Reviewer’s report

Title: Robust Metrics for Assessing the Performance of Different Verbal Autopsy Cause Assignment Methods in Validation Studies

Version: 1 Date: 22 May 2011

Reviewer: Heiko Becher

Reviewer’s report:

• Major Compulsory Revisions

I find this paper useful and thoughtfully written. However, I feel that too much of the interpretation is based on the choice of the two scenarios in table 1. I cannot really judge, and I think that goes beyond the task of a reviewer, how far the results in table 2 and in the figures can be generalized. In this example, \( P(B \text{ estimated} \mid C \text{ true} ) \) is considerably larger than \( P(C \text{ estimated} \mid C \text{ true} ) \). If this happens to a rare cause C, it may be uncritical. However, if C is very common, it is a bigger problem. Since the probabilities are taken from a Dirichlet distribution, both situations happen. I would suspect that the result would be less dramatic, if the value of \( P(B \text{ estimated} \mid C \text{ true} ) \) and \( P(C \text{ estimated} \mid C \text{ true} ) \) are interchanged. The authors do not necessarily have to run another simulation, but they could discuss this issue in some more detail.

• Minor Essential Revisions

P.3: Reference 10 is not yet available. I would prefer not to cite papers which are not yet submitted.

P.4: King and Lu is reference 36, not 38

P.9: 2nd equation: C has not been defined

Level of interest: An article of importance in its field

Quality of written English: Acceptable

Statistical review: Yes, and I have assessed the statistics in my report.