Reviewer’s report

Title: Development and validation of a clinical prediction model to risk stratify patients presenting with small pulmonary nodules. A research protocol

Version: 0 Date: 24 Sep 2018

Reviewer: Laura Bonnett

Reviewer's report:

Many thanks for submitting this interesting protocol. I have a few comments which I believe will enhance the manuscript and ensure that an independent research team could replicate/undertake the research. In particular, I think that the subheadings need rethinking as the text under each sub-heading was not what would traditionally be associated with that description in my opinion. Perhaps a longer methods section without subheadings may be more appropriate?

Other more minor comments are as follows:

1. Page 2, line 7 - Add "(BTS)" after "British Thoracic Society"

2. Page 2, line 16 - Add a reference to justify the percentages quoted in this paragraph.

3. Page 2, line 22 - I think the authors mean "multivariable" rather than "multivariate"? Multivariable implies multiple variables but one outcome, whilst multivariate implies multiple outcomes.

4. Page 2, line 26 - Define PET-CT (or at least point the reader to the list of abbreviations)

5. Page 2, line 33-34 - Add a reference to justify the statement about "unrepresentative of the risk in the wider population"

6. Page 2, line 37 - Quantify what you mean by "discriminate well". For example is there a c-statistic which is commonly reported in this clinical area?

7. Page 2, line 42 - It is not clear from the introduction why this novel model is required (it is more clear from reading the discussion however). Justify why a new model is needed, and thus why it is not appropriate to update an existing prognostic model.

8. Page 2, line 44 - Related to the previous point, what is the accuracy of current models?
9. Page 3, line 15 - Rather than "...the model has access..." I think it would be clearer to say "the model development team has access..."

10. Page 3, line 17 - Link the AI model to the current model better - should this text be in the introduction or discussion rather than the study design section perhaps?

11. Page 3, line 30 - Can you add a potential end date for data collection?

12. Page 3, line 43 - Link this section to the sample size section of text

13. Page 3, line 58 - How is a "technically inadequate CT" defined?

14. Page 4, table 1 - Justify this list of candidate predictors based on existing literature or similar

15. Page 4, line 35 - Link this text to the multiple imputation section

16. Page 4, line 35 - Have you considered using an additional category for these variables such as "not relevant"?

17. Page 4, line 59 - Justify why you want to build two models. Have you considered variable selection methods such as backward selection?

18. Page 5, line 11 - Add a reference for the fractional polynomials approach.

19. Page 5, line 17 - With an outcome of benign/malignant logistic regression would be the appropriate model - be clear about this when describing the generalised linear model framework.

20. Page 5, line 52 - Justify the use of the unreliability index by adding a reference. It is not "standard practice" within clinical prediction models in my opinion.

21. Page 5, line 56 - Justify the use of false negative criteria for producing risk groups. Harrell recommends using tertiles (or similar) of the linear predictors to produce risk groups. Ensure that the end user can establish probabilities of the event for patients in each risk groups.

22. Page 6, lines 22-30 - Consider including a calibration plot too

23. Page 7, line 27 - Resolve the ethics approval text

24. Page 7, line 31-34 - There are no data currently available so update this text accordingly
I look forward to reading the updated protocol.

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

An article whose findings are important to those with closely related research interests

**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.