PEER REVIEW HISTORY

BMJ Open publishes all reviews undertaken for accepted manuscripts. Reviewers are asked to complete a checklist review form (**http://bmjopen.bmj.com/site/about/resources/checklist.pdf**) and are provided with free text boxes to elaborate on their assessment. These free text comments are reproduced below.

ARTICLE DETAILS

TITLE (PROVISIONAL)	Ability of Verbal Autopsy data to detect deaths due to uncontrolled hyperglycaemia; testing existing methods and development and validation of a novel weighted score
AUTHORS	Blackstock, Sarah; Witham, Miles; Wade, Alisha; Crampin, Amelia; Beran, David; Ogle, Graham; Davies, Justine

VERSION 1 – REVIEW

REVIEWER	Srihari Cattamanchi Consultant and Adjunct Professor (Research), Department of Emergency Medicine, Sri Ramachandra Institute of Higher Education and Research, Chennai – 600116. Tamil Nadu. India.
REVIEW RETURNED	14-Oct-2018

GENERAL COMMENTS	Thank you for the chance to review this manuscript which assesses the utility of Verbal Autopsy statements in identify deaths due to uncontrolled hyperglycaemia and the development of a weighted score (WS) to explicitly distinguish deaths due to uncontrolled hyperglycemia. Uncontrolled hyperglycaemia is one of the common cause of non-communicable diseases deaths in LMICs where they are scares resources to diagnose and manage it. In the study, the Weighted Score has good specificity and negative predictive value. The study shows that the InterVA-4 algorithm works poorly in recognising deaths due to uncontrolled hyperglycaemia but enhances detection if diagnosed pre-mortem with diabetes mellitus. The study also observed late diagnosis of uncontrolled hyperglycemia in a large number of deaths. Readers from LMICs where data on the number of deaths due to uncontrolled hyperglycaemia is wanting and where Verbal Autopsy statements are available would be interested in knowing about the InterVA-4 algorithm. The InterVA-4 algorithm identifies the numbers of deaths due to uncontrolled hyperglycaemia, consequently identifying gaps in the health system in early diagnosis and management of uncontrolled hyperglycaemia. The research questions were explicitly defined, the design was appropriate; the methodology appropriately described; main outcome measures were clear; the results were credible and aptly answered the research question. The conclusion focused on the

	aims and well interpreted. The study is reported using the checklist from the TRIPOD guidelines. The discussion is concise. The manuscript is well-written and grammatically correct. There are a few typing mistakes that need corrections as well as the spelling should change to British from American. Uncontrolled hyperglycemia is common and appealing to a wide range of clinicians especially from LMICs; furthermore the high-quality of this research study, support the publication of this paper in the BMJ.

REVIEWER	Abraham Flaxman Institute for Health Metrics and Evaluation University of Washington USA I have led the development of SmartVA-Analyze, a freely available software tool for computer-coding of verbal autopsy interviews.
REVIEW RETURNED	05-Nov-2018

GENERAL COMMENTS	This paper develops a prediction model for identifying diabetes deaths from verbal autopsy interview data. Although the authors reject using the PHRMC validation database on the grounds that it did not include questions about polyuria and polydipsia, I would strongly suggest that they consider the including the predictors identified by this prior work. My work on the Tariff Method found responses on questions about sores and ulcers to be highly informative for identifying a subset of diabetes deaths, for example. The authors' choice to restrict analysis to deaths of age 49 and below strikes me as unnecessary and potentially misleading---I would expect the bulk of diabetes deaths to occur in individuals above age 49. Do the authors anticipate some fundamental difference in signs and symptoms for diabetes deaths in younger patients? I am concerned about this database introducing bias due to change in access to care over time. The data were collected from 1992 to 2016, and I suspect changes in wealth, in BMI, and in other determinants have been large over those 20+ years. It may be misleading to generalize historical model performance to predict performance on future data. A methodological shortcoming that is acknowledged, but seems to call into question the utility of these results, is the lack of clinical diagnostic criteria for identifying diabetes deaths. A review of VA data will not find diabetes deaths where the family members cannot answer key questions accurately. How accurate do you think the interviews are? For example, while the authors of this paper have highlighted polyuria as a key symptom in a diabetes death, the in PHMRC database, "Did [name] stop urinating?" was endorsed for 37% of deaths from Diabetes with Renal Failure. SPCC is likely to have similar accuracy to PCVA for the PHMRC validation data, where we found that doctors could identify diabetes deaths with a chance-corrected concordance of about 50% when answers including health care experience were included, but with less than 20% CCC if this information was not available.

