Reviewer's report

Title: Improved method for physician coded verbal autopsy reduces rate of discrepancy: Experience of Nouna Health and Demographic Surveillance System (NHDSS), Burkina Faso.

Version: 8 Date: 18 May 2011

Reviewer: Bruce Neal

Reviewer's report:

1. Abstract is very long

2. The conclusions are not well justified. If you are trying to identify a pattern of mortality it doesn't matter if you assign a cause to every death. Simply that those you assign a cause to give you the right pattern. Which the authors have demonstrated very nicely is the case with their first method. Specifically, while that first method fails to get agreement for a large proportion it still gives an almost identical pattern of mortality for the population as the second method for which agreement is much better. So while agreement has been improved the final data of interest have not.

3. Another point that is not really addressed is that better agreement doesn't mean a more reliable result. Having 98.5% agreement is almost implausibly high and almost certainly 98.5% of causes are not correctly assigned. Indeed, all the data that we saw at the recent Bali conference suggest that the reviewers are doing far far worse than this. This does not seem to be appreciated by the authors.

4. So, is this a better method. I don't think so. Sure you get more agreement but what does that mean? More agreement on some correct diagnoses, more agreement on some incorrect diagnoses, and no real impact on the final estimated mortality pattern. With greater cost and longer timelines attendant upon the additional review step. I think the authors need to be much more cautious about their conclusions.

5. It would also be helpful to know how many causes the physicians could code to. Was it a restricted list? Or could they assign any ICD code at any level? And if the latter how did they decide what agreement was? Same broad diagnosis or exactly the same code to every decimal point?

6. The authors state in the discussion that the narrative contains important information. While that might seem intuitive, I am not sure there is evidence to support that. And indeed, we again have evidence to the contrary. Use of a good structured questionnaire may well be as effective as adding in narrative data.

Level of interest: An article whose findings are important to those with closely related research interests
Quality of written English: Needs some language corrections before being published

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