Author's response to reviews

Title: Improved method for physician coded verbal autopsy reduces rate of discrepancy: Experience of Nouna Health and Demographic Surveillance System (NHDSS), Burkina Faso.

Authors:

Maurice YE (yemaure@yahoo.fr)
Eric Diboulo (dibouloeric@yahoo.fr)
Louis Niamba (niamba_louis@yahoo.fr)
Ali Sié (sieali@yahoo.fr)
Boubacar Coulibaly (boubacar.crsn@fasonet.bf)
Cheik Bagagnan (cheik.crsn@fasonet.bf)
Jonas Dembelé (sohandemb@yahoo.fr)
Heribert Ramroth (heribert.ramroth@uni-heidelberg.de)

Version: 5 Date: 3 May 2011

Author's response to reviews: see over
03 May 2011

Dr Maurice YE  
MD, MPH  
Nouna Health Research Centre  
Ministry of Health, Burkina Faso  
B.P. 02  
Nouna, Burkina Faso

To

PHM Editorial Team  
Tel: + 1-206-897-2881  
Email: editorial@pophealthmetrics.com

Dear Sir/Madam,

It was a great pleasure for me that the manuscript entitled “Improved method for physician coded verbal autopsy reduces rate of discrepancy: Experience of Nouna Health and Demographic Surveillance system (NHDSS), Burkina Faso”, which was submitted a few days ago, has been reviewed and returned to the authors for revision. With the submission of this cover and resubmission of the manuscript, we are taking in to consideration and respond to, point-by-point, the various comments by the two reviewers. Before we respond to the reviewers comments, I would like to draw to your attention that one additional co-author, who made substantial writing contributions to the manuscript during our most recent round of revisions, was added as last author. The author is Dr Heribert Ramroth, from the University of Heidelberg/Germany.

Another substantial change to note is that the authors decided to improve the title of the paper, even though this suggestion was not made by the reviewers. The new title is “An improved method for physician-coded verbal autopsy reduces the rate of discrepancy: Experience in the Nouna Health and Demographic Surveillance Site (NHDSS), Burkina Faso”.
Finally, the document was restructured by our technical person who corrected the English vocabulary, grammar and style. Most of the changes are marked in color. The previous figures 4 and 5 about leading causes of death and the table as well about group causes of death were replaced by simple figure on cause specific mortality fraction in the attached figures.

The point-by-point response is structured by reviewers in the order of their comments and suggestions:

**Reviewer 1: Rasika Rampatige**

**Remark 1: Terms death certification and coding are used interchangeably:**
Corrective action: Here we are referring to cause of death coding using the physician verbal autopsy method, which we agree is different to death certification. Corrective measure has been taken in the revised paper.

**Remark 2: The term verbal autopsy also needs to be properly introduced and defined**

We agree with the reviewer that the definition is not clearly outlined in the document or much information about verbal autopsy.

Corrective action has been taken in the paper and the background section was restructured.

Verbal autopsy was also more clearly defined as suggested by Reviewer 1 within the introduction section.

**Remark 3: Advantages and shortcomings of the VA methods should be discussed at the introduction and not in the discussion of the study**

These suggestions were also inserted at the introduction section taking into account the advantages and shortcoming of the method.

**Remark 4: Authors should be aware of the bias imposed by using deaths from two separate years for the comparison. The reason for adopting this method needs to be explained.**

The datasets used came from two different years. But, the fact that the coding is done in the same manner with the same physician coders should reduce the bias. Indeed, the idea to apply a revised coding method for the 2010 dataset is well justified by the high rate of discrepancy encountered for data from the years 2003-2009 data coded, which was above 50% by physicians when using the two independent physicians coding. Thus it was only a shift from an old method to a revised method (with the same physicians), with the major change being that a panel of physicians was introduced in the case of primary coder discrepancies, but applied in a comparable sample of VA questionnaire. Due to resource limitations, physicians could not recode the same dataset using the two methods. This also justifies why a T-test mean proportion for two independent samples was chosen. Some bias may have been introduced by this method, as during panel discussion one physician may try to impose his viewpoint to the others.
Reviewer 2: Bruce Neal

Remark 1: The term 'longitudinal death data' is a little difficult to comprehend: deaths collected over a longitudinal follow up. In our context, data are being collected in a regular basis over a time period continuously whenever a death has occurred within the HDSS target area. But for convenience, the term was drop out.

