Author's response to reviews

Title: Performance of four computer-coded verbal autopsy methods for cause of death assignment compared with physician coding on 24,000 deaths in low and middle-income countries

Authors:

Nikita Desai (nikita.desai@mail.utoronto.ca)
Lukasz Aleksandrowicz (aleksandrowl@smh.ca)
Pierre Miasnikof (MiasnikofP@smh.ca)
Ying Lu (ying.lu@nyu.edu)
Jordana Leitao (leitaojordana@googlemail.com)
Peter Byass (peter.byass@epiph.umu.se)
Stephen Tollman (Stephen.Tollman@wits.ac.za)
Paul Mee (paul.mee@wits.ac.za)
Dewan Alam (dsalam@icddrb.org)
Suresh K Rathi (RathiS@smh.ca)
Abhishek Singh (abhishek@iips.net)
Rajesh Kumar (dr.rajeshkumar@gmail.com)
Faujdar Ram (fram@iips.net)
Prabhat Jha (prabhat.jha@utoronto.ca)

Version: 2 Date: 2 October 2013

Author's response to reviews: see over
REVIEWER 1  
Major revisions  
1. Because the physician PCVA from the study site was used instead of redoing this diagnosis for the analysis more detail on this is needed. The sites vary in how much data is available to the physicians and how they use the data. Most sites have one underlying cause of death but I believe at least 1 ie Agincourt allows fractional causes when there is not agreement among physicians. I think additions to table 1 or a separate table should include how the original cause of death was classified eg 1 or more causes and how were these used in the comparisons in this paper.  
R1. We agree. We have added the details of physician coding to Table 1. All deaths in our study datasets, including Agincourt had one single physician-assigned underlying cause of death in the form of a 3-digit ICD code, which was used as the physician standard. The questionnaires also varied but were reasonably consistent in terms of the major symptoms and signs collected. We have added this point to the methods.

Overall, we noted that at the individual level, all CCVA methods performed similarly across the various datasets. However, at the population level, the King-Lu method performed the best. The differences in the field and coding methods across studies would, a priori, reduce the observed differences between PCVA and CCVA in our comparisons (more likely so at the individual than at the population level). This might have contributed to the observed similarity of the results for the four CCVA methods. We have added this point to the discussion.

2. They claim in the discussion that the PCVA was high quality but they should present data to show this eg the percentage of unknown causes and garbage codes by study.  
R2. Agreed. We have added these details to Table 1. Generally, there were low ill-defined proportions at the age groups studied.

3. The discussion mentions the limitations of using physician coding as the standard but should add more on how limiting this is in regard to accuracy of PCVA in comparison to gold standard known medical causes of death. Ultimately the latter must be used for comparison with CCVA to see if it is better or worse than PCVA.  
R3. We have added a brief point in the introduction on the PCVA as the main comparison. We request that the editors make available the review paper by Leitao et al that shows that PCVA has reasonably high specificity but variable sensitivity for most of the common causes of deaths when compared with hospital-assigned deaths. We have also made the point that comparisons to hospital records is not a “gold standard” as hospital deaths differ from home deaths in symptom patterns, family recall of events and other variables.  
This is further discussed in the discussion. We point out the inherent problems with the “gold standard”. Specifically, there is an extensive literature on major discrepancies between post-mortem vs. hospital causes of death in high-income countries. Nonetheless hospital-based COD has been very useful at the population level to track changes in major diseases (such as declines in breast cancer mortality in recent years). All VA methods, however good they may be, are also going to be subject to the similar difficulties-with variable agreement with hospital-based diagnosis at the individual level, but much better usefulness at the population level. The main question in this paper is to compare how closely CCVA replicates PCVA at the individual and population level.

REVIEWER 2  
2. In the background it would be useful for the paper to give an indication of how accurate PCVA is in identifying the true underlying cause of death. Clearly there are methodological issues around validation of
PCVA against the medical underlying causes but without some understanding of this it seems pointless to compare CCVA against PCVA.

R4. We agree. See Reply R3.

3. Three references in this paper are manuscripts which have been submitted for publication and are presumably under review and thus not available for scrutiny. This is not acceptable. The authors should either hold back publication until after those papers are reviewed or provide adequate information to explain the point with the supporting evidence in this paper.

R5. The reviewer was not likely aware that these four papers were part of a series. As requested by the editor, we have changed these to unpublished observations. If required, we would request the additional papers be shared.

2. Are the methods appropriate and well described and are sufficient details provided to replicate the work? There are limitations to the method, in particular, the lack of a gold standard which needs to be discussed more fully. Also, see specific comments below.