	Expert algorithms for coding VA data often overlook multiple source of noise introduced by the data generation process. A VA is not simply recording the presence or absence of a symptom like polyuria---this is filtered through the family member's ability to observe, remember, and endorse the presence/absence of this symptom. Although this paper augments expert opinion with beta coefficients derived from a modest number of identified diabetes deaths, it is likely to miss complex patterns present in VA data but not in the examination room. Additional specific feedback: I feel you have used too many acronyms in the abstract – WS, SPCC, SiPCC, CACC – these are not familiar to me and seem unnecessary In the abstract, the conclusion strikes me as an over-reach, since WS has PPV of only 60%. That means it will still be a substantial underestimate, so it seems to suffer from the same flaws that lead the authors to recommend against InterVA. I would like a paper like this to convincingly justify the utility of a prediction method that distinguishes diabetes and non-diabetes deaths, only. My working hypothesis is that public health decision makers need to know the other leading causes of death in this population, as well.

VERSION 1 – AUTHOR RESPONSE

Reviewer: 1

"Thank you for the chance to review this manuscript which assesses the utility of Verbal Autopsy statements in identify deaths due to uncontrolled hyperglycaemia and the development of a weighted score (WS) to explicitly distinguish deaths due to uncontrolled hyperglycemia.

Uncontrolled hyperglycaemia is one of the common cause of non-communicable diseases deaths in LMICs where they are scares resources to diagnose and manage it.

In the study, the Weighted Score has good specificity and negative predictive value.

The study shows that the InterVA-4 algorithm works poorly in recognising deaths due to uncontrolled hyperglycaemia but enhances detection if diagnosed pre-mortem with diabetes mellitus. The study also observed late diagnosis of uncontrolled hyperglycemia in a large number of deaths.

Readers from LMICs where data on the number of deaths due to uncontrolled hyperglycaemia is wanting and where Verbal Autopsy statements are available would be interested in knowing about the InterVA-4 algorithm. The InterVA-4 algorithm identifies the numbers of deaths due to uncontrolled hyperglycaemia, consequently identifying gaps in the health system in early diagnosis and management of uncontrolled hyperglycaemia.

The research questions were explicitly defined, the design was appropriate; the methodology appropriately described; main outcome measures were clear; the results were credible and aptly answered the research question. The conclusion focused on the aims and well interpreted. The study is reported using the checklist from the TRIPOD guidelines.

The discussion is concise. The manuscript is well-written and grammatically correct. There are a few typing mistakes that need corrections as well as the spelling should change to British from American.

Uncontrolled hyperglycemia is common and appealing to a wide range of clinicians especially from LMICs; furthermore the high-quality of this research study, support the publication of this paper in the BMJ."

Author response: Thank- you for your careful read and appraisal of our manuscript. We are grateful for your positive comments. We have read through the manuscript and have corrected any typing mistakes and ensured UK rather than US spelling.

Reviewer: 2
Thank-you for your detailed read of our paper.

"This paper develops a prediction model for identifying diabetes deaths from verbal autopsy interview data. Although the authors reject using the PHRMC validation database on the grounds that it did not include questions about polyuria and polydipsia, I would strongly suggest that they consider the including the predictors identified by this prior work. My work on the Tariff Method found responses on questions about sores and ulcers to be highly informative for identifying a subset of diabetes deaths, for example."

