Reviewer's report

**Title:** Formal and informal prediction of recurrent stroke and myocardial infarction after stroke: a systematic review and evaluation of clinical prediction models in a new cohort

**Version:** 1  **Date:** 17 December 2013

**Reviewer:** Derrick Bennett

**Reviewer's report:**

Major compulsory revisions

1. The authors should make it clearer c-statistic (described as AUROCC in the manuscript), obtained from Cox regression and logistic regression are slightly different but have a similar interpretation.

2. It is not clear to me why the authors have used a random-effects meta-analysis. If this is to account for heterogeneity, surely the authors should ascertain whether heterogeneity is present, and then try and investigate potential clinical or statistical sources of this heterogeneity. In addition, if there was publication bias or small study bias then the random effects model of DerSimonian and Laird would give more weight to the smaller studies. The use of a random effects model thus requires further justification.

3. I am not convinced that conventional funnel plots are good at identifying publication bias (or small study bias), especially when there are only a few studies. It would be better to perform “contour enhanced funnel plots” as suggested by Peters, JL et al. (J Clin Epidemiol 2008;61:991-996), or better still, simply order the studies by sample size, which is a useful way to visual whether effects sizes are attenuated in larger studies. It may be interesting to see each study’s model performance statistic ordered by the size of the model development sample.

4. The authors did not extract information on whether internal validation was used in the development of the prediction rule. Although I note that internal validation (via bootstrapping) was included as part of their quality criteria in Table 1 they prioritized external validation over this. However, good internal validation would have guarded against over optimism in the model fitting process. It would also be useful to see how many studies reported internally validating their model as part of the quality criteria figure as although external validation of a “prognostic model” is essential internal validation using bootstrapping is important particularly when the developmental sample is small and/or there are a large number of candidate predictor variables.

5. The authors only seem to report discrimination for the Edinburgh Stroke Study evaluation cohort. Would it not also be appropriate to assess calibration as well? Ideally, this could be assessed graphically with supplemental information on goodness of fit.
6. It is interesting that the “clinical gestalt” based expert opinion of was similar or sometimes slightly better in terms of discrimination than the “prognostic model”. The paper does not how many experts were consulted in order to derive the discrimination statistic nor does it mention how the formal and informal rules compared in terms of calibration.

Minor essential revisions

7. Prognostic models (even well developed and validated ones) have not been widely accepted by clinicians. The findings that the “clinical gestalt” is about the same (and sometimes better), than the formal prognostic models for recurrent stroke and myocardial infarction in the evaluation cohort used in this report is likely to increase scepticism about the value of prognostic models. The authors should make it clearer that it is poorly developed prognostic models with methodological weaknesses as outlined in the report are the ones that are most likely to be not much better than clinical judgement of an expert in the field.

8. In several places throughout the manuscript the authors use the term “multivariate” which should be “multivariable”.

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