Remark 2: Exactly what is meant by 'discrepancy' is not clear? We define a discrepancy as when both coders disagree on the final diagnosis of cause of death. When referring to Physician Coded Verbal Autopsy (PCVA), the most used coding method in our setting, completed VA questionnaires are independently reviewed by two experienced local physicians who assign a single COD based on ICD-10 coding. Two possibilities can come out: physician can reach a consensus in the underlying cause of death, in this case an agreement, but they can also reach non consensus in diagnosis, which we define as a discrepancy or discordance.

Remark 3: Likewise, the 'discordance rate' is not explained sufficiently for a reader to know what is meant: Discordance rate is “a quantity measured with respect to another measured quantity” it’s the number of time both coders disagree on the final diagnosis measured as a proportion. The rate is obtained by taking the total number of VA coded where there is not agreement among physician coders over the total number of VA coded.

Remark 4: The conclusion that suggests this is a good method for resource poor countries seems a little at odds with what has been done because the panel method actually requires more resource than the standard method. And it is not clear what the real gain from this additional investment is:

What we attempted to point out here is that: The locally-adapted method doesn’t require more resources; it is only more time-consuming as compared to the old method. However it helps achieve greater agreement rate between coders. To figure out this, from 2003 to 2009, the VA coded with local trained physicians has never achieved more than 50% of concordance rate or agreement among physician coders. This means otherwise that around 50% of VA coded was left aside because of discrepancy. When regarding the time spent by physicians on an activity that ends in success only 50% of the time, it is can also be seen as an inefficient use of resources. This is why the new method that used panel of physician in our views brought more gain by reducing substantially the discrepancy and therefore the undetermined cause of death.

Given the fact that more than 75% of death occurred out of formal health facilities, at home, this method should be encouraged to be used in the coding of all deaths.

Remark 5: Background
It might be helpful to specifically state the objectives of this research at the end of the background section. So that the reader knows exactly what it was you set out to do.

Ok, we agree with that, corrective action have been taken to state clearly the objectives of this research at the end of background section.

Remark 6:
What is a 'sub-Sahelian' climate? You might need to explain this if it is important.

A Sub-Sahelian climate is characterized by hot climate with short rainy season lasting from June to September with a rainfall varying between 400 to 1000 mm. The vegetation is mainly dry savannah comprised of scattered short trees. The mean temperature varies from 26 °c to 34°c with very often 40°c in April the hottest period. This was explained in the core document.

Remark 7: One question that arises is why you didn't apply the same methods independently to the same set of deaths. That would have decreased the risk of bias consequent upon other factors that might have varied form one year to the next.

The only thing that substantially varied between these two years is the coding method, the others components of the VA procedures remained unchanged, including the coders. Furthermore, the datasets used were different as coming from two different years. But, the fact that the coding is done in a regular way with the same physicians coders should reduce the bias. Indeed, the idea to apply a second method on 2010 dataset is well justify by the high rate of discrepancy so far encountered since 2003 to 2009 data above 50% by physician when using the two independents physicians coding. Thus it was only a shift from a old method to another method with the same physician with the difference that a panel physician was introduce, but apply in a comparable sample of VA questionnaire. For resource limitation, physicians could not recode the same dataset using the two methods. This also justify why a T-test proportion comparison for two independents samples was chosen. Some bias may have been introduced by this method, as during panel discussions one physician may try to impose his viewpoint on other participating physicians.

Remark 8: The statistics section is a bit short on detail. I was unable to understand from what is written here exactly what you did by way of quantitative comparisons. Or how your metrics were calculated. Ok more details were provided in the document.

Remark 9: It’s not clear what 'agreement' actually means. At what level of detail of diagnosis was there agreement? How did you calculate the 'concordance rate'? Indeed, how do you calculate concordance when you have a panel of many members? What is the concordance between?

