**Reviewer’s report**

**Title:** Automated versus Physician Assignment of Cause of Death for Verbal Autopsies: Randomized Trial of 9374 Deaths in 117 Villages in India

**Version:** 0 **Date:** 10 Feb 2019

**Reviewer:** Carina King

**Reviewer's report:**

This paper presents a prospective study, which randomly allocated VAs to be analysed using physician coded versus algorithm coded cause of death. The authors have a large volume of information they present, and have interesting findings!

My main comment is around the firmness of the conclusions in light of the method. From my interpretation of the paper, you compared two different groups of people that were demographically similar, for their cause of death. But the method of assigning the outcome and collecting the information to allocate the outcome was different between arms (i.e. the open narrative may have influenced the answers given to the closed questions, and created a different rapport between interviewers and respondents leading to different types of information being collected). Therefore, it is not clear to me that you can conclude the <100% concordance is due to limitations in the automated algorithms… but may be due to the method, or actual population differences in cause of death between the two groups. You could present the concordance between the physician coding in the automated arm and the physician arm as a key result more clearly, and also present the sensitivity for the two physicians between each other in Table 3, as a direct comparison. Additional file 11 suggests that there genuine differences in the distributions of death (e.g. ischaemic heart disease 17% in the physician arm, and 12% in the automated arm - classified by physicians)

Abstract:

- Line 15: add "automated" method for clarity
- In the methods section, can you state the algorithms that were being tested.
- The conclusion in the abstract suggests that the physician assigned cause of death was a genuine "gold standard". This statement needs to be tempered.
Introduction:

- In the first paragraph, this will differ hugely by region, but also by age group. E.g. as more women deliver in facilities, the majority of early neonatal deaths will occur in healthcare facilities. But this will not be the case for older groups necessarily. This could be clearer.

- Paragraph 3 - is there evidence that physician coding is still the most commonly used form of assigning COD? I would rephrase to include the word 'historically' unless you have an up to date citation for this.

- Paragraph 3 - in studies where the whole sample or population of VA s have been analysed by both physician review and automated algorithm, will not be subject to a bias related to lack to randomisation. This would only be an issue when a sub-set of the population is recruited using non-random approaches and allocated to different arms. It is not clear that you are referring to the later, in terms of previous studies.

Methods:

- Trial description: how were the villages selected? Purposive or random? And within these villages, it was not entirely clear, were all households eligible for interview? What about households with no response?

- Participants: the information about the results being similar according to time since death should be in the results section, not here. Same comment for the age of deaths. And why have you chosen not to show these results?

- Randomisation: You state that the surveyor was blind to the allocation, but they would have known, as you state that the physician arm included an open narrative and the algorithm arm did not. Therefore, as they conducted the interview they would be aware of the trial arm.

Results:

- Page 9, the comparison to other studies should be presented in the discussion, not in the results, and needs a reference.

- It would be helpful to show the cause of death concordance in the main text, rather than a supplementary file.

- Can you add the information about the 2 physician agreement, for all age groups, and state the number that required a third physician to arbitrate?
Discussion

- Page 12, in the limitations, points 2 and 3 appear to be making the same point to me.

- It would be good to see how the costs were generated - $3 is very low cost, so it would be good to see what amount of time and physician hourly rate in this setting was.

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.

No

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

Statement on potential review bias
Please complete a statement on potential review bias, considering the following questions:

1. Did you co-author any publication with an author of this manuscript in the last 5 years?

2. Are you currently or recently affiliated at the same institution as an author of this manuscript?

If you can answer no to all of the above, write 'I declare that I did not publish with these authors in the last 5 years and also meet the affiliation criteria”. If your reply is yes to any, please give details below.

I declare that I did not publish with these authors in the last 5 years and also meet the affiliation criteria

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