of COVID-19 confirmed cases from a city hospital in Wolverhampton, UK and also risk factors of the disease by ethnic groups. Since the present study was conducted using a high quality data which can represent local population of Wolverhampton, the manuscript provides helpful insights into the importance of locally focused public health strategies, especially in cities with diverse communities. However, there are some aspects of the manuscript that could be made clearer and I hope this suggestion make this manuscript to have the highest impact. [Major comments] 1. I am confused about the word “conveyance” in this manuscript. I would inclined that author should clearly clarify what kind of conveyance of risk was able to be shown using this cross-sectional data. 2. The authors assumed participants whose smoking information is unavailable as non-smokers, which can be plausible. However, since the proportion of those unknown cases are quite huge (i.e., 15%), I would inclined that authors should mention this assumption may provide biased results, especially in associations between outcomes and smoking history. [Minor comments] I would be inclined to mention the full name of IMD in the manuscript, before using the abbreviation. Also, I suggest authors to use a shorter sentence length for the reader-friendly manuscript (e.g., P11L23).

REVIEWER	Isabelle Bray UWE, Bristol England

REVIEW RETURNED	28-Nov-2020

| GENERAL COMMENTS | Thank you for the opportunity to read this paper, which I found very interesting. I agree with you that this type of analysis (of COVID outcomes for a defined population, with data on risk factors for the entire population) is exactly what has been lacking in the research to date on risk factors for COVID, and the relationship between these factors. My comments are mostly to do with how this important research is presented.
1. In the title (also used later in the paper), readers may not be familiar with the phrase 'health economy'. Could it be replaced by 'population'?
2. The objective stated in the Abstract is "To address the generalisability of covid-19's outcomes to the well-defined but diverse communities of a single city area". This is not completely clear. One suggestion is "To describe variations in Covid-19 outcomes within a well-defined but diverse single city area".
3. Also in the Abstract, please give some more detail about the outcome categories you study i.e. Covid admission, non-Covid admission, Covid death
4. In the Results section of the Abstract, please also specify baseline group when giving odds ratios. Similarly, when you state that the South Asian group had lower risks (4th sentence), please state lower than what.
5. The Introduction does not seem to fit with the rest of the paper. It states that "the objective was to establish a … multi-source …clinical data set used for the purpose of direct care,…to improve clinical coding and mortality recording accuracy, and to enable an informed understanding of factors influencing hospital activity, including admissions." This does not match the objective stated in the Abstract. Given the title and abstract, I would have expected to see a greater focus, with corresponding literature, on what we know about risk factors for Covid, some of the problems in studying this (which include not having the right datasets available) but also issues such as confounding.
6. The Methods section gives a thorough description of how the data were compiled. The methods of dealing with missing data are quite crude, was multiple imputation considered? This should be included as a limitation. Please give a range for the proportion of missing data across the variables you considered. The ethnicity groups are given as "Caucasian, S Asian, Black, Mixed ethnicity Chinese" and "unknown". Are "Mixed Ethnicity" and "Chinese" separate or combined? Was there an "other" category? Note that in Table 2, the categories are given as "White, S Asian, Black, other or unknown". It is not clear why these do not correspond to those in the Methods section.
7. The statistical methods section is rather brief, but describes to analyses – factor analysis and regression analysis. Please explain the purpose of each, as I am not clear how the factor analysis helps achieve the objective.
8. In the Results, I think the second sentence of the section on 'Hospital Admissions' should read "Compared to NA, there was an increased association of all variables with NCA and CA, including age,…".
9. In the same sentence, the "surrogate measures of dependency" are not defined in the Methods.
10. The next sentence states "Male gender, BMI, IMD and smoking status were significantly different" but does not explain different from what.
11. A general comment on the Results section – for me, the major |

	benefit of the analysis you have been able to do with this data set is that you can control for many important confounders simultaneously, but I don't think this comes across. It would help to make it clearer in the text and the tables where results are crude or adjusted, and adjusted for what. 12. The Discussion mentions the observed associations between smoking and better Covid outcomes, but I think this deserves further discussion/explanation. 13. In the Discussion, there is a section on 'the conveyance of risk'. Given that this was not stated earlier as a specific objective of the study, I think this would sit better as part of the 'implications' section. I believe that conveyance of risk has so far been based on the best data we had, but that was not very good. Your study and others like it will help us to better understand the complexities of the risks for different groups within our population, and therefore to improve the information about risk.

VERSION 1 – AUTHOR RESPONSE

Reviewer: 1

Comments to the Author
The authors describe the characteristics of COVID-19 confirmed cases from a city hospital in Wolverhampton, UK and also risk factors of the disease by ethnic groups. Since the present study was conducted using a high quality data which can represent local population of Wolverhampton, the manuscript provides helpful insights into the importance of locally focused public health strategies, especially in cities with diverse communities. However, there are some aspects of the manuscript that could be made clearer and I hope this suggestion make this manuscript to have the highest impact.

Thank you for these positive comments.

[Major comments]
1. I am confused about the word "conveyance" in this manuscript. I would inclined that author should clearly clarify what kind of conveyance of risk was able to be shown using this cross-sectional data.

