Author's response to reviews

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

Authors:

Evasius Bauni (ebauni@kilifi.kemri-wellcome.org)
Carolyne Ndiila (cndila@gmail.com)
George Mochamah (gmochamah@kilifi.kemri-wellcome.org)
Gideon Nyutu (gnyutu@kilifi.kemri-wellcome.org)
Lena Matata (lmatata@ke.cdc.gov)
Charles Ondieki (ondiekiotwori@yahoo.com)
Barbara Mambo (tuongee@gmail.com)
Maureen Mutinda (syongwa@gmail.com)
Benjamin Tsofa (btsofa@kilifi.kemri-wellcome.org)
Eric Maitha (maithabe@yahoo.com)
Anthony Etyang (aetyang@kilifi.kemri-wellcome.org)
Tom Williams (tom.n.williams@gmail.com)

Version: 2 Date: 6 May 2011

Author's response to reviews: see over
Dear Dr Muller,

Re: Our submission “Validating physician review and probabilistic modelling (InterVA) approaches to verbal autopsy interpretation using hospital causes of adult death.”

Many thanks for allowing us the opportunity to present a revised version of our manuscript in response to the reviewers’ comments. The reviewers made a number of important comments and suggestions. We hope we have been able to address them all to your satisfaction in this re-submission, which, as a result, we believe is a stronger and more balanced contribution.

We respond as follows:

Reviewer 1

Reviewer comment:
(1) “The authors do not explain why there were such high losses to follow up; i.e only 145 out of 438 potential cases were included in the study. This could have seriously implications for the generalisability of the study findings”.

Response:
Our main aim in this paper was to compare two methods (PR/InterVA) that are commonly used to interpret VA data with cause of death recorded in hospital by a physician - our “gold standard”. The study, therefore, only included those deaths that occurred in the hospital (and were coded by a physician). Deaths not meeting these criteria were dropped from the analysis. The fact that only 145 of 438 deaths entered our analysis does not, therefore, reflect loss to follow-up but rather failure to meet the inclusion criteria. We do not anticipate that this will introduce any particular bias or adverse implications to the study findings. We have clarified this point in the manuscript.
Reviewer comment:
(2) “The sample size and limited categories of causes of death are not sufficient to support the conclusions regarding utility of the InterVA method. There is no information on several important causes of adult deaths; such as site specific cancers, chronic obstructive pulmonary disease, major subdivisions of cardiovascular disease - e.g. ischaemic heart disease/stroke; or other conditions such as liver cirrhosis; renal failure; or pneumonia”.

Response:
Our study would certainly have been more definitive with a larger sample size and acknowledged this as a limitation of our study in the manuscript. Nevertheless, the overall picture of CSMF for the major causes of death in our study population was similarly determined by both methods. Although there are other important causes of adult deaths, our study had 2 cases of cancer (1 cancer of the cervix and 1 leukaemia), 1 case of chronic obstructive pulmonary disease (asthma), 1 case of ischaemic heart disease/stroke (stroke cases were due to other underlying causes such hypertension), 1 case of liver cirrhosis (alcoholic liver disease), 1 case of renal failure and 2 cases of pneumonia. These frequencies were so low that a massive study would be required to meaningfully investigate the performance of the different models for these conditions or sub-divisions thereof. While we acknowledge that this would be a valuable aim of future and larger studies this was not possible under the power and budgetary constraints implicit in this pilot single site study. We have not made any specific changes to the text in this regard but would be happy to expand the discussion to include this if you consider it important to do so.

Reviewer comment:
(2) “...the methods used do not produce directly comparable data; since the HCOD and physician review yield underlying causes of death (presumably based on the ICD principles, although not stated as such in the article); while the InterVA yields only the likelihood of one or more causes based on probability, without taking into consideration any causal relationship (or absence thereof) between the multiple causes of death”.

Response:
We apologise for lack of clarity regarding this point. The PR and HCOD are based on the ICD 10 guidelines, which we have made more clear in our revised manuscript.

In this study InterVA model assigned most of the cases 118 (81%) to a single cause and 20 (14%) cases to two causes, of which 8 cases agreed with either HCOD or PR making the effect of causal relationship minimal.
Reviewer comment:
(3) “The inferences derived from the apparent similarity in cause-specific fractions observed from the different methods are misleading. These similarities seem fortuitous; when you take the misclassification patterns into account. Additional files 3 and 4 indicate that physician review demonstrates higher validity and markedly less misclassification than InterVA, at the individual level.”

Response
We agree that the physician review performed better than the other methods at an individual level; however, as the reviewer will appreciate, the purpose of VA is to describe the underlying cause of death structure at the population rather than at the individual level. It was more important to us that the model and the physicians arrived at broad agreement in identifying cause of death groups with the greatest public health importance at the population level. We have not made any specific changes in response to this comment but would be happy to do so if considered editorially important.

Reviewer comment:
(4) “The kappa scores also suggest much higher accuracy for the physician review (0.52) as compared to the InterVA (0.32)”.

Response
We agree with this statement and believe that we have presented the data in a very open and honest way in the submitted paper. We have not altered the manuscript in response to this comment; however, we will happily be guided in this by your editorial decision.

Reviewer comment:
(5) “In Table 2, the physician review also demonstrate higher sensitivities within narrow 95% CI; as compared to the InterVA model.”

Response
We agree that the physician review performed better than the model and we have accommodated this by editing the concluding statement. We hope this addresses the reviewer’s concern.

