Reviewer’s report

Title: Association between long-term exposure to ambient air pollution and prevalence of diabetes mellitus among Malaysian adults

Version: 0 Date: 20 Aug 2019

Reviewer: Evi Dons

Reviewer's report:

In this manuscript, the authors associate exposure to air pollution to the prevalence of diabetes mellitus in the general population of Malaysia. The sample is relatively big (28,000 to 56,000) and three cross-sectional surveys are available (2006, 2011, 2015), each time in a representative sample of Malaysian adults. In general, the paper is well written and easy to read, but at some points it needs restructuring. There is lots of literature on the topic of onset of diabetes, prevalence of diabetes, or diabetes-related morbidity and mortality associated to air pollution. This paper adds to the literature as this is the first study in Malaysia.

Major comments:

- It's hard to judge the exposure measurement error resulting from the use of fixed monitoring stations. How large is the area? How many monitoring sites are there? What's the mean distance from the homes to the monitoring stations? Nowadays, modeled air pollution at the residential address is the standard, if not also accounting for travel behavior. Given this, I would prefer not to call it 'individual exposure' in the manuscript.

- It would make more sense to either study lifelong air pollution exposure (including for example moving houses), or relate onset of diabetes to the air pollution exposure during the same year. Now annual exposure estimates are hypothesized to be representative for lifelong exposure - or at least this is what I assume as it is not specified in the paper. Please comment on this in the discussion and justify your choice.

- Results/Discussion: the (statistically significant) effects that are seen, are not properly interpreted. For example, on line 226 you mention a statistically significant association between exposure to NO2 and DM for 2015. However, the effect is in the other direction than initially hypothesized (higher air pollution, lower risk) - please discuss.

Minor comments:

- Revise the abstract. The results-section does not contain actual results from the association between DM and air pollution exposure. Please specify exposure assessment in the abstract in more detail (annual air pollution at the nearest fixed monitor).
- Graphical abstract: Replace SO3 with SO2. The graphical abstract suggests that mechanisms are studied in this manuscript which is not the case.

- Line 135: Based on which variables does the author conclude that this is a representative sample?

- Both undiagnosed and underdiagnosed are used throughout the manuscript. Please be consistent and use the right term.

- Line 155-156: Could the change in physical activity assessment impact the results? Please discuss. Self-reported BMI may also lead to bias?

- Line 168-169: Specify the number of stations. From Fig.2 it looks like 14 stations, but then how do you get 25% of industrial stations? Are the industrial sites representative for the exposure of the population?

- Line 178: In case of multiple stations per state, you average the air pollution by state. At the same time, you state that the exposure of an individual is based on the most adjacent monitoring station. The station located within the same state as the residential location of the individual is not necessarily closest station for the individual - in this case which station is used? Please clarify.

- Line 178: Change places: 'residential coordinate' and 'air monitoring station'.

- Table 1: It should be discussed in the manuscript that the sample size was larger in 2006. How may this impact the odds ratios or the confidence interval?

- Figure 2: I'm not sure this is really necessary. Trim down the figures to fit on one page, move to the supplemental information, or delete.

- Figure 3: I believe you overinterpreted these figures. I don't feel like the trends for NO2, SO2 and PM10 are similar, for example there is a steep increase for PM10 while for NO2 the curve is almost flat.

- Lines 210-217: This paragraph does not belong here: it's not a result of this study. Move it to the Methods or to the Discussion.

- Lines 234-240: Be consistent in the reference to the survey: both NHMS 2011 and 2011 NHMS are being used.

- Figure 4: The diagonal line in the figures should be removed.

- Figure 4, panel (b): The CI for PM10 in 2006 in the graph does not seem to correspond to the numbers presented.

- Lines 353-362: This part should be moved to the introduction.
According to the authors, the different results in 2006/2011 and 2015 could be explained by the presence of bias in the exposure assessment. However, why would this error be different in different years?

One option to tackle this would be to only include individuals that live close to a monitor in a sensitivity analysis.

The authors state that they did not correct for BMI and physical activity. However, I understood from the Methods-section that the authors did correct for BMI and physical activity?

The authors refer to PM10 as coarse particulate matter. This is not correct. PM10 includes ultrafine particles, fine particles, and coarse particles - usually PM10-2.5 are considered coarse particles.

Line 475: Rephrase: You do not study the onset of DM, so rephrase 'developing DM'.

Line 476: Typo: 'increased' should be 'increase'.

Level of interest
Please indicate how interesting you found the manuscript:

An article of limited interest

Quality of written English
Please indicate the quality of language in the manuscript:

Needs some language corrections before being published

Declaration of competing interests
Please complete a declaration of competing interests, considering the following questions:

1. Have you in the past five years received reimbursements, fees, funding, or salary from an organisation that may in any way gain or lose financially from the publication of this manuscript, either now or in the future?

2. Do you hold any stocks or shares in an organisation that may in any way gain or lose financially from the publication of this manuscript, either now or in the future?

3. Do you hold or are you currently applying for any patents relating to the content of the manuscript?

4. Have you received reimbursements, fees, funding, or salary from an organization that holds or has applied for patents relating to the content of the manuscript?
5. Do you have any other financial competing interests?

6. Do you have any non-financial competing interests in relation to this paper?

If you can answer no to all of the above, write 'I declare that I have no competing interests' below. If your reply is yes to any, please give details below.

I declare that I have no competing interests

I agree to the open peer review policy of the journal. I understand that my name will be included on my report to the authors and, if the manuscript is accepted for publication, my named report including any attachments I upload will be posted on the website along with the authors' responses. I agree for my report to be made available under an Open Access Creative Commons CC-BY license (http://creativecommons.org/licenses/by/4.0/). I understand that any comments which I do not wish to be included in my named report can be included as confidential comments to the editors, which will not be published.

I agree to the open peer review policy of the journal