

## Peer Review Report

# Review Report on Sex associations between air pollution and estimated atherosclerotic cardiovascular disease risk determination

Original Article, Int J Public Health

Reviewer: Jiawei Zhang

Submitted on: 19 Jul 2023

Article DOI: 10.3389/ijph.2023.1606328

### EVALUATION

#### **Q 1** Please summarize the main findings of the study.

This study explored the association between long-term exposure to air pollution and ASCVD risk using the UK Biobank cohort. Benefiting from the large cohort sample and detailed information, the author provided good evidence in supporting the association and further explore the heterogeneity of the association by sex and urbanicity. It is a nice and well-conducted study, and the overall results are interesting. I recommended it be accepted after some writing and small methodological issues to be solved below.

#### **Q 2** Please highlight the limitations and strengths.

The strength of the study benefits from the relatively large sample size, with more than 300 thousand participants, and detailed recorded information to explore the association.

The main limitation of the study raised from its cross-sectional study design, which hinders its further causal inference to the underlying causal link between long-term air pollution exposure and ASCVD risk. Besides, like many other studies using data from the UK biobank, the study population in this study is less representative, with only 5% of the response rate, and participants were clustered into 22 large cities in the UK.

#### **Q 3** Please provide your detailed review report to the authors. The editors prefer to receive your review structured in major and minor comments. Please consider in your review the methods (statistical methods valid and correctly applied (e.g. sample size, choice of test), is the study replicable based on the method description?), results, data interpretation and references. If there are any objective errors, or if the conclusions are not supported, you should detail your concerns.

Major:

1. Method: estimated 10-year ASCVD risk

in line 128-130: the author state "This model ... is calculated based on death due to coronary heart disease or nonfatal myocardial infarction or fatal or nonfatal stroke over a 10-year period for participants without a history of established CV diseases.". If the model is based on the cause of death, why no cause of death data is used in the later analysis? I think it should be a typo from the author.

2. Method: air pollution

Air pollution is estimated in 2010 for the majority of them. Given the participants were recruited from 2006 to 2010, such temporal differences in enrollment and air pollution exposure should be considered and proper methods should be applied to explore the issue, or at least for air pollution with data in multiple years, like NO<sub>2</sub> and PM<sub>10</sub>.

3. Method: Air pollution score

It is less clear how the Air pollution score is calculated, especially how the coef. of each pollution from the APS equation is obtained. If the coef is obtained from the coef in table 4-5, in col of model 2, given the different distribution of each air pollution, (as seen in table 1), please justify why it is appropriate to use the coef from

the 1 units increase for each pollution. Given the source of different pollution, 1 units increase in PM2.5 would be very different from 1 units increase in PM10.

4. Method: Statistical analysis: descriptive

in line 201, the author state "The study population's characteristics were presented using means and standard deviations (SD) for continuous variables", but in line 233 and table 1, the author summarized the distribution of air pollution as median and quantile. Please keep consistency in the methods and results.

5. Method: Statistical analysis: collinearity

in line 237 and Table 2, the authors show the results of collinearity, but it is not mentioned in the methods, especially which method was used here, please clarify.

6. Method: Statistical analysis: collinearity

Given the high collinearity between certain pairs of pollution, eg. NO<sub>2</sub> and PM<sub>10</sub>, PM<sub>2.5</sub> and NO<sub>x</sub>, et al. using multivariate regression to include all those pollution simultaneously would cause problems for the model fitting, and may generate bias estimation for all of the pollutions.

7. Method: Statistical analysis: linearity assumption for exposure

The author estimated the effect of air pollution under a linearity assumption. A linearity test should be conducted, as can be seen from the regression of air pollution in quantile, that Q<sub>2</sub> vs Q<sub>1</sub> has an inverse effect as Q<sub>4</sub> vs Q<sub>1</sub>. This indicated a possible non-linear association.

8. Method: Statistical analysis: potential clustering

The UK Biobank cohort enrolled participants from 22 cities, with a distinguished cluster structure among the participants, which affects both their exposure and healthcare access. Such clusters in the cohort should be considered in the statistical model.

9. Result: interpretation

In lines 243–245, after adjusting the covariates, the author said "Adjustment for all covariates did not affect this association", but from the results, the coef obviously changed, and changed a lot (eg. from 0.28 to 0.09 for ASCVD risk and PM<sub>2.5</sub>). Such 'not affect' interpretation should be reworded. Same in line 256 and others.

10. Result: the presence of results

The response variable 'ASCVD risk' is a metric to measure the risk, with the range of 0.00–1.00, or 0% to 100%. It may be not appropriate to only report the two digits of the coef for ASCVD risk. It is suggested to be reported in four digits as either 0.00% or 0.0000.

11. Discussion

The discussion of the study results is short and should be extended, In this study, the author estimated the sex-specific association, as shown in the study design and results presentation, But such content is not discussed in the paper.

Overall

The manuscript needs a thorough and carefully proofread. There are quite a lot of grammatical and language issues, which hinder the understanding of the manuscript.

Minor:

line 76, please add a citation for the source of data.

line 82, what do you mean "controlled environment studies"?

line 65, 66, 87 and more, please check the proper use of all the abbreviations, and make sure they have the right subscript.

line 99, I think it is an error here, that half a million participants are not the 5.5% of the overall UK biobank cohort but all. Please checked.

line 119-123, the paragraph is confusing here. As stated early that 399k participants are who did not have missing data, but later, 114k participants were further excluded due to missing data. Although the Figure 1 is clear, but the paragraph is confusing.

line 153, what is B coef? please change all the B coef to  $\beta$  coef.

line 223, change 'smocking' to 'smoking'

line 229, presented a higher 'mean' estimated .... 'mean' is missing here.

line 235, delete 'for'

line 267, what does 'localization' mean here? Should be 'location'

#### PLEASE COMMENT

**Q 4** Is the title appropriate, concise, attractive?

Yes

**Q 5** Are the keywords appropriate?

YES

**Q 6** Is the English language of sufficient quality?

No

**Q 7** Is the quality of the figures and tables satisfactory?

Yes.

**Q 8** Does the reference list cover the relevant literature adequately and in an unbiased manner?

Yes

#### QUALITY ASSESSMENT

**Q 9** Originality



**Q 10** Rigor



**Q 11** Significance to the field



**Q 12** Interest to a general audience



**Q 13** Quality of the writing



**Q 14** Overall scientific quality of the study



#### REVISION LEVEL

**Q 15** Please make a recommendation based on your comments:

---

Major revisions.