Reviewer comment:
(6) “The manuscript does not give compelling evidence on the 'high' costs for physician review; to justify that speed and costs are sufficient reasons to accept a compromise on accuracy in cause-attribution at the individual level; and thereby adopt the InterVA method and reject physician review”

Response:
Physician review has been shown to be a reliable tool for VA interpretation but is subject to standardization, change of expertise and results in considerable work in that the data have to be coded by two different physicians. On the other hand, the InterVA model permits automation of the coding process and hence has the
advantage of speed and cost as it is freely available. Our favourable view of the InterVA model followed our finding of no differences between physician interpretation and the InterVA model which might have led to substantially different public health policy conclusions at the population level. We acknowledge that our phraseology in favour of adopting the InterVA method over physician review may have overstated the case and we have therefore altered the sentence to read: “We hope that our study, albeit small, provides new and useful data that will stimulate further definitive work on methods of interpreting VA data”.

Reviewer comment:
(7) “A study with a larger sample, and which yields directly comparable data would be required to provide robust evidence to support the InterVA method”

Response:
As above (comment 2) we agree that our study would have been more definitive given a larger sample size; however, we hope that our study, albeit small, provides new and useful data that will stimulate further definitive work of the type suggested by the reviewer. Like all scientific innovations, interpreting VA data with algorithms requires large samples and rigorous testing with data sets from various communities. In the 2010 INDEPTH meeting in Ghana, member sites with VA data agreed to pilot the InterVA model and we hope the outcome of this exercise will provide further insights into the validity and robustness of the method. Although we have not made any specific changes in the light of this comment, changes made in response to other comments (above) have improved the overall balance of the paper.

Reviewer 2

Reviewer comment:
(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.

Response:
The InterVA model was developed using an expert panel and was deliberately designed to be generic and not context dependent, and to produce relatively broad cause of death categories. As a result we re-categorised our data to compare with the sub-list of the 35 causes of death that are included in the InterVA model. Details are provided in additional material file 2. We have added this clarification in our revised manuscript.
(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.”

Response:
We have added the Kappa statistics 95% confidence intervals on Table 1 as advised. Our interpretation has not changed.

Reviewer comment
(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.”

Response:
We agree that post-mortem examination is the most accurate way to determine cause of death; however, such data are unavailable at the Kilifi site. Although we agree that the Kilifi District Hospital data will not be 100% accurate, they are the only data we have to validate PR and InterVA. We have acknowledged this as a limitation of our study in the revised version of our manuscript and hope that this change addresses the reviewer’s concern.

Reviewer comment:
(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”.

Response:
We agree with the reviewer that the coders can distinguish TB from HIV using ICD 10 code B20.0 and this is the case with our current study. In contrast, the InterVA assign TB and HIV as separate entities, which makes direct comparisons difficult in situations where TB and HIV occur together. TB cases reported in this current study are cases that the physicians diagnosed as TB only.
Reviewer comment:
(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.”

Response:
Open-ended sections of the VA questionnaires in which verbatim descriptions of key events and symptoms leading up to an individual's death were not entered into InterVA. While it would be possible to use such information in InterVA, previous evidence suggests that little to no additional information is provided in the open-ended sections of VA questionnaires that is not also recorded in response to the closed questions, thus having little effect on resulting cause of death profiles and little return for the considerable effort required to extract relevant information from these free-text sections. The InterVA input contains several variables covering the “chronic conditions” section and so were mapped accordingly. We do not believe that this omission is likely to have any significant effect on the resulting cause of death profiles.

Reviewer comment:
(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.”

Response:
The adaptations made to the data to fit the model were: The same VA data were compiled into an input file for the InterVA model and processed into cause of death data. The model also expects an input of “high” or “low” to reflect the local prevalence of two specific causes which often vary by more than an order of magnitude between settings: HIV and malaria; In this study these were set to “high” and “high” respectively. These statements have been added on the revised version of the paper.

Reviewer comment:
(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?.”

Response:
In general, the rate of paediatric admission to hospital in Kilifi with malaria has declined by more than 90% in the last 10 years. Nevertheless, irrespective of this decline, death from malaria has always been the preserve of children. As for most other malaria-endemic areas in Africa death from malaria in adulthood in Kilifi has always been rare. This is in contrast to many areas of lower endemicity (the Far East etc) where adults bear the brunt of malaria mortality. We have not made any specific changes in this regard but would be happy to do so if considered important.
Reviewer comment:
(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”

Response:
Although InterVA did not score as high as the PR it did identify accurately the five main causes of adult death, the key targets for public health intervention at unprecedented speed. At a community level and for the purpose of informing public health intervention, our decision did not compromise quality/accuracy. In addition, InterVA was readily available and free.

Reviewer comment:
(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”

Response:
The use of InterVA was to be limited to adults in this current study. Several previous studies have applied InterVA model in children and the data seems to be comparable. We are also planning to evaluate the performance of the model on infants and neonates categories separately.

Reviewer comment:
(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.”

Response:
Scientific innovations change with time and our conclusion does not exclude the use of better methods in the future. At the moment, few of the alternative methods are user friendly and freely available in the public domain. We are eager to test their applicability and validity with our data as soon as they become available. We have altered the concluding statement to cater for the reviewer’s concern.
Conclusion
In addition to the changes outlined above, we have made a number of minor changes to the text that we hope improves both flow and balance. The changes are highlighted in the revised document.

We hope you will find the revised version acceptable for publication in Population Health Metrics Journal and look forward to your positive response.

With kind regards,

Yours sincerely

Dr. Evasius Bauni.