Reviewer's report

**Title:** Variable performance of models for predicting methicillin-resistant Staphylococcus aureus carriage in European surgical wards

**Version:** 2  **Date:** 27 October 2014

**Reviewer:** Tjibbe Donker

**Reviewer's report:**

Predictive models are often used to optimize screening efforts for health care-associated infections, such as MRSA, by identifying high risk patients on admission. The authors describe how model selection influences the utility of these predictive models in clinical practice. They show that commonly used stepwise variable selection models can give poor predictive results, and advocate the use of models that take uncertainty into account.

The paper is well-written and the analysis is excellent, I enjoyed reading it. The authors have been thorough in analyzing the differences between the model results, and present their result meticulously. The paper holds an important message on the disadvantages of a commonly used method, and is therefore of great importance to both the scientific community and policy makers.

I am left with only one major question, about the methods, and a few minor points throughout the paper.

**Major Compulsory Revisions**

The second paragraph of the methods “To evaluate the predictive performance...” starts with a description of the random partitioning of the cohort in a derivation and validation dataset in a 1:1 ratio. These are used to respectively construct the models and test their predictive performance, calculating the c-statistic. This description is clear and the analysis is sound, however, the third paragraph of the methods starts again with a partitioning of the dataset (or cohort), this time described as 50% derivation and 50% validation, to test the clinical utility.

I might misunderstand this, but this second division seems to be redundant. Couldn’t the original partitioning be used to calculate sensitivity, PPV, etc.? The 1:1 ratio and 50%-50% are exactly the same, and the models are constructed the same way. The original validation dataset(s) could just as well be used for assessing the clinical utility, or am I mistaken?

**Minor Essential Revisions**

In Methods, Statistical analysis, the second sentence describes why a stepwise model would perform inferior, citing Babyak 2004. In my view, this is typically something mentioned in the discussion. However, I have a question related to
this sentence: would the predictive performance of the stepwise model increase if the number of positive patients increased? Wouldn’t this decrease the chance of overfitting? In other words, how are the presented results dependent on the incidence of MRSA on admission?

How many times was the cross-validation performed? The methods describe a repeated random sub-sample, but don’t specify the number of repeats.

In Methods, Statistical analysis, second paragraph, 5th sentence, it says the c-statistic ranges from 0.5 to 1. Although 0.5 does mean that there is no predictive ability, the statistic itself has a theoretical range from 0 to 1, with values below 0.5 being less meaningful (they would imply negative predictive ability, meaning something is seriously wrong with the model). Nonetheless, I would advice stating the correct theoretical range of the statistic.

Discussion, 2nd paragraph, the advantage of targeted screening (reduced screening burden by 2/3) is mentioned. This is compared to universal screening, I assume? It is not obvious from the text here.

The conclusion of the paper focuses on the difference in performance between the different models, and rightfully points out that model uncertainty should be taken into account. However, all models show poor predictive power, is the difference between them not a small effect compared to the overall poor predictive power?

Discretionary Revisions

Results, 4th paragraph, last sentence: Why is the c-statistic value for the hospital Barcelona so much higher? Is it caused by the low number of cases, or are these cases all “typical” patients, recently hospitalized, nursing home residents with wounds?

Level of interest: An article of outstanding merit and interest in its field

Quality of written English: Acceptable

Statistical review: Yes, and I have assessed the statistics in my report.

Declaration of competing interests:

I declare that I have no competing interests