Validating the InterVA Model to Estimate the Burden of Mortality from Verbal Autopsy Data: A Population-Based Cross-Sectional Study

Sebsibe Tadesse*

Institute of Public Health, the University of Gondar, Gondar, Ethiopia

Abstract

Background: In countries with incomplete or no vital registration systems, verbal autopsy data are often reviewed by physicians in order to assign the probable cause of death. But in addition to being time and energy consuming, the method is liable to produce inconsistent results. The aim of this study is to validate the InterVA model for estimating the burden of mortality from verbal autopsy data by using physician review as a reference standard.

Methods and Findings: A population-based cross-sectional study was conducted from March to April, 2012. All adults aged ≥14 years and died between 01 January, 2010 and 15 February, 2012 were included in the study. The verbal autopsy interviews were reviewed by the InterVA model and physicians to estimate cause-specific mortality fractions. Cohen's kappa statistic, sensitivity, specificity, positive predictive value, and negative predictive value were applied to compare the agreement between the InterVA model and the physician review. A total of 408 adult deaths were studied. There was a general similarity and just slight differences between the InterVA model and the physicians in assigning cause-specific mortality. Both approaches showed an overall agreement in 298 (73%) cases [kappa = 0.49, 95% CI: 0.37-0.60]. The observed sensitivities and specificities across causes of death categories varied from 13.3% to 81.9% and 77.7% to 99.5%, respectively.

Conclusions: In understanding the burden of disease and setting health intervention priorities in areas that lack reliable vital registration systems, an accurate analysis of verbal autopsies is essential. Therefore, users should be aware of the suboptimal performance of the InterVA model. Similar validation studies need to be undertaken considering the limitation of the physician review as gold standard since physicians may misinterpret some of the verbal autopsy data and finally reach a wrong conclusion of the cause of death.

Introduction

Developing countries generally lack consistent, timely, and reliable information on the level of cause-specific mortality fractions (CSMFs) in their populations [1]. Vital registration data are incomplete and contain only few physician-certified deaths [2]. Nevertheless, any meaningful health intervention policy and/or program must be informed by the cause of deaths (CODs) that are of the greatest importance locally. Verbal autopsy (VA) is a useful tool in such settings to establish the probable COD by interviewing a close caregiver or anyone who can provide witness to the death event [3].

There have been various attempts at validating physician reviews to interpret VA data [4-7]. However, the methodology is known to have several limitations. For example, physicians

may differ systematically in their methods of interpreting VA data owing to their training, experience, and/or perceptions of local epidemiology, particularly when diagnostic criteria are not standardized amongst different physicians [8-10]. Hence, there may be inter and intra-reviewer variability among physicians that may lead to inconsistencies in COD data, hindering reliable temporal and spatial comparisons of mortality. They mostly use open history to reach a decision and may not account consistently for all indicators. They may also be influenced by their own biases, particularly for less obvious CODs for which decisions have to be made between equally likely diagnoses [11-15]. Moreover, the physician review process incurs remunerative costs, consumes time, and requires the involvement of physicians who are an already overstretched resource in low-income countries [8,16].

Furthermore, a large percentage of CODs assigned by VAs remain undetermined as physicians often disagree over a final COD classification, especially for deaths for which VAs were not successfully completed [10,17-21].

Different alternative methods to the physician review process for interpreting VA data have remained of limited use [22-24]. However, the use of the InterVA model to interpret VA data is a relatively new methodology that has just been explored to have the advantage of achieving the maximum spatial and temporal consistency [25-27]. Moreover, it requires minimal time and labor resources, especially in comparison with the physician review method. Also, it is freely available in the public domain, making it an ideal option for resource-constrained settings [28]. A new version of InterVA, InterVA-4, was launched in August 2012 along with the new WHO standards for VA. It was designed to incorporate the more specialized previous versions of the model for maternal and neonatal deaths, and to build on the experience from InterVA-3 and preceding models [29]. Further details of the approach used in InterVA models are available in a range of peer-reviewed publications which can be found under the "more info" section of its website (www.interva.net).

In order to design appropriate promotive, curative, and rehabilitative health services and to influence policy decisions, information on the burden of mortality at a population level is critically important. In response to this, the current study is designed to evaluate the performance of the InterVA-3 model as the physician alternative method for generating cause-specific mortality data from VAs in northern Ethiopia.

Methods

A population-based cross-sectional study was conducted from March to April, 2012, in Dabat Health and Demographic Surveillance System site (HDSSs) hosted by the University of Gondar. The site is located in a district known as Dabat, northern Ethiopia, and has an estimated population of 46,165 living in 7 rural and 3 urban "kebeles" (the smallest administrative units in Ethiopia). The local communities largely depend on subsistence agriculture and information on vital events, like birth, death, and migration are collected quarterly [30].

Study population and data collection

All adults aged ≥14 years and died between 01 January, 2010, and 15 February, 2012, in the area were included in the study. This period was preferred in order to obtain an adequate number of deaths without marked recall bias. It is believed that adult deaths were remembered very well.

