Reviewer’s report

Title: The neglected burden of tuberculosis disease among health workers: a decade-long cohort study in South Africa

Version: 0 Date: 08 Mar 2017

Reviewer: Christiaan Mulder

Reviewer's report:

Dear authors,

With great interest I have read your manuscript. I appreciate the design chosen in which surveillance data has being used to the compare the risk of TB between healthcare workers and the general population in South Africa.

I have the following suggestions:

1) Line 115: why were being a contact of a TB patient and previous TB not included as a covariate? These could have affected both the risk for TB disease as it could have stimulated people to work on TB control.

2) Line 135: what was the basis of these score thresholds? Is this based on any evidence?

3) Line 158: Are HW's included in the TB incidence rate of the general population? If so, this has inflated the TB incidence in the general population and subsequently have led to an underestimation of the differences in risk of TB between the general population and the HW's. Would it be possible to repeat the analyses excluding the HW's from the general population?

4) Figure 1: size and colors of shapes are unclear. The arrows of the three shapes below point to nowhere. This figure should be revised.

5) Line 178: That should be listed 'Figure 2'

6) Table 2: Put the overall TB incidence among HW in the correct column

7) Could the authors explain why the adjusted RR in HW working 1-5 years was almost twice as high as the unadjusted RR?
8) It's not exactly clear from the methods and table 4 if occupation and duration of employment were or were not included in the adjusted regression model.

9) Line 255: I would be good if the authors could elaborate on the possible changes in the data capture system or reporting practices.

10) The basis of the analyses in this study is the probabilistic record linkage. Although the authors try to explain how this works I think a technical appendix explaining more about the theory of Newcombe could be useful. What is lacking is how the linkage weight is exactly given, and also why 70% was used as the threshold for further analyses. For example were many records close to this 70%, and how would have affected this the outcomes if they would have been included?

**Are the methods appropriate and well described?**
If not, please specify what is required in your comments to the authors.

Yes

**Does the work include the necessary controls?**
If not, please specify which controls are required in your comments to the authors.

Yes

**Are the conclusions drawn adequately supported by the data shown?**
If not, please explain in your comments to the authors.

Yes

**Are you able to assess any statistics in the manuscript or would you recommend an additional statistical review?**
If an additional statistical review is recommended, please specify what aspects require further assessment in your comments to the editors.

I am able to assess the statistics

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

Acceptable

**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