Reviewer’s report

Title: Multiple aneurysms in subarachnoid hemorrhage - Identification of the ruptured aneurysm, when the bleeding pattern is not self-explanatory. Development of a novel prediction score.

Version: 0 Date: 08 Oct 2019

Reviewer: James Burke

Reviewer's report:

Please include all comments for the authors in this box rather than uploading your report as an attachment. Please only upload as attachments annotated versions of manuscripts, graphs, supporting materials or other aspects of your report which cannot be included in a text format.

Please overwrite this text when adding your comments to the authors.

Hadjiantanasiou et al report the results of a novel machine learning model to predict symptomatic aneurysm in patients with MIA using data from a retrospective single-center case series. Their model retained 3 factors — aneurysm size, shape and location — and correlated with their gold-standard assessment in all cases from their small external validation.

I think that the clinical problem is a real one (albeit likely applicable to a small number of cases) and the general approach and dataset are reasonable. However, I suspect that the model's performance is overstated and have concerns about the accuracy of their gold standard.

Major Issues:

1. Model development context — As the authors argue in the discussion, the key question is not to predict which aneurysm in clinical cases where it's obvious — but rather to predict in the borderline cases. However, to predict in that context, the prediction model should be evaluated in non-obvious or troublesome cases?

It's reasonable to argue that the model should be developed on all cases because it is likely that similar factors predict rupture in troublesome and non-troublesome cases (although this should be explicitly spelled out). The problem arises with statements about how the model performed. Saying, for example, that it was accurate on all 32 external validation cases is non-informative. If 28 of those cases were obvious, then its model only predicted correctly in 4 potentially informative cases.

2. Gold standard limitations — The big problem here is the conclusion that all of the endovascular cases were assigned correctly in the derivation population. First, if that is true, then it argues that there is no need for a clinical score — standard care is right every time! Second, it can't be true. Assuredly sometimes, the wrong aneurysm was coiled — its just that there is no
way to know if the right one was coiled unless an untreated aneurysm re-ruptures and that case is captured by the records. (There is no documentation of how the follow-up was performed, so I have less than complete confidence this was the case) It seems unlikely to be a coincidence that all of the cases in the derivation and validation datasets where there was a disagreement on which aneurysm was symptomatic were surgical.

Given that the gold standard is probably right in most cases, this isn't a fatal flaw, but rather a source of measurement error that should be discussed and something that should lead the authors to framing their conclusions very cautiously.

3. Unjustified complexity of the statistical modeling approach — The authors applied a component-wise gradient boosting model to fit these data. Yet, there is no justification for why this approach was selected. In general, the simplest and most transparent method with the smallest number of ad hoc assumptions should be used. So, why not just use logistic regression with a random patient level intercept? I'm not sure what their model adds above a LR model.

More to the point — their approach requires additional assumptions, is known to lead to overfit models unless trained on enormous datasets, is not transparent, and is likely less reproducible than LR. So, why use it?

4. Reliability of model predictors — How reliable is the assessment of the data elements that go into the prediction? While size and location should be reasonably reliable, I'm mostly worried about "regular" vs. "irregular" which necessarily implies some judgement.

Minor Issues:

1. Incomplete reporting of the internal validation — How much did the model differ across the internal validation sample? How much evidence of overfitting is there in the base model?

2. More description is needed on the process for dealing with missing data. How much data was missing and for each of the elements? Why average imputation as opposed to multiple imputation?

3. I'm not following the logic for the "random guessing" bars in figure 2. with random guessing, shouldn't a guess be right 50% of the time? it looks like its closer to 1/3…

4. To the understand the clinical applicability of the score, it would be helpful to see how often there is a "close call" between multiple aneurysms.

5. Overall model discrimination seems like it should be estimable with this model using a c-statistic or other conventional techniques.
Are the methods appropriate and well described?
If not, please specify what is required in your comments to the authors.

No

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.

No

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:

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.

None

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