Reviewer’s report

Title: Large-scale external validation and comparison of prognostic models: An application to chronic obstructive pulmonary disease

Version: 0 Date: 14 Nov 2017

Reviewer: Orestis Efthimiou

Reviewer’s report:

This paper compares several prognostic models for 3-year all-cause mortality in chronic obstructive pulmonary disease. Thus, it has the potential to be of particular importance for this disease. Moreover, the authors present their analysis step by step (including software codes), so as to facilitate researchers that would like to perform a similar analysis. In addition, the paper is well written, and the conclusions drawn are clear, and supported by the data.

I am a biostatistician, so I will focus my review mainly on the statistical methods of this paper and not on the clinical context. My main concern regarding the paper is that the statistical methods that the authors used have not been published yet. In the methods' section (line 185) the authors state that "We followed a pre-specified study protocol (please see supplementary material) and described the detailed statistical methods elsewhere (submitted to the American Journal of Epidemiology)".

Before considering this paper for publication I would wait until after these methods have been published. It is rather irregular to publish an application of some methods before these methods have been peer-reviewed. E.g. what if there is a change in the methods during their revision? I would wait before publishing this paper, not because it uses wrong methods or because it is not interesting enough, but because these methods need to first be published in a peer-reviewed statistical journal.

Apart from (and in addition to…) the fact that the methods used are still unpublished, the authors do not describe these methods in a 'user-friendly' manner. The main paper only has few details, while the appendix is very big and not so well structured.

In what follows I discuss some more specific comments to the authors, most of which regard clarifications and are rather easy to address in a revision.

Major comments:

1. The methods' section of this paper should give sufficient level of details of the methods used, and also refer to the appendix for any additional information. The authors do present some of the statistical details in the appendix, e.g. where they give the protocol, but this is not done in a structured way, meaning that there is no single section of the appendix providing all the details
needed. E.g. several aspects of the methods that the authors used are not in the protocol, but are scattered all over the appendix (see following comments).

2. The appendix is huge (266 pages), and is poorly connected with the main paper. In order to help your readers find their way through this large appendix, I would suggest that in the main paper you can refer to specific parts or pages of the appendix. E.g. instead of writing "R codes are available in the supplementary material" (main paper, line 256) you could write "R codes are available at Section XXX or page YYY of the supplementary material".

3. The structure of the appendix is rather chaotic. E.g. in line 281 of the main paper it writes "Heterogeneity in the MSC analysis was evaluated by the pooled heterogeneity variance among groups (τ2_pooled)", but it doesn't write what this pooled heterogeneity is. I was then trying to find the definition of tau_pooled in the methods section of the appendix, but it was actually 100 pages later. And then, methods for inconsistency was another 50 pages later. I.e. the description of the methods is intertwined with presentation of results. This makes difficult keeping track of the statistical methods that the authors used. Again, it would help if all methods were in a single section of the appendix (not the protocol). Also, as previously mentioned, it would help if in the main paper when you refer to a method/model/result you also refer to the part of the appendix where this is defined/shown.

4. Table 3: direction of effects is not clear. Does a positive number mean that the score in the column or the score in the row is better? E.g. for ADO vs. GOLD, in the lower triangle of table 3, I read the result 0.083. Does this mean ADO has a higher or lower AUC from GOLD? Also, Table 3 is quite busy, with lots of information. It might help readers if the order of the scores in rows and columns was rearranged, from best to worst (or the opposite).

Minor comments:

1. Line 290, page 14: The authors say that the performed multiple imputations with missing data, but do not give much detail. In the appendix it is mentioned that variance-covariance matrices were calculated using bootstrapping. It is unclear to me how these two aspects are combined. Did the authors bootstrap on each imputed dataset separately to calculate the variance-covariance matrices and then combined them with Rubin's rules? In the appendix this seems to be implied but it is not 100% clear, or I might have missed it (and again there is no mention in the main paper). In any case, I would welcome a clarification in the section.

2. Line 162, page 7: in the Background section it might be useful for some readers to add some text describing in broad terms how a prognostic score works.

3. Table 3 (main paper) and Table 17 of the appendix are both titled "League table presenting the Mixed Score Comparison (MSC) meta-analysis (lower-left half of the table) and the direct random-effects meta-analysis (upper-right half of the table)" but they give different numbers. Please clarify why there are differences.
4. Line 180, page 7: Some of the papers cited at this point (it writes "similar to network meta-analysis (NMA) of randomized trials [43-50]") are just NMA applications, and are thus not very useful for the reader who would like to know what NMA is. I would consider removing some of the citations that are practical implementations of NMA and maybe adding some methodological/didactical papers that describe the NMA model, its assumptions etc.

5. Line 271, page 13: it writes that "We used Multiple Score Comparison meta-analysis, a frequentist two-stage network meta analytic approach to compare all scores within a single analytical framework, inspired by similar approaches developed for treatment effectiveness comparison [56], also referred to as "design inconsistency approach"[61].' The last part ("…also referred to as…") is ambiguous and rather misleading. Design inconsistency is a method for assessing inconsistency in NMA. I am not sure which approach the authors say that it is also referred to as "design inconsistency".

6. Line 276, page 13: it writes that "In the Stage I of the MSC meta-analysis, estimates are averaged within each group…". It would be helpful to clarify here what are the estimates that are averaged within each group (i.e. ΔAUC)

7. Line 378, page 22: it writes "We did not use a τ2 for each group since this is not recommended for groups with a single cohort". I don't really understand this sentence, if there are groups with a single cohort then you wouldn't be able to estimate tau anyway, correct? Also, it is not recommended by whom? Please add a citation to clarify the source of this recommendation.

8. Line 415, page 23: there is a typo in the phrase "Nowadays, there are no treatments to lower the risk of mortality are available for patients with COPD yet…"

9. Figure 1 in the main paper is a hazy, the groups are unreadable.

10. Figure 1 in the Appendix is very difficult to read, to the extend that it is pretty much unusable. The lines are overlapping, I couldn't figure out what is connected to what. Putting the scores in a circle would be much better, I guess.

11. Appendix page 221: some references are not displayed properly

12. The appendix needs some editing for minor typos here and there.

**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 recommend additional statistical review

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