Author's response to reviews

Title: Performance of Physician Certification of Verbal Autopsies: Multi-Site Validation Study using Clinical Diagnostic Gold Standards

Authors:

Rafael Lozano (rlozano@uw.edu)
Alan D Lopez (a.lopez@sph.uq.edu.au)
Charles Atkinson (atkinsct@uw.edu)
Mohsen Naghavi (nagham@uw.edu)
Abraham D Flaxman (abie@uw.edu)
Christopher JL Murray (cjlm@uw.edu)

Version: 2 Date: 2 June 2011

Author's response to reviews: see over
Reviewer's report
Title: Performance of Physician Certification of Verbal Autopsies: Multi-Site Validation Study using Clinical Diagnostic Gold Standards
Version: 1 Date: 2 May 2011
Reviewer: Peter Byass

Major compulsory revisions:

1. The major problem here, as with other papers in this supplement, is that the analyses and conclusions depend crucially on the validity of the PHMRC reference dataset and the newly proposed metrics for VA performance (references 34 and 38 in this paper). Since those references are not yet published or well-established, this creates an obvious difficulty for papers based on them - irrespective of their actual validity. Thus it is essential to at least indirectly mention possible factors which may compromise the validity of the so-called gold standard data and metrics, which of course should be more fully explored in refs 34 and 38 respectively.

   We agree with the reviewer that the paper as written would need to incorporate extensive information on the PHMRC study and the metrics of validity unless it is published alongside the two references 34 and 38. We suggest to the editors that the paper should be published simultaneous or after the publication of the two references. If for some reason this cannot occur, we will be happy to modify the paper to incorporate more of the limitations as suggested by the reviewer.

2. On page 9, the statement "PCVA provides less accurate measurement than medical certification" seems somewhat misleading. In fact you have just established CSMF comparisons of 0.675 to 0.780 for adults, 0.682 to 0.683 for children and 0.733 to 0.756 for neonates - to me this feels much more like approximate equivalence. It does, however, raise the much more fundamental issue of the validity of medical certification - maybe this is discussed in ref 40, which I have not seen?

   We agree with the reviewer that the statement as written is misleading. PCVA does markedly worse for adults but nearly equivalently for children and neonates. With regards to the interpretation of differences in CSMF accuracy. A difference between 0.675 and 0.780 is actually quite large. For the adult cause list, this means that the average error in the estimated CSMF for PCVA is more than 0.5% larger for every cause. Another way to interpret this is that one cause could be wrong by 10.5 absolute CSMF percentage points more to yield a CSMF accuracy of 0.675 compared to 0.780. We have modified the discussion to reflect this.

Minor essential revisions:

3. The sentence beginning "The reasons behind the decision..." on page 6 does not make sense.
We have clarified the language to highlight that for PCVA a shorter cause list had to be used for neonates.

4. page 7: "Figure 6 shows..." the odds ratio for what? Please make clear what an OR of 1 represents here.
   This has been clarified to say the odds ratio of assigning the correct cause as a function of the physician reading the VA.

5. Legend to fig 6: the odds ratios of what?
   This has been corrected.
Reviewer's report
Title: Performance of Physician Certification of Verbal Autopsies: Multi-Site Validation Study using Clinical Diagnostic Gold Standards
Version: 1 Date: 13 May 2011
Reviewer: Elisabeth França
Reviewer's report:
The focus of this study is to describe the performance of the physician certification of causes of death on verbal autopsy (PCVA) using an expressive size of dataset (12,542 deaths) collected at six sites in four countries. It presents results of a validation study of the PCVA comparing different procedures: VAs with and without healthcare experience (HCE), physicians with and without local information, and also using two physicians’ readers. So its subject is of great interest in the field of verbal autopsy (VA) due to the wide use of PCVA and the careful methodology used in the study. The points raised by the authors will certainly contribute to better understanding of the role of the PCVA.

1. In the Abstract Methods, please consider including the criteria of the gold standard causes of deaths used in the study because it was mentioned in the Results section (and also in the Background section) without specific definition.

Due to limitations of space, we are unable to include detail on the criteria for gold standard deaths. We have however modified the abstract to refer to the PHMRC study and its more complete documentation.

