Reviewer's report

Title: How much do the physician review and InterVA model agree in assigning probable causes of death? A Comparative analysis from rural Ethiopia

Version: 1
Date: 18 April 2015
Reviewer: Peter Byass

Reviewer's report:

This paper addresses an important question: how did the results of physician interpretation and automated (InterVA-4) interpretation of cause of death compare, over a case series of 434 deaths.

Major issues for essential revisions

The authors appropriately use kappa statistics as a means of making their comparisons, but they seem to be a bit confused in the interpretation of the results. The implied null hypothesis is that the two approaches to VA interpretation are equivalent, in the absence of any source of "truth". In other words, if there are differences, nothing can be inferred about which approach is "better" or "more accurate". In some instances this clarity is not entirely reflected in the text. For example, in line 81 "the reliability of the diagnosis that can be reached by using these methods" is not something that can be addressed in this type of comparison. The final conclusion "further refinement to the InterVA model" (line 266) may be true, but is not supported by the results presented here, in the same way that a conclusion "retrain Tigran physicians" would not be supported by these findings.

The Kilite Awlaelo site has actually been using both physician and InterVA-4 methods since 2009, as can be seen from several INDEPTH publications in which Kilite participated using InterVA-4 that are not referenced here (Global Health Action 2014; 7:25362 and following papers). Therefore the statement in lines 87-89 is not entirely correct.

Line 116 onwards describes the process by which two independent physicians reviewed each case. It is important to also present kappa statistics for the agreement between the first and second physicians to understand the effect of inter-physician variation.

Line 125 onwards - it is not clear how the fractional likelihoods of different causes produced by the InterVA-4 model were handled, as well as the residual uncertainty associated with cases have a total likelihood of less than 100%. These methods have been clearly described in previous papers using InterVA and need to be stated and followed more carefully here. The InterVA-4 findings should align with the Kilite Awlaelo data presented in the INDEPTH series of papers and this should be checked.
Line 156 - a complex process has been used to reconcile the different sets of causes of death assigned by InterVA-4 and the physicians. What is not made clear is that InterVA-4 classifies causes according to the WHO VA 2012 standard (described in detail in Global Health Action 2013; 6:21518), and so it would be more informative for an international audience to conform to this standard categorisation for both the physician and InterVA-4 results.

Line 172 - the study population here is very small, at least when it comes to comparing sub-groups or rare causes, and this is not well discussed as a limitation of the study. No assessment of the statistical power of the study is presented. The authors also do not refer a recent large international study that made similar comparisons on a much larger dataset (Journal of Global Health 2015; 5:010402).

Minor essential revisions

There is huge inconsistency in the way that kappa is referred to in the paper - Kappa, kappa, K, k, etc. It would be more correct to use the Greek character kappa throughout.

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.

Declaration of competing interests:

I have no financial competing interests - the InterVA models which I have developed are freely available public-domain resources.

I have had some interactions with the Kilite Awlaelo team through the INDEPTH Network and provided generic training on VA methods, but have not worked in any way on the analysis or writing of this manuscript.