Thank you for this comment. We believe that our findings demonstrate that appropriate communication, or conveyance, of risk is an important message arising from our study as it shows that local factors have a significant impact on assessed risk which can influence policy makers, clinicians and patients potentially to their detriment. We have clarified this as suggested by yourself and also moved the relevant section to the "implications" section (page 18, revised m/s) of the discussion as suggested by reviewer 2.

2. The authors assumed participants whose smoking information is unavailable as non-smokers, which can be plausible. However, since the proportion of those unknown cases are quite huge (i.e., 15%), I would inclined that authors should mention this assumption may provide biased results, especially in associations between outcomes and smoking history.

Thank you for this point – we agree that it is plausible, but certainly an assumption, that most of the smoking status-unknown subjects will be non-smokers. We have clarified that this is an assumption and also strengthened our discussion of this point (under "General Associations")

[Minor comments]
I would be inclined to mention the full name of IMD in the manuscript, before using

the abbreviation. Thank you for pointing this out – we have defined IMD in full at the first use on p5 (Methods). Also, I suggest authors to use a shorter sentence length for the reader-friendly manuscript (e.g., P11L23).

Thank you – we have removed a number of semi-colons and replaced the long sentences with shorter ones in several places including your example on page 11 of the original version.

Reviewer: 2

Comments to the Author
Thank you for the opportunity to read this paper, which I found very interesting. I agree with you that this type of analysis (of COVID outcomes for a defined population, with data on risk factors for the entire population) is exactly what has been lacking in the research to date on risk factors for COVID, and the relationship between these factors. My comments are mostly to do with how this important research is presented.

Thank you for your positive feedback and support for our assessment.

1. In the title (also used later in the paper), readers may not be familiar with the phrase 'health economy'. Could it be replaced by 'population'?

Thank you for this relevant point. We do believe that health economy has a specific relevance to our study but agree that clarification of the phrase would be helpful to readers and have therefore explained that this is a reference to local population health at first use in the body of the paper (discussion, original m/s p10 – "ethnicity", line 6).

2. The objective stated in the Abstract is "To address the generalisability of covid-19's outcomes to the well-defined but diverse communities of a single city area". This is not completely clear. One suggestion is "To describe variations in Covid-19 outcomes within a well-defined but diverse single city area".

Thank you for this suggestion which we have adopted.

3. Also in the Abstract, please give some more detail about the outcome categories you study i.e. Covid admission, non-Covid admission, Covid death

Thank you – we have clarified our outcome measures in terms of COVID and non-COVID admissions and deaths.

4. In the Results section of the Abstract, please also specify baseline group when giving odds ratios. Similarly, when you state that the South Asian group had lower risks (4th sentence), please state lower than what.

We have clarified that these ORs are in comparison to the White group

5. The Introduction does not seem to fit with the rest of the paper. It states that "the objective was to establish a … multi-source …clinical data set used for the purpose of direct care,…to improve clinical coding and mortality recording accuracy, and to enable an informed understanding of factors influencing hospital activity, including admissions." This does not match the objective stated in the Abstract. Given the title and abstract, I would have expected to see a greater focus, with corresponding literature, on what we know about risk factors for Covid, some of the problems in studying this (which include not having the right datasets available) but also issues such as confounding.

Thank you we have amended this and added some further context to the abstract objectives (although constrained by the journal's abstract word-limits), to improve clarity, and an addition (with

further references) to the first paragraph of the introduction in respect of COVID-19 mortality as well as a further addition to the second paragraph.

6. The Methods section gives a thorough description of how the data were compiled. The methods of dealing with missing data are quite crude, was multiple imputation considered? This should be included as a limitation. Please give a range for the proportion of missing data across the variables you considered.

A weakness of the study is that there was some missing data but this was very limited in magnitude and only affected 3 variables: BMI, smoking and ethnicity. We feel these were dealt with appropriately; for BMI as described in Methods; we coded all unknown smoking as non-smokers on the very likely assumption that the vastly greater majority were non-smokers, whist missing ethnicity was coded as unknown and analysed as such. Given the degree of completeness rather than incompleteness of our data, we consider our approach approximates to a complete case analysis, arising from significant effort on multisource data accrual, integration and quality. We thus do not feel that multiple imputation should be applied to replace missing data, since we do not feel this can possibly improve precision. In so doing, we are thus also avoiding the greater and well-recognized potential to introduce bias from a poorly fitting imputation models. We consider this to be a strength of the paper.

We have addressed this by adding a statement (with a reference on imputation models) to this effect to Strengths and Limitations.

The ethnicity groups are given as “Caucasian, S Asian, Black, Mixed ethnicity Chinese” and “unknown”. Are “Mixed Ethnicity” and “Chinese” separate or combined? Thank you for pointing out this confusion which has arisen because there’s a comma missing after Mixed ethnicity –we have inserted this to clarify that Mixed and Chinese are separate ethnic categories. Was there an “other” category? There was no “other” category beyond “Not Known” – the combined “other” was only introduced solely for the purpose of clarity and brevity in Table 2, in order to group all other categories including Chinese and Mixed because each individual group had no significant association and their numbers in these categories were small.

