Reviewer's report

Title: Validation of prediction models: examining temporal and geographic stability of baseline risk and estimated covariate effects

Version: 0 Date: 01 Feb 2017

Reviewer: Maarten van Smeden

Reviewer's report:

The authors present an interesting paper, illustrating how to evaluate temporal and geographic validity of a prediction model. The methods are illustrated with a heart failure case study. I believe this paper is a good fit for BMC diagnostic and prognostic research. I do have some remarks that have to do with the writing and presentation. In addition, the presented methods rely much on p-values calculated with unknown procedures. In my view, these procedures should be clarified more and better acknowledged as a limitation of the presented methods.

- p4 l21 "geographic" should be "geographic validity"

- p4 l31 "in a sample different from that in the model was developed" should be "in a sample different from that in which the model was developed"

- p6 l7 "From this model, we extracted from the fitted" should be "From this model, we extracted the fitted"

- p6 l26 I assume the authors are not talking about the results of their analyses yet, but only about the assumptions made that come with these models? I guess then it should read: ".. variable is assumed constant across …". Same holds for l41.

- p6 l33 I suggest that the authors replace "above" by "from Model 1"

- p7 l36-41: can the authors be a bit more specific than just mentioning "for computational reasons, we were unable to fit a full random coefficients model…"? I wonder what was the real reason for not fitting this model: computational time? Software errors? Or something more fundamental as in a non-identifiable model problem? Same holds for p10 l26.
- p10 l9-14: I am not sure that I can interpret the magnitude of effect by looking at the range in linear predictor effects. The authors claim this range (0.76,1.26) is evidence of only modest effect. Is there any guidance on what constitutes a modest effect?

- p11 l21-26: can the authors provide a point estimate for the interaction effect next to the p-value?

- p12 l4-12: I think the conclusion of that time effect does not vary across hospitals is rather strong (P=.10). First of all, it is unclear what the power is of the test for the 3-variance terms (or even which test was used for calculation for this p-value). Second, there is no measure of effect size provided (effect was claimed for the effect of the linear predictor (P=.03)). Especially in these cases, using the P=.05 cut-off criterion for claiming effect seems rather arbitrary. I think these limitations should be made more clearly. Ideally, a measure of effect size would be provided (is there one?).

- Results section: there are some p-values reported without mentioning of the test statistic or procedure. As this paper aims to provide guidance, the information about which test statistic/procedure is used is relevant for readers who want to use the same approach. I also think it is good practice to provide information on test statistics and degrees of freedom when providing p-values.

- Results section: some (single) p-values are given with a ">"-sign. Why is that?

- Table 1: 95% CI of Model 2 has more precision (4 digits) than Model 1 (3 digits)

- Appendix: something went wrong with the notation of model 5.

**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:

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