Reviewer’s report

Title: A novel method for interrogating Receiver Operating Characteristic curves for assessing prognostic tests

Version: 0 Date: 21 Jun 2017

Reviewer: Mariska Leeflang

Reviewer's report:

The authors present a novel and interesting method for assessing clinical utility of diagnostic tests. In itself, the method is not very difficult and the reasoning behind it makes sense. So if we address this study as a demonstration of a method, then the manuscript could either have been much shorter or some more application examples could have been provided.

On a different level, the manuscript raises a lot of questions about underlying mechanisms (do sensitivity, AUC and specificity indeed remain stable when the prevalence changes) and terminology (is a screening test the same as a prognostic test). I have tried to address all these concerns as well, but I am not sure if they should be treated as real concerns.

1. **Title:** I would have liked ROC in full in the title (Receiver Operating Characteristic curve) and probably also "of a prognostic test" at the end.

2. **Clinical utility** is a term that is not well defined. Both CDC and Bossuyt et al (Clin Chem 2012) define it as the likelihood of or a measure for improved outcomes for patients when the test has been used. I do not think predictive values cover such a definition, so maybe another term (for example predictive value?) is more helpful. In any case, the authors should provide a definition of the term utility if they use it.

3. The authors claim that this approach is novel. Which is surprising, given its relative simplicity. But after checking the literature, I couldn't find any similar articles either (which is not a guarantee of course). I did find a paper that looks remotely the same: Mandic et al, J Cardiopulm Rehab and Prev 2008; 28(6): 415-19. Maybe it is worthwhile to have a look at it and discuss it in this manuscript.

4. The authors start with the Jungner and Wilson criteria for screening tests. I am not sure if the claim really can be made that these criteria functioned as gatekeepers to the introduction of novel tests. Any test can be brought on the market without rigorous regulation. If I remember correctly (but I haven't checked), then these criteria are mainly useful to decide whether a screening test may indeed be useful to be used in screening programs. I think removing the part "and acted as gatekeepers to the introduction of novel tests" should suffice.
5. In line with that, I think the authors mix up screening and prognosis. I know that these terms as well are used in different ways, so perhaps it would be good here as well to define what the authors exactly mean with screening test or prognostic test. To me, a screening test detects disease that is already there, but perhaps in an early stage. While a prognostic test to me predicts whether something will occur in the future. I know that some readers may think the same, but I also see these terms being used interchangeably. So the more reason to use clear terminology, to provide definitions and to be consistent in terminology.

6. Page 6, line 2: first write ROC in full.

7. On page 7, the authors state that a screening test is fundamentally different from a diagnostic test in that sense that screening tests can only estimate a probability. I disagree with this statement, as I think that a screening test is not necessarily different from a diagnostic test (but a real prognostic test may be different); and because I think that a diagnostic test in itself is rarely capable of confirmation presence or absence of disease.

8. Page 7, lines 10+11: I do not understand the last part of this sentence ("these statistics therefore..."). Maybe it can be removed? It seems a bit out of place and irrelevant here.


10. Equations A and B should be fairly well known by the audience of this journal. So these can be shortened or even removed. Same may be true for C and D, but I see that these are useful to understand the later equations E, F, G.

11. Part of the Results section seems to be more appropriate under the Methods section, as it explains the methods they used to derive the data from their example.

12. Page 13, lines 10 to 15: so the idea is to have two tests or testing algorithms; one with a high PPV and one with a high NPV. What would be the consequences if these two tests were used in the same patient and then both turn out to be positive? Or the one with the high PPV is negative and the one with the high NPV is positive? Maybe elaborate on this a bit more in the Discussion section (I know the authors explain that the two tests should be used in relation to a specific context, but a bit more explanation is required here).

13. Page 13, lines 16-22 and Figure 5. It is not clear to me how the sub-figures in Figure 5 are derived and what they add to the storyline. Could the authors please add some text to the Methods paragraph explaining all this? And are Figures 5A and 5B really necessary? I also read from the Figure caption that these figures are based on hypothetical data: this should have been made explicit in the Methods section.

14. Perhaps another real-life example, including practical considerations, would have been helpful.
15. One more general comment: these methods imply that the sensitivity, specificity and AUROC of a test are the same, irrespective of whether it is used in a 5% prevalence situation or a 20% prevalence situation. Although I do understand that this is mathematically indeed the case, in practice the population with a prevalence of 5% will be a different population than that of 20%. And these different populations may cause differences in test accuracy (sensitivity, specificity). Could the authors please address this point in their Discussion section?

16. Are there any limitations to this method or study that should be mentioned in the Discussion section?

17. I have checked the website the authors mention at the end of the Conclusion section. It would have been helpful to mention this website a bit earlier and perhaps to use some more pictures from this website in the manuscript. It seems to be very helpful. Only point is: the bottom right figure states the AUC that belongs to the ROC, but it is depicted next to another graph. So it could be interpreted as the area under that bottom-right curve.

18. Are there any benefits for the authors if readers access this website? In that case, it should be mentioned, I think.

19. Figure 3 and 4: I have checked the calculations in Excel. These seem to be correct, but for Figure 3A they are only correct if you start with the sensitivity of 50%. But that is not how the methodology is described. In the Methods section, the authors describe that the PPV and specificity are known and that the sensitivity is calculated. However, if I use equation F to calculate the sensitivity for the numbers presented in figure 3A, I get three different sensitivities: 48%, 54% and 57%, while the figure shows three times sensitivity of 50%. So apparently, the authors used the sensitivity of 50% to calculate the corresponding specificity and after rounding, this results in the same numbers as in the Figure. But this is not what was stated in the Methods section.

**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:

Needs some language corrections before being published
**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