Reviewer’s report

Title: Performance of InterVA for Assigning Causes of Death to Verbal Autopsies: Multi-Site Validation Study using Clinical Diagnostic Gold Standards

Version: 1 Date: 30 May 2011

Reviewer: Daniel Chandramohan

Reviewer’s report:

I think this is the first time the accuracy of interVA has been tested against a gold standard. The previous “validation” of interVA was done by assessing the correlation between PCVA and interVA. Given that the accuracy of PCVA itself is debatable assessing the concordance between PCVA and interVA is not the appropriate way to determine the accuracy of interVA.

It is not surprising that the chance corrected concordance rate of interVA was lower compared to PCVA and SPmethod. The expert opinion based algorithms developed previously performed as good as or lower than PCVA. However data driven algorithms particularly artificial neural network performed better than expert opinion based algorithms. Loronzo et al demonstrate well the key reason for the lower performance of interVA by comparing the probabilities of interVA versus SPmethod for selected causes given the symptom acute cough. It is not uncommon that expert opinion based probabilities are often wrong as demonstrated in Figure 6 that the probability of dying from chronic respiratory disease, other acute infection, maternity related death, suicide and drowning is the same given that some had acute cough according to interVA.

The comparison of chance corrected concordance rates of interVA, PCVA and SPmethod is useful to determine which of these methods perform well. However, this is not comparable to the previously reported “validation studies” of interVA. An analysis of the concordance between PCVA and interVA (individual causes of death and CSMF) consistent with the previous studies would be informative to understand whether the poor performance of interVA is due to the fact that the comparator in this study is clinically confirmed goldstandard.

Proponents of interVA would argue that there was only one African site in the PHMRC validation study and the expert opinion based probabilities of interVA are primarily developed and tested against PCVA in datasets from Africa. However the accuracy of any algorithm based VA should not be depended on the cultural and epidemiological context. An analysis of the concordance between PCVA and interVA in the PHMRC dataset from Tanzania could be carried out to examine whether interVA performs as good as PCVA as claimed by previous studies.

The limitations of using clinically confirmed causes of death occurring at hospital setting as a gold standard to assess the validity of VA have been cited by some investigators to justify the concordance between PCVA and a second method of VA interpretation such as interVA. Lozano et al have shown the limitation of validation studies not having a goldstandard. A brief discussion of the limitations
of methodology of validation of VA without a goldstandard should be included in the discussion section.

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

**Quality of written English:** Acceptable

**Statistical review:** Yes, but I do not feel adequately qualified to assess the statistics.