Pre-tested and modified WHO and INDEPTH [31,32] designed VA questionnaire was used to collect the data. The VA questionnaire included an open narrative, medical history, and closed questions. The narrative section was used to record free explanations of the circumstances of death; the medical history sections were used to extract data from medical certificates, and the closed section dealt with specific signs, symptoms, and conditions leading to death. Three trained supervisors and nine data collectors who had rich experience

in the job participated in the data collection processes. After obtaining an informed written consent, the data collectors interviewed a close relative, friend, or neighbor of the deceased person who witnessed the death. Considering the usual mourning period in the study area, data were collected after 45 days for recent death events.

The VA questionnaire was translated into "Amharic" (the local language) and back to English to maintain the consistency of the questions. The training of data collectors and supervisors emphasized issues, such as the selection of eligible respondents, approaching grieving respondents, time of interviews, and compiling narrative responses (ensuring that duration, frequency, severity, and the sequence of symptoms were mentioned). The principal investigator and the supervisors coordinated the interview process, made spot-checks, and reviewed the completed questionnaires daily to ensure the completeness and consistency of the data collected. They also conducted random quality checks by re-interviewing about 10% of the respondents. The VA questionnaire was pre-tested on 25 respondents who lived near Dabat and had similar characteristics with the study population in the district. Based on the pre-test results, the questionnaire was adjusted contextually. Data entry was carried out by the principal investigator and another independent data clerk and was then compared to check for any variations in results.

Interpretation of VA data

The InterVA-3 model and the physician reviewed the same basic data from the VA questionnaire independently. That is, both methods utilized information collected in the open narrative and medical histories section together with the closed-ended section to assign the probable COD.

Physician interpretation

Two independent physicians reviewed each VA questionnaire independently to assign a single COD based on ICD-10. The ICD-10 list had unique codes for diseases, signs, symptoms, abnormal findings, complaints, social circumstances, and external causes of injury [27]. The physicians met subsequently to reach consensus on cases where there were differences of opinion. If no physician consensus was reached after discussion, the COD was regarded as indeterminate. The physicians were trained in procedures on assigning COD and given details of the study area and study population. However, they were not given any special briefing on the probabilistic model so as not to encroach on their professional freedom. In spite of that however, their review process was closely monitored and that they be not direct beneficiaries of the research output was ensured.

Interpretation of the InterVA model

The model relates a range of input indicators, such as age, sex, physical signs and symptoms, medical history, and the circumstances of death to likely CODs using Bayesian probabilities [27]. The model results in up to three likely causes per case when possible; each associated with a quantified

likelihood. To assign an estimate of the overall certainty for that patient, the model gives the average likelihood for a maximum of three CODs [28]. In this study, a high prevalence of Malaria and HIV/AIDS were used as basic epidemiological parameters for the model as their prevalence varies from place to place. Data were entered case-by-case into Microsoft visual FoxPro window of the InterVA version 3.2 to assign the possible COD responsible for the death of each individual.

Comparison of the InterVA model with the physician

The most probable CODs assigned by the model were considered to facilitate comparison with the single CODs which were assigned by the physician. All CODs in both methods were re-categorized into 9 main groups for two reasons. The first reason was to have meaningfully comparable COD categories between both methods. Second, it was more important that the model and the physician arrive at a broad agreement in identifying COD groups with the greatest public health importance at a population level, rather than individual level causes. The list of the 9 main categories used in this study was: pulmonary tuberculosis (TB), HIV/AIDS-related deaths, diabetes, other infectious diseases, digestive diseases, cardiovascular problems, maternity-related deaths, other non-communicable diseases, and injuries/accidents.

Then deaths were aggregated case-by-case to their respective COD categories to determine the CSMFs at the community level by using both the InterVA model and the physician review. Cohen's kappa statistic, sensitivity, specificity, positive predictive value (PPV), and negative predictive value (NPV) were applied to compare the agreement between the InterVA model and the physician review.

Ethical considerations

The study protocol was reviewed and approved by the Institutional Ethical Review Board of the University of Gondar. Then, informed written consent was obtained from the study participants who were close relatives, friends, or neighbors of the deceased after explaining the purpose and the procedures of the study. Confidentiality was granted for information collected from each study participant. Study participants found sick at the time of data collection were referred to the nearest health institution for medical treatment. There was no remuneration for family.

Finally, for the purpose of completeness, findings of the previous study on population characteristics, interpretations of VA data, and others which were specific to pulmonary TB were included in this study [17]. The current and the previous studies were conducted in the same study area and study period using the same data source.

Results

Characteristics of the study population

A total of 408 VA interviews were successfully completed and reviewed by both the InterVA model and the physicians. Of the deceased, 222 (54.4%) were females. Two hundred eighty-one (68.9%) of the deceased were 50 and above years of age.

Most of the deceased, 325 (79.7%), and 298 (90.0%), were married and farmers, respectively. As far as education is concerned, 308 (73.0%) of them were illiterate. The Majority, 306 (75.0%), of the deceased were rural dwellers.

