Reviewer’s report

Title: Development and validation of a novel prediction model to identify patients in need of specialized trauma care during field triage: design and rationale of the GOAT study

Version: 0 Date: 28 Jan 2019

Reviewer: Alexander Hapfelmeier

Reviewer’s report:

This is a very well written manuscript about the design and analysis of a model development and validation study. The authors have also addressed most of the major issues of such studies. I'd still like to give two recommendations which might be crucial for the development of a well performing prediction model.

First, the authors might want to address the imbalance of the binary outcome classes by over- or undersampling or even more sophisticated methods for imbalanced classification problems. These approaches can also be part of the tuning process to optimize the sampling of classes.

Second, the main idea of the study is to replace a "simple" model by a more sophisticated one. Still the authors chose to investigate only a single alternative, which is the gradient boosting decision tree. This is a severe limitation. Though it could be discussed as such in the "Limitations" section, I strongly recommend performing a benchmark study of several approaches, for example including Lasso, Random Forests, Neural Networks, Support Vector Machines etc. Otherwise the authors risk ending up with a model which is too simple again. There are several software implementations, for example in R, that provide all the tools one needs to perform such a benchmark study.

Further minor issues:

- Line 83: "All hospitals …" Is it a reasonable assumption that all hospitals will participate in the study and that all outcomes will be available?

- Line 85: "Patients that die during transportation will be excluded". To my understanding it is possible that the death of such patients is related to a wrong decision on where to go. The authors should discuss whether use of a combined endpoint "death or ISS > 16" is more meaningful.

- Line 131: Please do also state that variable selection will not be related to the prediction performance, if not performed within the training data.

- Line 132: How will the additional features be "engineered"?
- Table 1: I don't see the importance of the "anticipated functional form". Why is this information needed?

- Line 197-198: It is not clear which "studies" it is referred to.

Level of interest
Please indicate how interesting you found the manuscript:

An article of importance in its field

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