Reviewer's report

Title: A Survey of Variable Selection Methods in Two Chinese Epidemiology Journals

Version: 1 Date: 14 July 2010

Reviewer: Rolf Groenwold

Reviewer's report:

This is a concise, well-written, and easy-to-read manuscript on the reporting of variable selection methods in two Chinese epidemiology journals. Please find my review comments below.

Major compulsory revisions

1. I think two issues are mixed up throughout the manuscript. What is actually reviewed is the reporting of the use of variable selection methods. The actual use of those methods is not reviewed. This is most striking at the first sentence of the Results section: only 15% of the screened articles ‘employed’ variable selection methods. To me, this number is absurdly low. In any epidemiological study variables are selected, either in the design or in the analysis of the study. The fact that one decides not to measure a certain variable, can already be considered variable selection (category E, using the categorization of the authors). What this reviews shows, however, is that selection of variables is reported in only 15% of the reviewed manuscripts, which is still very low and worth noticing.

2. It would be of interest to make a distinction between etiologic / observational intervention research on the one hand and prediction research (diagnostic or prognostic research) on the other. In the former, variables are selected to adjust for confounding, whereas in the latter the ‘best’ predictors of the outcome are selected. Hence, I would expect different approaches to select variables.

3. The authors indicate shortcomings of the different methods to select variables, but do not clearly indicate what would be the preferred method (if there is any).

4. What is the distinction between category A / B and D. Is any paper in D also part of either A or B, or did the papers report both bivariate and multivariate analyses and selected variables on either one of the two? Please clarify. Furthermore, the e.g. (at category B) in the abstract is helpful. Consider including such small examples in the Methods section as well (the latter is a discretionary remark).

5. Have the authors considered to include international top-epidemiology journals as well (e.g., Am J Epidemiology), to allow for a comparison?

Minor essential revisions
1. Some of the percentages in the running text are a bit unclear (what is the denominator?). For example, in the abstract 150 (80%), and 97 (44%) are reported, but I don’t see how these percentages are calculated. Please clarify.

Discretionary revisions

1. I’m not familiar with the two Chinese journals that are reviewed. Are these in English language, or Chinese?

2. Last sentence of the abstract: “…. to safeguard the scientific validity of Chinese epidemiological research”. If any, I think this should be “… to safeguard the scientific validity of papers published in Chinese epidemiology journals”.

3. Methods. Articles were selected using an electronic search algorithm. Additionally, all other papers were hand searched. So, all papers were actually screened? How many were initially retrieved by the electronic algorithm?

**Level of interest:** An article of limited interest

**Quality of written English:** Acceptable

**Statistical review:** No, the manuscript does not need to be seen by a statistician.

**Declaration of competing interests:**

I declare that I have no competing interests