Author’s response to reviews

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.

Authors:

Alexis Hadjiathanasiou (alexis.hadjiathanasiou@ukb.uni-bonn.de)
Patrick Schuss (patrick.schuss@ukb.uni-bonn.de)
Simon Brandecker (Simon.Brandecker@ukb.uni-bonn.de)
Thomas Welchowski (welchow@imbie.meb.uni-bonn.de)
Matthias Schmid (matthias.schmid@imbie.uni-bonn.de)
Hartmut Vatter (hartmut.vatter@ukb.uni-bonn.de)
Erdem Güresir (erdem.gueresir@ukb.uni-bonn.de)

Version: 2 Date: 23 Jan 2020

Author’s response to reviews:

Dear Reviewers,

We would like to thank you for thoroughly reviewing our manuscript after the major revision and your additional comments! We have addressed each of the concerns mentioned point-by-point below and hope that they are adequate in this regard. Please note that our responses follow each of the reviewer’s initial comments and all changes in the manuscript are tracked. All corrections from the first revision are still highlighted with red color.

Response to reviewer

Reviewer #1, Mr. James Burk, MD, MS:

Point 1

1. My primary concern is the model's performance is overstated. It’s been externally validated on 5 tricky cases — that's great — but, it’s hardly persuasive evidence that it should be widely applied and/or that it is likely to have robust predictiveness outside that setting.
So, for example, in the results of the abstract its essential to not merely say, "predicted correctly in all cases, even in the most debatable and challenging cases", but to also point out how many such cases there are (n=5)

We explicitly referred to the 5 challenging cases as well in the Abstract as in the main manuscript and pointed out that because of the very small number of patients with such constellation greater volumes of challenging cases are need for a better validation and calculation of the performance of the prediction score.

Abstract, Results: Lines 39-42: ‘Even in the five most debatable and challenging cases assessed in the period of prospective validation, for which the score was designed for, the ruptured aneurysm was predicted correctly’

Abstract, Conclusions: Lines 45-47: ‘In a small prospective sample, the prediction score performed with high accuracy but larger cohorts for external validation are warranted.’

Discussion:

Lines 316-317: ‘…and the troublesome cases with non-definite bleeding pattern are infrequent.’

Lines 324-325: ‘Given the fact that the challenging cases needed for prospective validation are overall rare, multicenter independent data is necessary.’

Point 2

In terms of modeling complexity — I think the authors raise several very reasonable points. but, if the primary virtue of their algorithm (compared to logistic regression) is to help in identifying which variables to include in the model — they need an argument about why that is a virtue in this case. My sense is that this is a situation where we have reasonably strong prior from earlier work that gives us a good idea of what variables should be included in the model. So, why is an algorithm that helps with variable selection helpful? The main concern here, in a small sample, is that the model will be overfit. And the more parameters that the model requires (and this approach effectively fits more parameters than LR) the more likely it is to be overfit. The fact that there is no evidence of overfitting, in a very small external validation dataset, is NOT strong evidence that over-fitting doesn't exist. if there were an enormous external validation on a large, relevant sample, this problem would be mitigated. But, given that the external validation is on a tiny sample…it means there is no way to realistically test for overfitting. if so, isn't avoiding overfitting a greater priority than optimizing variable selection?

We agree that a large external validation sample would allow for a much better evaluation of overfitting than the smaller sample considered for our analysis. On the other hand, we have used the best available methods to make model validation in our paper as fair as possible. In particular, we used nested cross-validation to determine the number of boosting iterations and the performance of our prediction model. It must be emphasized that the performance of logistic
regression would have suffered from the small sample size as well, since the coefficients of this model would also have to be estimated from the small sample and would thus be subject to potential overfitting (even if all variables have been pre-specified correctly). Furthermore, based on our experience, we are confident that overfitting due to variable selection is not a severe problem, since the number of patients (n=252) was relatively large compared to the number of candidate variables (only 10).

Statistical analysis: Lines 124-128: ‘We note that major overfitting issues due to variable selection are unlikely in this case, as the number of patients (n=252) is sufficiently large compared to the number of candidate variables (10, i.e. 25 patients per variable). In addition, to counteract the variability induced by variable selection, the coefficient estimates obtained from gradient boosting are regularized such that they automatically improve predictive performance and reduce overfitting.’

Again, we would like to thank you for thoroughly reading the manuscript after the revision and for the final suggestions for further improvement of our manuscript.

Yours sincerely,

Alexis Hadjiathanasiou, M.D.

(on behalf of the authors)