Note that in Table 2, the categories are given as “White, S Asian, Black, other or unknown”. It is not clear why these do not correspond to those in the Methods section. Please see explanation above - we have added this clarification to the legend in Table 2.

7. The statistical methods section is rather brief, but describes two analyses – factor analysis and regression analysis. Please explain the purpose of each, as I am not clear how the factor analysis helps achieve the objective.

Thank you for this point. In many studies the temptation is to use “all variables” without an *a priori* consideration of how those variables interact with each and if they do, how they congregate. So, for example, in this situation, where much has been made of deprivation and ethnicity separately, factorial analysis clearly show the 2 variables to be intertwined (and we go on to explore that in detail in this paper) as were (for obvious reasons perhaps) mortality and hospital admissions (thus making the point about studies that may focus on, or be weighted in favour of, hospital admission) and again palliative care registration and nursing home residency were co-associated – there were 9 such blocks. It is not common practice by many to show such detail, but our contention would be that such things really must be considered *a priori* in order to avoid “confounding effects and redundancy”. In truth and if reviewer 2 will forgive us, we only really put this in here, to make the point so that others who have published might reflect on how to avoid the “throw it all in and see what it shows approach ”. However the referee is kindly right to point out that we were too brief in explaining binary and multinomial regression analysis. We have made changes accordingly to the statistical methods section.

8. In the Results, I think the second sentence of the section on ‘Hospital Admissions’ should read

“Compared to NA, there was an increased association of all variables with NCA and CA, including age,…”.

Thank you, we have changed the wording exactly as suggested.

9. In the same sentence, the “surrogate measures of dependency” are not defined in the Methods.

Again thank you and we have also clarified in the text of the results (second sentence in “Hospital Admissions”).

10. The next sentence states “Male gender, BMI, IMD and smoking status were significantly different” but does not explain different from what.

Again thank you and we have clarified in the text (3rd sentence of the section named above)

11. A general comment on the Results section – for me, the major benefit of the analysis you have been able to do with this data set is that you can control for many important confounders simultaneously, but I don’t think this comes across. It would help to make it clearer in the text and the tables where results are crude or adjusted, and adjusted for what.

Once again thanks for highlighting this . The crude rates or absolute rates do not need clarification . The problem arises with Odds Ratios. Where a single independent variable is considered against a categorical dependent variable, then the resulting Odds ratio is unadjusted in the sense that it is not influenced by multiple other factors and in this circumstance equates to the relative risk . However, as in most of the analyses here, when multiple independent variables are considered against a dependent categorical variable, then the Odds Ratio for any single variable is influenced by the interaction with other factors and thus becomes adjusted, although with the large numbers we have used , the deviation of that from actual relative risk, whilst its exists, is usually small. We think this is a fine distinction and the best thing to do here is to remove the very few references to “unadjusted” when this is applied to ORs, which we have done , since this neither influences the intended meaning or the statistical confidence.

12. The Discussion mentions the observed associations between smoking and better Covid outcomes, but I think this deserves further discussion/explanation.

As mentioned in our response to reviewer 1 we have added further comment on this association to the discussion (General Associations) and included a reference to an article highlighting the ongoing study of this area.

13. In the Discussion, there is a section on ‘the conveyance of risk’. Given that this was not stated earlier as a specific objective of the study, I think this would sit better as part of the ‘implications’ section.

Thank you for this suggestion, we have moved this to implications.

I believe that conveyance of risk has so far been based on the best data we had, but that was not very good. Your study and others like it will help us to better understand the complexities of the risks for different groups within our population, and therefore to improve the information about risk. Thank you very much for this comment.

VERSION 2 – REVIEW

REVIEWER	Sung-mok Jung Kyoto University School of Public Health
REVIEW RETURNED	10-Jan-2021

GENERAL COMMENTS	The manuscript was well revised and I sincerely appreciate authors' efforts to reflect my comments.

REVIEWER	Isabelle Bray UWE Bristol, UK
REVIEW RETURNED	16-Jan-2021

GENERAL COMMENTS	Thank you for addressing my comments. I have just a few minor things to mention: 1. The text in italics just above Table 2 should say "other or unknown" rather than "other" or "unknown" 2. 4th line of discussion, "we raises" should be "we raise" 3. The first sentence of 'Strengths and weaknesses' could be clearer

VERSION 2 – AUTHOR RESPONSE

Thank you very much. We are most grateful to the editorial team and the reviewers for their highly professional review of our paper. We are pleased to resubmit with the minor revisions requested by reviewer 2, as follows:

> 1. The text in italics just above Table 2 should say "other or unknown" rather than "other" or "unknown"

Thank you for pointing this out - we have made this change.

> 2. 4th line of discussion, "we raises" should be "we raise"

Thank you for pointing this out - we have made this change.

> 3. The first sentence of 'Strengths and weaknesses' could be clearer

Thank you - apologies for the lack of clarity (and good grammar!) - we have amended the first sentence and this should now be much clearer to the reader.