Reviewer's report

Title: Classifying perinatal mortality using verbal autopsy: Is there a role for non-physicians?

Version: 2 Date: 27 April 2011

Reviewer: Timothy Colbourn

Reviewer's report:

Major Compulsory Revisions

1. The calculation of Relative Difference (RD) is based on the rounded figures of CSMF for PC and CC. Using unrounded figures alters the reported RD sometimes above the 20% threshold (maternal accidents goes to 29%) or below/level with it (Tetanus goes to 20% and ‘other (SB) goes to 17%). Should redo with unrounded figures (see attached Excel spreadsheet ‘Tables 2 and 4 calcs.xlsx’).

2. Also, regarding the causes with small RD: The reasons these causes are in fairly good agreement (small RD) may be because they are very small CSMFs - if you look at the sensitivity of CC diagnosis for these causes it is actually pretty low (53% for maternal accidents and 50% for cord prolapse). Worth mentioning given that these small causes are less relevant overall.

3. Methods: The exclusion criteria seem overly strict and need to be explained to the reader. Why exclude hospital deaths (are they already covered by VA in all cases?) Why exclude those where a birth attendant was not present - especially if you are not interviewing the birth attendant? Also, 7 days seems too short a period to be worried about recall bias, especially given that the death of a baby is such a major event that it is bound to stick in the mothers mind. Given your sample is relatively small it could have served you better to have a more inclusive inclusion criteria. Not including 145 cases because the mother was not available in the first 7 days lost you 33% (145/437) of your sample.

4. Methods: Training and VA methodology: ‘Train the trainer’: Please expand. It is not clear to me why a ‘train the trainer’ method was required when only 13 physicians and 40 non-physicians were trained in total. Also what is the breakdown of the 53 trained people (and subsequent VA done) by country given that this study was done in 4 countries?

5. Methods: Why not compare Community Coordinator PANELS vs. Physician panels? What happens if you analyse by single Physicians rather than panels as the gold standard? Also, it would be useful to the reader if you present the data on the % agreement between members of each physician panel (inter-rater reliability). This would give an indication of the reliability of individual physician diagnosis and would give the reader a sense of the reliability of the ‘gold standard’. Comparing panels of Physicians with single community coordinators seems unfair too. I realise you are after a cost-effective intervention (the
community co-ordinators) - but it wouldn't cost much more to have panels of coordinators rather than single coordinators as the comparison group.

6. Discussion: strength of study being that it is from a variety of countries: But you have not broke down the results to show how they differ by each of the 4 countries.

7. Discussion: comparison with Chowdhury et al: you need to calculate Kappa so as to compare with their results.

Minor Essential Revisions
8. Abstract: Methods: It is not clear who the Physician panels are. You could be referring to the 13 physicians or a separate panel of experts? From the background it is apparent that a panel of 2-3 physicians (presumably from the 13) are used - need to make this clearer here in the abstract.


10. Methods: calculation of RD: Should mention that this is always on a positive scale i.e. negative differences are reported as positive Relative Differences.

11. Methods: calculation of overall concordance in COD across all causes of SB (57%) and END (47%): Would be good to show this calculation in the Tables.

12. Author contributions: If all 16 authors had significant input to the data analysis and draft writing, how come you all missed the rounding errors for the RD, no-one calculated Kappa statistics to compare with Chowdhury et al; no-one looked at inter-rater reliability of physicians within panels, no-one looked at the differences between countries etc? I'm not convinced that all of the 16 authors participated as stated.

13. Figure 1: split of 518 PD to 229 END and 289 SB: How was this split determined (given that no VA was done) - maybe worth adding a footnote here and also a footnote for the bottom 134 SB and 118 END split saying that that is according to physician diagnosis following VA.

14. Table 1 and 2 footnotes: should change the 'other' within the 'other' category to spell out what these causes were or state that they were unknown. Also, what is the difference between the main category 'Trauma' and the causes 'birth trauma' and 'neonatal accident' which are categorised under 'other'. Should make this clear as currently the categorisation seems inconsistent.

Discretionary Revisions
15. Abstract: Methods: would be better if you briefly described the robustness criteria in the abstract.

16. RD calculations: Is it possible for you to calculate 95%CI for the RD? As a conservative estimate you could simply base it on the RD between the upper 95%CI for CSMF(PC) and the lower 95%CI of CSMF(CC) as the lower limit, and the lower 95%CI for CSMF(PC) and the upper 95%CI of CSMF(CC) as the upper limit.

17. Conclusion: I think the way forward is actually computer programs like
InterVA. They are more accurate, are consistently reliable, valid (based on state of the art expert medical opinion) don't rely on one cause of death and are far quicker and far cheaper and thus more cost-effective. Would be good if you could expand upon the sentence in the discussion on the advantages and disadvantages of computer programs (or even simple algorithms for the SB / END split).

**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