Reviewer's report

Title: Simplified Symptom Pattern Method for Verbal Autopsy Analysis: Multi-Site Validation Study Using Clinical Diagnostic Gold Standards

Version: 1 Date: 10 May 2011

Reviewer: D. R. Mani

Reviewer's report:

The manuscript "Simplified Symptom Pattern Method for Verbal Autopsy Analysis: Multi-Site Validation Study Using Clinical Diagnostic Gold Standards" by Christopher JL Murray, Spencer L James, Jeanette K Birnbaum, Michael Freeman, Rafael Lozano and Alan D Lopez proposes a simplified and more effective implementation of the (previously published) Symptom Pattern method for computer coded verbal autopsy. The new Simplified Symptom Pattern (SSP) approach is rigorously validated using accepted train/test methodology based on gold standard VA data.

Major Compulsory Revisions

1. In its current incarnation, the manuscript is excessively dependent on companion papers that are all referenced as being "In submission". Unless many of these are co-published as part of a special issue, it would greatly benefit the reader to replicate relevant aspects from these papers. Some key aspects that fall under this category include the design of the datasets, the tariff calculation, and chance-corrected concordance. Having these definitions and details available in this manuscript will make this document more self contained, and enable interested readers to replicate the methodology laid out here.

2. There are several questions and comments related to the Section titled "Validation Using the PHMRC Gold Standard Train-Test Datasets".
   i. Paragraph 2 in this section states that 100 data subsets were analyzed. The next paragraph claims to assess performance on 500 test datasets. This is confusing to the reader and should be explained in more detail
   ii. Paragraph 2, sentence 2. Shouldn't the last two words be "test dataset"?
   iii. The authors also state that a set of 500 train/test splits was developed. Are these a FIXED set of splits? Or, are the splits created randomly as and when needed? If a fixed set of splits was created, an explanation of the rationale would be very helpful.
   iv. 12 SP variants are evaluated using the train/test datasets, and the best combination is chosen. It is possible that the performance of the best variant is biased by the set of train/test splits used, especially if a fixed set of splits was used for all the variants. In such a case, the choice of the variant can be based on the train/test splits, but the final performance should be evaluated using an unseen validation dataset(s) that was set aside exclusively for this purpose. At
the very least, random subsets of the 500 train/test splits should be used for each variant.

3. What are the implications for deployment of the SSP model? Other than computational cost, would the model be applicable to all regions that the PHMRC data originated from? Would tuning/retraining be necessary? Will the code and/or model be publicly available?

Minor Compulsory Revisions

1. Byass et al (2003) should be included in the list of references.

**Level of interest:** An article of importance in its field

**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.