Reviewer's report

Title: Validating physician review and probabilistic modelling (InterVA) approaches to verbal autopsy interpretation using hospital causes of death

Version: 1 Date: 20 April 2011

Reviewer: Robert RM Mswia

Reviewer's report:

Overall Aim of the paper
The paper aims at exploring the performance of interVA compared to physician’s review of VA in interpreting verbal autopsy data collected in Kilifi Epidemiology and Demographic Surveillance site in Kenya. The comparison is made against the cause of death information certified at the Kilifi district hospital, which is used as a gold standard in this case.

Minor Essential Revisions:
1. (Material and Methods – The probabilistic InterVA model section (Page 7)) - How is the sub-list of 35 possible causes of death from the full list of ICD-10 codes. My understanding of the InterVA is that is it not based on ICD classification, rather on the experts opinion on a list of causes of death. An explanation from the authors would have provided an insight.

2. (Table 1: Kappa statistics for agreement of the 3 methods among the 145 adult deaths). Authors could have presented the Kappa statistics with 95% confidence intervals to see if there are any overlaps

3. (Material and Methods – Hospital cause of death-the gold standard (Page 8)) – Use of hospital cause of death may not necessarily be a gold standard, as studies in countries with more accurate and state of the art diagnostic procedures have shown major differences between medically certified deaths compared to pathology results. However, in the absence of such findings from pathology reports, use of hospital records may be the best alternative to gold standard. Authors should acknowledge the possibility of flaws with their KDH data as the “accurate” gold standard for comparison

4. Data management and statistical analysis – 4th paragraph (Page 10): With the new VA standard tools and methodology, I believe it is possible to identify HIV resulting into TB from the physicians review of VAs. The coders can use an ICD-10 combination code (B20.0) for HIV resulting into TB. If there is a category in interVA for TB/HIV, the PR combination code could be used for such categorization

5. It is not clear whether or how the authors have utilized narrative and the chronic conditions sections of the VA questionnaire in setting up their interVA model. This section may provide additional information to help in the assignment of probable cause of death.
6. Other than re-categorization of the causes of death to fit in with interVA model, the authors did not state what other adaptations were made to the data to fit the model. Is the interVA model adaptable to local context? A statement to explain any adaptations would be helpful.

7. I do not see malaria as among the leading causes of death. Is malaria-related mortality no longer a burden in Kenya? Could it be because the authors have only explored adult VA deaths?

8. The authors have concluded that the interVA is their preferred choice for determining causes of death at community level. While I agree with them regarding the need for automating the process for death certification because of the many reasons they have mentioned, I find that it is a bit premature to make the kind of definitive statement or decision because of the following reasons:
   a. It appears that the decision was largely based on the need speedy generation of the information rather than the quality/accuracy of data. Based on the authors’ results, interVA consistently performed less than physician-coded VAs, even for the top most conditions they explored) against their “gold” standard, as evident in their tables 1 and 2). However, in general the low levels of agreement between physicians review and interVA against the gold standard suggests limitations with both methods, it is more so with interVA model
   b. The validation and/or comparison was only done for adult deaths aged 15 and above, the picture for interVA performance for infant and child causes of deaths may be different.
   c. There are other emerging and promising automation methods that have suggested to perform better that either the interVA or the physicians’ review of VA (such as the machine Learning methods developed by IHME, University of Washington), and hence the authors could potential use such approaches to test their applicability and validity in Kilifi prior to making a definite conclusion of adopting interVA as their choice.

**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