Reviewer’s report

Title: The Brier score does not evaluate the clinical utility of diagnostic tests or prediction models

Version: 1 Date: 24 Jul 2017

Reviewer: Kathleen Kerr

Reviewer’s report:

The paper is greatly improved over the first version. However, in my opinion it could benefit from further simplification. The paper's main point is that the Brier score is similar to the misclassification rate (and equal to MCR in some contexts) in that it treats the two types of errors (false positives and false negatives) as being equally bad. In the vast majority of clinical contexts the two types of errors are not equally bad, and therefore the Brier score is not appropriate. The fact that the Brier score is a proper scoring rule means it satisfies a necessary condition for being a good metric, but that is not enough (not sufficient).

Specific comments:

1. Line 107. I don't think you want the first E[ ]. I think you want to define the brier score as BS=E[D=X]^2.

2. The expression at line 114 is incorrect.

3. Line 115. It is not necessary clear what is meant by "equivalent to the misclassification rate." It is cleaner and clearer to simply say "equals."

4. The material on lines 123-129 does not really contribute anything. It seems to just state some "textbook" results.

5. There is some confusion in the paper distinguishing estimands and estimators. For example, the expressions at lines 142 and 144 are presented as the definition of Net Benefit but actually give estimates of Net Benefit (and in standard statistical notation, "Net Benefit" should have a "hat" on it).

6. line 171-173: you might comment that what the authors call "Method 1" is just a miscalibrated version of what the authors call "Method 2"

7. I think it would be helpful to readers not familiar with Net Benefit to make explicit the costs and benefits corresponding to different p_t. Using a risk threshold of 5% implies that the benefit of treating an event patients is X times greater than the harm of treating a non-event patients. I recommend doing this every time a specific p_t is used in an example.
8. Page 10, top. "considering all patients either positive or negative." Are the authors assigning a predicted risk of 1 when all patients are considered positive and a predicted risk of 0 when all patients are considered negative?

9. Line 197. How are the authors computing AUC for binary tests?

10. Line 281. I think it is too imprecise to say that the Brier score is an "unsound" metric. It is a mathematically "sound" metric in the sense that it is proper, but it does not summarize the utility of a risk model for recommending treatment.

11. It would be worthwhile to note in the paper the assumptions implicit in Net Benefit: that all cases have the same expected benefit of treatment and all cases have the same expected harm of treatment.

12. In my previous review I suggested the authors cite their reference [7] (Kerr, 2014) as being the most comprehensive evaluation of NRI statistics. For the particular criticism the authors call out (that NRI statistics tend to overstate the incremental value of a new predictor), a more appropriate citation would be the authors' reference [5] or Pepe, Janes, and Li, JNCI 2014.

13. Line 60. I don't think "accuracy" is well-defined here. Consider replacing with "performance."

14. The draft contains quite a few typos and grammatical mistakes.

**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.