Reviewer’s report

Title: The effects of misclassification in routine health care databases on the accuracy of prognostic prediction models: A case study of the CHA2DS2-VASc score in atrial fibrillation

Version: 0 Date: 02 Oct 2017

Reviewer: Richard Stevens

Reviewer's report:

The aim, to examine how misclassification in electronic health records affects validation of a prediction model, is extremely important. Validation studies in routine health data are increasingly common. UK readers will be conscious that the QRISK scores replaced previous cardiovascular risk tools nationwide following successful validation in routine healthcare data.

There are three parts to this paper: (1) a study of the amount of error in the electronic health record; (2) a study of the effect of the error on hazard ratios; and (3) a study of the effect of the error on prediction rule validation studies. The first and the third are both publishable in their own right and together they make a unique and powerful paper. Including (2) as well overburdens the paper and makes it hard to follow - I would move this material to an appendix or possibly a separate paper, or remove it altogether as less novel than (3).

I struggled to follow the rationale for the analyses of hazard ratios. In the Methods and the Results they seemed to be a minor distraction in a paper that is actually about validation rather than model fitting. In the Discussion I found that the 'hazard ratio' angle was actively distracting and confusing. For example on page 14, the remarks about misclassification "averageing out" in multivariate analyses - is this a reference to the Cox models or to the validation analyses or both? While trying to decide I lost the thread.

There is an established literature about how classification error affects regression models e.g. Zucker & Spiegelman (2008) Statistics in Medicine 20:1911-1933. If you are going to include the Cox model analyses in this paper then it seems an omission not to discuss the results - in the Discussion section - in terms of this established literature.

In the Introduction, you cite several references 3-7 for measurement error models. I have usually seen the terminology of 'measurement error' used for the case of continuous variables and the term 'classification error' for categorical or binary variables. Your reference 5 is the Carroll book - I only have the first edition to hand, which includes less than a page on classification error, but perhaps the second edition has more? Likewise the book by Wayne Fuller (which your software has accidentally cited as Wayne AF instead of WA Fuller!) emphasizes continuous variables, doesn't it? I don't have my copy of Fuller to hand so accept my apologies if I'm wrong about
that; if I'm right, could you reference specifically the chapters or sections that are most relevant to dichotomous variables?

The first paragraph on page 14 ends with a remark that "validation studies ... commonly report the observed risk per score" and that at score 1 "this risk was nearly twice as high" as the reference value. This doesn't seem to match with Table 4, and in fact Table 4 suggests that numbers of events are too small, at the lower score levels, to make any comment on observed risk.

When calculating c-statistics for the model, did you use a c-index for censored data or did you assume follow-up was complete to two years?

You might attract criticism for using mortality, rather than stroke, as the outcome in a CHADS2 project, but I agree with you that for this methodological study of misclassification error validation against mortality can still be informative. There are already clinical papers assessing the usefulness of CHADS2 for predicting mortality (e.g. Kurtul & Acikgoz 2017 A J Cardiology 120:8-14 or Yang et al 2016 PLOS One) albeit in more acutely ill patients.

Given that the main aim of the paper (according to the abstract) is about validation, rather than model fitting, it seems disproportionate that 2 of the 5 items under "Data analysis" are for model fitting. I would strongly recommend you give more detail on the validation methods. Firstly, on page 9 item 4 under "Data analyses" says "We then used Cox proportional hazards models to analyze the predictive performance of both scores." Many readers may be aware of the Cox modelling for deriving a model but not aware of its use in calibration so you may need to give slightly more detail and a reference e.g. chapter 15 of the Steyerberg book.

Likewise, there are surprisingly few validation metrics in the results (have I missed some?). There are numerical values for c-statistics, in the text, the concordance table in Figure 1, and a graphical representation of calibration in Figure 2. It would greatly enhance the applicability of the paper to include the P/O statistic for overall calibration, and the currently fashionable measures calibration slope and calibration intercept. Although I agree with Royston & Altman (2013, BMC Med Res Methodology, 13:33) that the former is actually a measure of discrimination or fit rather than calibration, it is informative and increasingly widely used even if it is not well-named.

Perhaps consider too whether you wish to report on sensitivity & specificity (or PPV) for the CHADS2 score at a specific threshold.

Although my comments above are extensive, this is an excellent piece of research. The depth of your work on the data and the decision to address validation make a unique and important manuscript.
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.