Author response: Thank-you for this comment and for sending the additional results from PHRMC through. We would have been delighted to have tested our weighted score using the PHRMC database. However, two of the strongest positive predictive symptoms and the strongest negative predictive symptom from the weighted score were not available in this dataset. Thus, it was not possible to use it as a validation tool.
It could be possible to develop another weighted score to detect diabetes using the data contained in PHRCM however, that would need to be the subject of a separate study. Our aim in this study was to develop an algorithm for use with information collected as part of the widely used WHO VA tool. Given that data on classic symptoms for diabetes are less available in the PHRCM data set than in the VA datasets that we used, we don't feel that developing a separate algorithm using this tool would add anything to our current study.

The authors' choice to restrict analysis to deaths of age 49 and below strikes me as unnecessary and potentially misleading---I would expect the bulk of diabetes deaths to occur in individuals above age 49. Do the authors anticipate some fundamental difference in signs and symptoms for diabetes deaths in younger patients?

Author response: Thank-you for this comment; we agree that deaths from diabetes and its sequalae are more likely in older people where prevalence will be greater. However, our aim was to detect deaths due to uncontrolled hyperglycaemia rather than general diabetes-related deaths. Thus, we a priori limited our age range to detect premature mortality where multiple co-morbidities and their associated symptoms did not confound the results. However, we acknowledge that testing the score in older persons needs to be the subject of future study. We have added this as a limitation as follows: "For this weighted-score development study, we limited the age range of cases to between one and 49 years to ensure that we detected premature mortality and to avoid confounding from competing symptoms that may be seen in older people who likely have multiple co-morbidities. We may have missed cases in older deaths, and how this weighted score performs in older age groups needs to be the subject of separate study".

I am concerned about this database introducing bias due to change in access to care over time. The data were collected from 1992 to 2016, and I suspect changes in wealth, in BMI, and in other determinants have been large over those 20+ years. It may be misleading to generalize historical model performance to predict performance on future data.

Author response: We agree that change in the determinants of diabetes and access to care will have changed over the past 20 years. However, our weighted score used only diagnoses and symptoms of diabetes. The symptoms have not changed in the years over which VA data are available. The reviewer is correct in that access to care has changed, and we have already commented that the main contributor of our score – diagnosis of diabetes – requires access to care, which may not be available in many LMICs. We have commented on access to care for diabetes in terms of prior diagnosis, but given the small numbers of cases, felt that further comment on temporal trends in access would be misleading. Likewise, although determinants of diabetes have increased, given the small numbers of deaths discovered, we have made no comments on temporal trends in numbers of deaths per se.
Regarding access to care, we have commented in the discussion of the manuscript: "Unfortunately, given the small numbers of deaths found in this study, we were not able to reliably look at temporal trends in access to care."

A methodological shortcoming that is acknowledged, but seems to call into question the utility of these results, is the lack of clinical diagnostic criteria for identifying diabetes deaths. A review of VA data will not find diabetes deaths where the family members cannot answer key questions accurately. How accurate do you think the interviews are? For example, while the authors of this paper have highlighted polyuria as a key symptom in a diabetes death, the in PHMRC database, "Did [name] stop urinating?" was endorsed for 37% of deaths from Diabetes with Renal Failure. SPCC is likely to have similar accuracy to PCVA for the PHMRC validation data, where we found that doctors could identify diabetes deaths with a chance-corrected concordance of about 50% when answers including health care experience were included, but with less than 20% CCC if this information was not available.

Author response: We completely agree, and the lack of clinical confirmation of cause of deaths in the VA data is why we wished to use the PHMRC database as validation. Given that was not possible (see above), five clinicians with experience of work in LMIC settings and dealing with diabetes were involved in assessment of the cases, and agreement on diagnosis between these physicians was high. This mitigates to a reasonable extent against the lack of confirmed laboratory diagnoses.
As the reviewer points out, we have also highlighted this as a limitation of the study. But, and in accordance with reviewer one, VA methods are not meant to give definitive diagnoses, but an indication of deaths that are likely due to uncontrolled hyperglycaemia in situations where quality vital statistics recordings are absent. The issue that this reviewer points out is likely to apply to all VA analyses, but until there are more accurate, global, CoD data, VA methods remain a useful tool, and one where the results are also inline with the GBD estimates at IHME (Byass et al, Lancet Global Health 2017).

