Author’s response to reviews

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

Authors:

Jason Oke (Jason.oke@phc.ox.ac.uk)
Lyndsey Pickup (lyndsey.pickup@optellum.com)
Jerome Declerck (jerome.declerck@optellum.com)
Matthew Callister (matthew.callister@nhs.net)
David Baldwin (david.baldwin@nuh.nhs.uk)
Jennifer Gustafson (jennifer.gustafson@ouh.nhs.uk)
Heiko Peschl (heiko.peschl@ouh.nhs.uk)
Sarim Ather (sarim.ather@ouh.nhs.uk)
Maria Tsakok (maria.tsakok@ouh.nhs.net)
Alan Exell (Alan.exell@ouh.nhs.uk)
Fergus Gleeson (fergus.gleeson@ouh.nhs.net)

Version: 1 Date: 02 Nov 2018

Author’s response to reviews:

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

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

We assume you mean these “…that were malignant in the Yonemori, Li, VA and Mayo cohorts (75%, 62%, 54% and 23%)”

We have added the relevant references

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.
We take your point and have changed this.

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

We have done this now.

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

We have added a reference for this.

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?

We have changed this to

Whilst these models discriminate well (C-statistic ranging from 73.5% to 91.6% [10] they differ considerably in their estimate of risk for smaller nodules.

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.

We thank the reviewer for this comment. We feel that the justification is given in the paragraph that starts “To date, five studies“ but take the point that it is not clear to the reader.

We have added this sentence to the beginning of this paragraph in an attempt to make this clearer.

“Existing nodule prediction models have been developed in highly selected patient groups with high rates of malignancy and give very different estimates of risk for smaller nodules.”

8. Page 2, line 44 - Related to the previous point, what is the accuracy of current models?

We think that this has been answered somewhat with the response to point 6

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..."

Agreed, I think this was a typo and has been corrected. Thanks for spotting this.

“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?

We are not sure about this but will amend the introduction if the editor feels it is necessary.
11. Page 3, line 30 - Can you add a potential end date for data collection?

We cannot guarantee the exact date but we have removed the “predicted to be” part of the sentence.

12. Page 3, line 43 - Link this section to the sample size section of text. We have added the following sentence - (see sample size section for justification of sample size).

13. Page 3, line 58 - How is a "technically inadequate CT" defined? The definition is quite long so we have added this to the appendix.

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

We have added the following to sign post the multiple imputation section

(see multiple imputation section for details of our method to handle missing data).

16. Page 4, line 35 - Have you considered using an additional category for these variables such as "not relevant"? No we had not considered using a “missingness” indicator.

This is an interesting idea and we have amended the imputation section to include this as a potential method for dealing with missing data.

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

Good point. On reflection we may want to explore a parsimonious model (potentially more than nodule size alone) using backward selection. We have amended the text accordingly – including the “Criteria for statistical evaluation of the model” section.

Please note that we have pluralised model throughout the manuscript now.

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

A reference has been added.

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.

Thanks – we have changed this now to

Regression coefficients for both models will be estimated using maximum likelihood estimation in a logistic regression model.

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.
On reflection we are going to remove this measure for this reason.

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.

The reason for this is context specific. The prediction rule will need to have good rule out properties and make reliable negative predictions. False positives are less important as further surveillance and testing will be carried out if the risk prediction is high. We have added words to this effect justifying our choice of risk groups.

22. Page 6, lines 22-30 - Consider including a calibration plot too We have added calibration plot to the list of results we intend to report.

23. Page 7, line 27 - Resolve the ethics approval text Thanks for picking this up and we apologise for not spotting this ourselves.

We have changed this now to

In keeping with the Governance Arrangements for Research Ethics Committees (GAfREC) and University Oxford Policy, research undertaken on data collected before formulation of a study where data are anonymous to the researcher, does not require ethics approval. In keeping with the requirements of the Research Governance Framework, both sponsorship and Trust Management approval will be sought before the research is undertaken.

24. Page 7, line 31-34 - There are no data currently available so update this text accordingly

Not entirely correct. The model will be built on the retrospective data (which has already been collected) and validated on the prospective study data (which has started to recruit). The statement in the manuscript was the text suggested by the journal. See here -

https://diagnprognres.biomedcentral.com/submission-guidelines/preparing-your-manuscript/research
Reviewer 2

1) The introduction section recites Fleishner and BTS guidelines for monitoring of lung nodules according to the size. I understand the rationale for including this in the manuscript but it occupies a large part of the introduction and can be hard to follow for non-specialists. I would propose to include this information in a table, which is easier to comprehend.

We have simplified this and put the pertinent information into a table and reduced the text accordingly.

2) It would be helpful to deemphasize on the size of the lung nodule, as stated in the previous comment, and mention other factors they use in their prediction model, eg other radiographic characteristics of lung nodules (size, density, etc) as well as patient characteristics (age, history, etc) and their significance as predictors of malignancy in the introduction section.

We think that the action taken for the previous suggestion has done enough to deemphasize nodules size. We have also modified the sentence

“Risk stratification tools that incorporate the age of the patient, their smoking history and their respiratory health could assist clinical decision making and reduce unnecessary investigations and quickly identify those at higher risk.”

3) The authors mention a number of other prediction models previously developed for the same outcome. They mention only validation study of one of the prediction tools in the UK. It would be useful to add countries/continents, clinical setting and ethnicity of participants used for each model in the development and the validation study.

We have amended the paragraph starting with “Existing nodule prediction models…” to include the country of origin. We feel that the clinical setting is already described (e.g. Brock was screening cohort, The VA model participants were mostly male smokers and the Mayo cohort were a single cohort managed in the 1980's. The percentage of nodules that were malignant in the Yonemori, Li, VA and Mayo cohorts)

Ethnicity is not reported in any of the prediction model papers or the UK validation study so we don’t think it is worth commenting on in the protocol.

We have added more information about the clinical setting of the UK validation study.
4) Given that this study is based in UK patients, it is important to mention the procedure for a referral for a CT in the national health system. This is to help appraise the external validity of findings, considering that regulation of referrals for computed tomography is dependent on the structure of the national health system.

This is an interesting point but we don’t feel it is feasible to mention all of the possible reasons for people to undergo CT scanning in the UK. CT’s can be ordered from anywhere in the health service including A&E or primary care and nodules can be detected from images not necessarily directed at the chest. CT imaging of the chest itself (which frequently discovers incidental pulmonary nodules) is ordered for many reasons and may be done in patients with traumatic injury, for suspicion of pulmonary embolism or screening for coronary artery disease. CT of other areas may also lead to incidental detection of nodules, e.g. colonography will find pulmonary nodules that will need further evaluation. Therefore, our nodules could arise from a very wide range of clinical scenarios and this should ensure wide applicability of the finished model.

5) Please elaborate on your rationale for including the specific variables for your prediction model as well as you will each variable (e.g. categorical, continuous)

These have been selected either because they have been shown to be associated with the risk of nodule malignancy or benignity or have been used in other nodule prediction models.

For the assessment of the outcome please elaborate on how will nodule will be assessed (follow up period and size comparison)

This was described in sub-section ‘Outcome’.

We have added more text to make this clearer.