Author’s response to reviews

Title: Automated versus Physician Assignment of Cause of Death for Verbal Autopsies: Randomized Trial of 9374 Deaths in 117 Villages in India

Authors:

Prabhat Jha (prabhat.jha@utoronto.ca)
Dinesh Kumar (drdineshkl@charutarhealth.org)
Rajesh Dikshit (dixr24@hotmail.com)
Atul Budukh (atul.budukh@gmail.com)
Rehana Begum (BegumR@smh.ca)
Prabha Sati (SatiP@smh.ca)
Patrycja Kolpak (patrycja.kolpak@mail.utoronto.ca)
Richard Wen (WenR@smh.ca)
Shyamsundar Raithatha (shyamjr@charutarhealth.org)
Utkarsh Shah (drutkarsh2537@gmail.com)
Zehang Li (lizehang@gmail.com)
Lukasz Aleksandrowicz (lk.aleksandrowicz@gmail.com)
Prakash Shah (ShahPr@smh.ca)
Kapila Piyasena (delpepiyasena@gmail.com)
Tyler McCormick (tylermc@uw.edu)
Hellen Gelband (hellengelband@gmail.com)
Samuel Clark (sinafala@gmail.com)

Version: 3 Date: 21 May 2019

Author’s response to reviews:
BMED-D-18-01907R2 Automated versus Physician Assignment of Cause of Death for Verbal Autopsies: Randomized Trial of 9374 Deaths in 117 Villages in India:

REPLY to 3RD round of reviews

C1. I agree with reviewer # 3 that the two arms are not totally identical, but I disagree that they are very different. They are not identical, but they are similar, so that the distributions of causes are likely to be close to each other in the two arms. In other words, major differences (as shown in the analysis) most likely reveal differences in the diagnosis method (automated versus manual diagnosis). To sort out this issue, I propose that the authors add a cautionary word in the discussion.

REPLY 1: Agreed. We have added page 14. … the physician and automated arms were well-balanced in the overall distribution of key predictors of the distribution of COD

C2. About the presentation of the paper, I think that the authors should avoid strong statements such as:

Page 12: “Hence claims of superiority of any one algorithm (13) are scientifically untrustworthy”. Rather show the strong or weak points of each method, as none is ideal.

REPLY 2: We have edited to state: “No one algorithm consistently performed better than the others, with variation in specific diseases. Hence claims of superiority of any one algorithm (13) carry little scientific credibility.”

However, we believe it is important to challenge directly the claims made by Murray et al in an earlier BMC Medicine paper. See our reply in an earlier review round which we have posted again:

Our reply #48 is worth drawing your attention as it provides the key justification for the trial-“claims for algorithm performance (drawn on weaker non randomised data) likely overstate strengths. In the case of SmartVA, the lead author went as far as writing “there would seem to be little scientific, financial or moral justification to continue with [physician coding of verbal autopsies].” (Murray 2014, BMC Ref 11)

This is of course unscientific chest thumping, with little justification. This claim makes it all the more important to establish the deficits in reliability of current algorithms. We believe that algorithms can be improved, but currently they lack validity in a lot of areas. We provide options to improve future studies, including use of randomised designs that avoid the obvious biases and preferences of algorithm producers”
In the context of the above over-the-top claim (published in this same journal!), we believe we have written our paper with suitable caution.

C3. Page 13: “It would be specious to claim physicians as a ‘gold’ standard, as none exists.” The gold standard for cause of death exists: formal autopsy (a recent study in Mozambique compared formal autopsy with verbal autopsy). Rather say “we used physician diagnosis as the reference standard, which has its own limitations”

REPLY 3: We prefer to retain this sentence with only slight changes written: it would be misleading to claim physicians as a ‘gold’ standard, as none exists.(1, 7, 12) Unattended deaths, by definition, cannot be conclusively categorized, and hospital-based deaths cannot adequately reflect home deaths.(1).

We do not agree with the reviewer’s suggestion that formal autopsy with verbal autopsy is an appropriate “gold” standard. As mentioned in the ensuing paragraph (now expanded), hospital-based deaths with a distinct COD profile, education patterns and likely microbiological aetiologies cannot be used to “validate” home deaths which differ clearly not only in these three known variables, but likely several other unmeasured ones (see Jha BMC Med 2014). We do mention that the promise of the “MITS/CaDMIA” data such as that in Mozambique offers some promise of improving the verbal autopsy tools, but this is restricted only to children and, in the case of the Mozambique study, is confined only to hospital deaths.

This is a long debate in the literature on appropriate “gold” standards (much of it published in BMC Medicine), so we prefer not to have an extended debate in the discussion of this paper. Rather our contribution is to argue that randomised designs should be used to evaluate newer algorithms, and that bigger, more representative datasets of home deaths should be developed to improve the training of such algorithms.

C4. Also, a table of agreement by cause for children and neonates (the equivalent of Table 3 for adults, now as Additional file 10) could be put in text and not in the annexes.

REPLY 4: We would respectfully decline this suggestion. The sample sizes of neonatal and child deaths are smaller than for adults, limiting the ability to look at results across algorithms. Moreover, we put table 3 on adult on the insistence of an earlier reviewer, as we are concerned that such a large table may reduce the paper’s readability. Thus, we prefer to avoid additional large tables in the main text.