Expert algorithms for coding VA data often overlook multiple source of noise introduced by the data generation process. A VA is not simply recording the presence or absence of a symptom like polyuria---this is filtered through the family member's ability to observe, remember, and endorse the presence/absence of this symptom. Although this paper augments expert opinion with beta coefficients derived from a modest number of identified diabetes deaths, it is likely to miss complex patterns present in VA data but not in the examination room.

Author response: Agreed and please see comment above. This reviewer comment is not specific to this analysis, but pertains to the use of VA methods more generally. Whilst more accurate methods

are lacking, VA remains a useful tool to determine cause of death and multiple studies reported in multiple publications have used this.
To address this and the above point specifically in the manuscript, we have added the following to the limitations: “Lastly, VA tools to ascertain cause of death are not as accurate as vital statistics reporting which are based on clinical diagnoses. However, such reporting is lacking in many populations, especially in lower and middle income countries. In these situations, VA is proven to be a reliable alternative method way of ascertaining cause of death”.

Additional specific feedback:

I feel you have used too many acronyms in the abstract – WS, SPCC, SiPCC, CACC – these are not familiar to me and seem unnecessary
Author response: We have spelt out these acronyms in full.

In the abstract, the conclusion strikes me as an over-reach, since WS has PPV of only 60%. That means it will still be a substantial underestimate, so it seems to suffer from the same flaws that lead the authors to recommend against InterVA.

Author response: Our abstract conclusion reads as follows – “Our results suggest that widely used VA methodologies may be missing deaths due to uncontrolled hyperglycaemia. our WS may offer improved ability to detect deaths due to uncontrolled hyperglycaemia in large populations studies where no other means exist”. We have tried to be both correct and circumspect in writing this. We have clearly stated the PPV in the abstract results. Our weighted score does perform better that INTERVA-4 in detecting deaths due to uncontrolled hyperglycaemia but we simply state in the abstract conclusion that it “may offer improved ability to detect deaths”. We have not stated that this should be used as an alternative or even that it does offer improved ability over INTERVA-4 (which it does). If the editor can suggest a replacement to “may” in this sentence (might?) we would be glad to use it and will be happy to be guided by the editor if they agree that we should be even more circumspect.

I would like a paper like this to convincingly justify the utility of a prediction method that distinguishes diabetes and non-diabetes deaths, only. My working hypothesis is that public health decision makers need to know the other leading causes of death in this population, as well.

Author response: Whilst we agree that it is important that policy makers know the hierarchy of causes of death, it is widely acknowledged that deaths due to uncontrolled hyperglycaemia are largely unquantified and, importantly, completely avoidable at little healthcare expense. Thus, they are an unknown but potentially important contribution to causes of death. We hope that the use of this weighted score will allow researchers and policy makers to start to quantify deaths due to uncontrolled hyperglycaemia so that they can consider if investment in this condition is worthwhile.

VERSION 2 – REVIEW

REVIEWER	Srihari Cattamanchi Consultant and Adjunct Professor (Research), Department of Emergency Medicine, Sri Ramachandra Institute of Higher Education and Research, Chennai – 600116. Tamil Nadu. India.
REVIEW RETURNED	11-Jul-2019

GENERAL COMMENTS	As I wrote earlier, the paper addresses an important issue, the identification of deaths due to uncontrolled hyperglycemia, which will be of great value in low middle-income countries where they are scares resources to diagnose and manage it. The authors have adequately addressed all the concerns raised both by Reviewer 1 and 2. I thank the authors for addressing a critical topic of importance to physicians from LMICs and addressing it through a high quality research paper and providing a tool.

REVIEWER	Abraham D. Flaxman Institute for Health Metrics and Evaluation Dept of Health Metrics Sciences University of Washington USA
REVIEW RETURNED	26-Jun-2019

GENERAL COMMENTS	I described a number of concerns in my initial review of this manuscript, and little has been done to address them in this revision. The most substantial change has been removing some acronyms, and it is a welcome simplification. The other changes seem to be the addition of three sentences on the limitations of the study. I feel that this is a small improvement but insufficient to address the major concerns I raised in my initial review.