2. In reference to the gold standard above mentioned, we are informed that the number of adults, child and neonates deaths by cause used in this research and presented in Annex 1 comprises of both levels 1 and 2 gold standard causes of death. Given the importance of this issue which is considered one of the most important contributions of this article, the addition of the number of deaths by levels of diagnostic certainty would be welcomed by readers.

The reviewer has a good point. However, we hope that this paper will be published in Population Health Metrics at the same time as the detailed paper on the PHMRC gold standard VA study. The paper on the PHMRC study includes extensive detail on the number of deaths by diagnostic criteria and cause.

3. In the Methods section, I would appreciate a more specific and detailed description of the cause of death classification list mentioned on page 5, first paragraph of Data analysis. The specific diagnoses can be seen in figures 3-5 and in the additional files (Annex 1, 2 and 3), but as this information has important relevance to the analysis I think it would be better to present them also in the text of the Methods section.

Given that the PHMRC gold standard VA study paper includes all this detail and we hope it will be published at the same time in Population Health Metrics, we believe it could potentially be highly repetitive to have the same table appear multiple times e.g. in this paper, other methods papers and the
PHMRC paper. If for some reason the editors would like us to present this material in this paper, we will be happy to do so.

4. In reference to the list of causes of death cited above and specified in the annexes, I suggest changing the title “neonates causes” to “perinatal causes” and presenting the specific stillbirth’s causes (pre-eclampsia? infections?) because the use of the term “stillbirth” as a neonate cause of death is not adequate. Besides, if the cause of stillbirth was not specified in the study I suggest using the term “unexplained stillbirth” or a similar one.

In the broader literature, some analysts such as Black et al. have recommended that the term perinatal conditions be avoided because this ICD label is often confused with all deaths during the perinatal period (Black R. Cousens S, Johson H, Law J, Bassani D, Jha P, Campell H, Fisher C, Cibulskis R, Eisele T, Mathers C for the Child Health Epidemiology Reference Group of WHO and UNICEF. Global, regional, and national causes of child mortality in 2008: a systematic analysis. Lancet 2010, 375:1969-87). We have followed in this study, the growing trend towards using the term neonatal conditions plus stillbirths. We agree with the reviewer that there is no exactly correct answer here but feel that the most explicit approach is to use the label neonatal causes and stillbirths. In this study, we have not made any attempt to identify specific causes of stillbirths.

6. I was a little confused when comparing Figure 1 and the explanation presented in the item “Organization of Physician review of VAs”. For example, please clarify why 90% of VAs were single-review according to this figure, and in the text it was mentioned that “all VAs were read by one physician” (page 4).

This has been corrected in the text.

6. Also on page 4 it was stated that 50% of VAs were reviewed by a “different physician” using complete information, i.e. VA with HCE. As observed by the authors, “physicians vary markedly in their ability to assign the true cause controlling for cause of death, availability of HCE, and whether the physician is from the site or another location”. Thus, using a different physician when comparing PCVA for VAs with and without HCE could introduce errors due to different observers which can be a confounding factor when comparing the different procedures. This could lead to a misunderstanding of the process of the physician review of VAs presented, and the authors should clarify these issues.

As we have noted physicians vary by their ability to interpret VAs. However, we believe the comparison with and without HCE is valid as physicians have been assigned at random. Further the logistic regression analysis includes random effects for physician so the coefficient on HCE in this case is controlling for physician which further strengthens the comparison. We have clarified this in the text.

Discretionary Revisions
1. In the Abstract, please consider changing or specifying the term “local prior information” due to the difficulty in understanding it in this section.

We have dropped the word prior and believe the wording local information will be easier to interpret.

2. I think it would be useful for readers who would not read the specific reference mentioned in the text an explanation about the meaning of the term “priority causes” (page 4, first paragraph of Methods).

We have clarified this language. Priority cause only means the specific causes included in the study. The residual categories include deaths from all other causes in the ICD chapter.

3. Maybe the sub-title “PHMRC Study Overview” (p. 4) could be changed to “Definition of the gold standard cause of death” or similar.

We have made the suggested changes.

4. The definition of the variable HCE has been found in different parts of the Methods section (pages 4 and 5) and also in the Results (p. 6, first paragraph), and should be reviewed and standardized. Does the term “medical information” mentioned in Figure 1 have the same meaning as HCE?

This has been standardized in the text

5. Page 7, fourth paragraph: the table mentioned should be table 3.

This has been corrected in the text