Reviewer's report

Title: Risk Score to Predict Gastrointestinal Bleeding after Acute Ischemic Stroke

Version: 2
Date: 6 June 2014

Reviewer: Colin Crooks

Reviewer's report:

Major compulsory revisions are listed below:

The authors do not describe in the methods how long patients were followed up for an occurrence of GIB. Were post discharge bleeds captured? What happens if patients present to a different hospital with GIB after discharge? Table 1 and the discussion suggest only in patient bleeds were included. If so this is problematic as patients with longer periods in hospital have more time to have a GIB event observed. Therefore predictors of longer inpatient stay will be predictors of GIB in this study, even if not associated with GIB. This could explain or least seriously confound the findings of this study.

For the same reasons Including length of hospital stay as a predictor of bleeding is unhelpful as clearly the longer of hospital stay the longer the period in which to capture a GIB and therefore it will itself become a 'risk factor'. Additionally the length of hospital stay is not known at the start of the an admission which contradicts the authors own statement that only risk factors known at the time of the stroke were included.

There is a comment in the discussion that seems to suggest that the time of GIB was not recorded? If this is the case how can the researchers know for certain whether the GIB was before or after the stroke? Or whether the recording of risk factors was definitely before the stroke and GIB (this is not fully described in the methods but the discussion mentions information collected in the period after admission was included in the study)?

Aren't the responses of the GCS score covered in the NIHSS, which also includes an assessment of consciousness? So why include the GCS in addition to the NIHSS?

Age was included as a linear variable in the model, yet the final score uses categories. How were these categories selected? Was there any test done for a departure from linear trend for age?

How reliable is the recording of the risk factors and outcomes within the datasets? Has there been any validation of the variables used in the dataset?

What is the distribution of the score in the study population? Do most people have a score of zero or is it evenly distributed?

What is the cross over between the CNSR and the CICAS? If these datasets
come from the same hospitals then the CICAs would not be an external validation.

The detail for the comments above was lacking in the methods. I had to deduce it from the discussion and tables but this should have been in the methods.

**Level of interest:** An article whose findings are important to those with closely related research interests

**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