R6. See reply R3, where we argue clearly for the need to avoid “gold standard” approaches.

3. Are the data sound and well controlled? Difficult to assess until specific comments are addressed.

R7. We believe that the re-write substantially addresses all the points raised by both reviewers.

5. Are the discussion and conclusions well balanced and adequately supported by the data? No, see specific comments below

R8. No reply.

Specific comments (Major compulsory revision)

Methods
1. CCVA methods - it is not clear why the open-source results in Figures 1 and 2 are the same as the PHMRC without HCE results rather than falling somewhere between the PHMRC with and without HCE. It would be useful if the authors could explain or comment.

R9. The open-source method is tested on a sub-sample open-source data of about 1500 deaths from the full dataset of 12,000 deaths. The details of the sub-set are quite opaque, and our understanding is that the sub-set contains some, but not all, of the health care experience (HCE) variables (our writing to the Dr. Flaxman did not yield any clarity on the sub-set as he did not agree to disclose which of the HCE variables were included in the sub-set). Thus, for Figure 1 and 2, we have shown the results of the sub-sample using our open source ORF and OTM methods with the published results on the larger dataset with and without HCEs. A footnote to the figure now explains these points more clearly. Despite the fact that our ORF and OTM are not identical to the IHME methods (which are still not open-source), the agreement with the published IHME results is quite reasonable.

2. The characteristics of the five VA datasets used in the study are presented in Table 1. It would be important to describe the similarities and differences between these datasets in terms of the data collection (were the same VA questionnaires used, who collected the data etc) as well as the physician coding of the data (quality control measures and procedures followed when physicians disagreed on the cause of death etc).

R10. See R1. The questionnaires varied but were reasonably consistent in terms of the major symptoms and signs they collect.
3. It is not clear whether in the case of the IHME data, the CCVA were compared with the true underlying cause as determined by medical records and diagnostic tests as done in the Flaxman paper or whether it has been compared with the PCVA as done for the other data sets. If compared with the underlying cause determined by medical records and diagnostic tests, it would not be comparable to other approaches.

R11. The full IHME dataset used physician judgment which relied on medical records/diagnostic tests, as does our IHME dataset subset (This was confirmed in correspondence with Dr. Flaxman). We disagree somewhat with the reviewer in that even with differences in methods to determine underlying cause of death (COD); the overall results are quite similar across various methods. The impact of these differences in COD is now added to the discussion- See also R1.

Results

1. In Table 4 the results of the ORF and OTM are reported as being similar to the IHME tests, however, it is not clear whether the IHME data sample includes or excludes the health care experience (HCE) referred to in the paper by Flaxman et al. and quite different results are reported when HCE is included (37.7 vs 48.0) for adults. Refer to point 2 above.

R12. As noted above, it is our understanding that the sub-set contains some of the HCE variables. Figure 1 shows the range of the IHME results. For clarity, we have removed the statement about similarity of our ORF and OTM methods from Table 3 and 4 and instead emphasized these in the Figure 1 and 2, which show the published results with/without HCE.

Discussion

1. In the discussion section mention is made about the high levels of quality control of the physician coding in the datasets used. This should be described in more detail in the methods. Importantly, the discussion should give a clear indication about the validity of the PCVA in terms of the true underlying causes of death.

R13. See R2. We have expanded the discussion to make this point.

2. The statements made in the last two sentences of para 2, page 10 need to be explained further as it is not clear what is meant by these statements. In addition, the references are not published and are manuscripts which have been submitted for publication. It is thus not possible to assess the evidence for these statements.

R14. We have re-written for clarity and changed the references to unpublished observations (see R5). The key point is that generating usable COD statistics requires a true random sample of deaths, and that a true random sample is the more important determinant than choice of field instrument, choice of coding method (CCVA or PCVA), even though PCVA is preferred.

Conclusion

While the possibility of developing an approach that combines the strengths of CCVA and OCVA is suggested, it is not obvious how this would be done. According to the STROBE guideline, it is advised to give a cautious overall interpretation of results considering the objectives, limitations, multiplicity of analyses, results from similar studies, and other relevant evidence.

R15. We have re-written to be more cautious. CCVA might well be used to define a “short list” which then is coded by PCVA. Specific analyses in larger datasets could be to various age-and sex-specific CODS (eg neonatal deaths, 1-59 month deaths, maternal age group deaths, adult male deaths, injuries, etc). The key is that far more experimentation with larger data sets, open source tools is needed to encourage innovation. We hope that this paper will spur such collaborations by demonstrating that cooperation across various countries and use of methods does yield helpful insights.