There is an agreement when 2 coders make the same final diagnosis for a given VA. Concordance rate which is here the same as “agreement rate” is the number of time when at least 2 independent coders make the same final diagnosis. We say that there is an agreement when a panel discussion leads to a final cause of death between the coders, and the death is not coded as “unclassifiable”.

---

*Téléphone : (226) 20 53 70 55/43, Fax : (226) 20 53 70 55*
In other words, agreement considered here as proportion can be calculated by computing the number of total VA coded where a consensus was reached other the total VA coded by physicians.

Remark 10: I was unable to locate table 2 (or table 1, although I could see no reference to that either):
This was a mistake on behalf of the authors. In fact in a previous version of the manuscript there were two tables, but in the version submitted we only presented one, although mention of Table 2 remained in the text. This has been corrected.

Remark 11: I wanted to see a direct side by side visual comparison of the cause specific mortality fractions by method. Perhaps a plot so that I could get an immediate impression of how the two methods compared.
A cause specific mortality fraction graph was provided (Figure 4) showing the mortality patterns using the two methods.

Remark 12: I was not convinced that I had seen results to support the first statement made in your Discussion.

Results have been oriented to more focus our research objective so that to cover the statement where introducing an improved method of VA coding could increase the agreement among physicians coders.

Remark 13
I wanted to hear a bit about the strengths and weaknesses of the research approach you had used. Are you likely to have precise and reliable results? If so why? If there are some weaknesses to the design what are they and what might the impact be?

We are likely to provide precise and reliable results given that the only aspect that substantially changes between the 2 approaches is the method. One of the weaknesses of the new method is that it introduces an extra step in the coding process and thus results in a longer and more time consuming process. It is also well documented that physician coded verbal autopsy has some limitations due to recall bias, and that local physicians’ knowledge on the location’s epidemiological profile can influence the diagnosis. Despite its limitations, verbal autopsy provides useful information on the importance of causes of death, and also enables the evaluation of the impact of public health intervention in terms of mortality due to preventable diseases like (poliomyelitis, measles…). This suggestion have been taking into account in the revised paper

Remark 14
Conclusions
I was not convinced by the data that you presented that this was actually a method that was going to be a good choice for widespread adoption in resource poor settings. I guess I could be perceived to be biased in my perception by my own work with Joshi that has suggested two reviewers do little better than one. Which seems to me to suggest that a panel is likely to offer rather little real advantage over two. And in the absence of convincing stats in this report of yours I see little reason to change this perception of mine. I think you need to try and express your reasons for this conclusion much more
clearly because I don't think you make an especially convincing argument at the moment.

Although the previous study done highlights some gain with a simple coder, I think we may consider the specific country context when applying the VA methods that really varies greatly from site to site. As pointed out by Frank Baidel and al, (year?) in the paper “Setting international standards for verbal autopsy”, “Despite such a widespread use of verbal autopsy, we are unable to assess how consistent and reliable the data are. Because verbal autopsy data sets are not widely shared” indeed, more than 36 HDSS within the INDETPH Network is doing VA to meet the information needs, but there is still a failure to agree on a standard approach for doing VA and the tools used still varies among sites.”

The key result of our study, which is the message we would like to pass on to people who read this manuscript, is that adding one step in the coding process could bring an added value in terms of improving agreement between coders. This is clearly supported by the fact that the discrepancy rate, which for a long time was over 50%, has been reduced to less than 1.5% by the new method. This method should then be encouraged, even if the current trends are to move on computerized methods, which have not yet been validated at on large scale.

Final comments

We tried our best to respond to the reviewers comments and make the necessary changes given the scope of our knowledge. I hope that all inquiries have been sufficiently responded to by my team. We remain available for further needs in clarification. All the authors have contributed to the revision of the manuscript according to the remarks of the reviewers.

Sincerely, yours

Dr Maurice Yé
MD, MPH
Centre de Recherche en Santé de Nouna
PO BOX 02 Nouna
Tel : +226 70244811, +22620537043, e-mail : yemaure@yahoo.fr
While hoping that our manuscript will fulfill your need and interest