Physician interpretation

Out of the 408 deaths, 329 (80.6%), were successfully assigned a single cause at the first attempt by two physicians. After holding consensus meetings, the physicians successfully assigned a single COD to 61 (15%) more cases. Therefore, on the whole, physicians assigned a single COD to 390 (95.6%) cases. No consensus was reached on 18 (4.4%) cases which were coded as "indeterminate" by the physicians.

Interpretation of the InterVA model

The InterVA-3 model assigned a single COD to 356 (87.3%) cases, two CODs to 52 (12.8%) cases, and three causes to 5 (1.2%) cases. In 10 (2.5%) cases, the InterVA model assigned the COD as "indeterminate". The probabilistic model assigned the likely CODs to all the VAs with a certainty of 75.0% and standard deviation of 2.8.

Comparison of the InterVA model with the physician

There was a general similarity and just slight differences between the InterVA model and the physicians in assigning cause-specific mortality. Out of all deaths in this population, two major groups of causes, pulmonary TB and other non-communicable diseases, accounted for about half of the overall mortality, as determined by both approaches. It is noteworthy that the InterVA model assigned significantly more CODs to pulmonary TB [147 (36.0%)] compared to physicians [94 (23.0%)]. On the other hand, physicians identified HIV/AIDS as a COD more frequently [46 (11.3%)] than the model [31 (7.6%)], (Figure 1).

A direct comparison of the CODs assigned by the physicians to the first CODs assigned by the InterVA model showed that there was an overall agreement in 298 (73%) cases [kappa = 0.49, 95% CI: 0.37-0.60]. The observed level of agreement across the COD categories varied from kappa value of 0.17 to 0.83. A poor level of agreement was observed only for digestive disease, (Table 1).

The results for sensitivities, specificities, PPV, and NPV of the InterVA model in comparison with the physicians were presented for COD categories. The observed sensitivities and specificities across the COD categories varied from 13.3% to 81.9% and 77.7% to 99.5%, respectively, (Table 2).

Discussion

In this study, the probabilistic InterVA model found out very similar results with the physicians for assigning cause-specific mortalities from VA data at the population level. This was true with other studies [7,25,33-35]. The frequencies of mortalities revealed were consistent with the existing knowledge on the burden of diseases among an underdeveloped population in sub-Saharan Africa [36-39], indicating good performance of the InterVA model for generating cause-specific mortality data from VA.
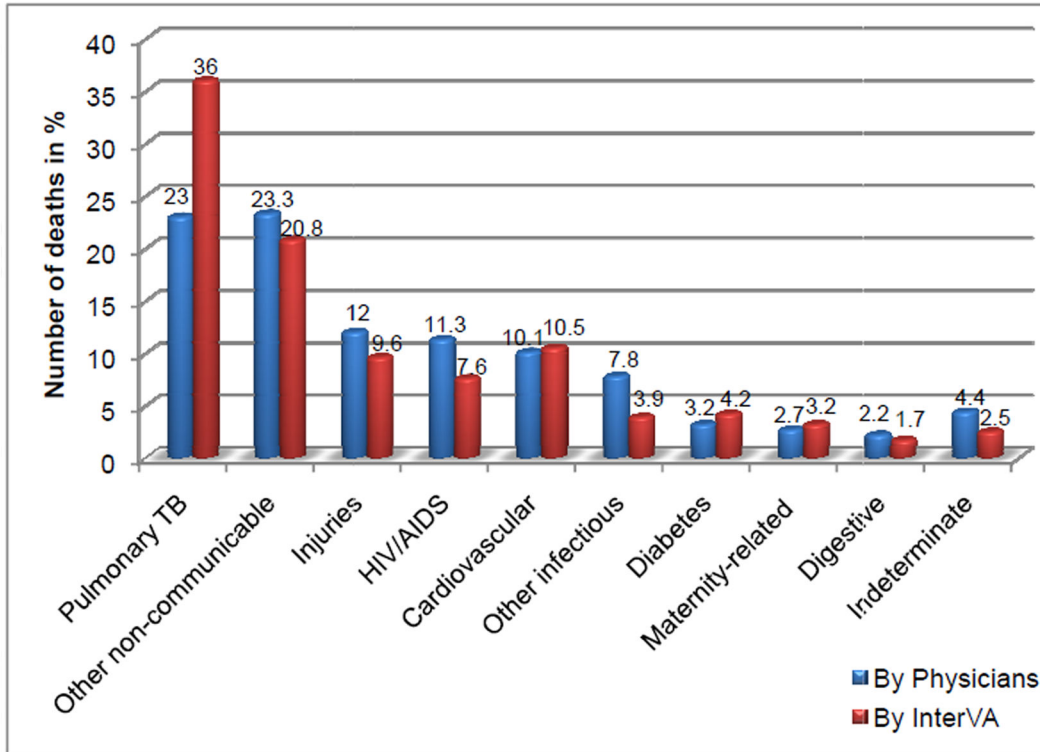
Figure 1. The CSMFs derived from physician review and InterVA model interpretation.
doi: 10.1371/journal.pone.0073463.g001

Table 1. Case-by-case agreement between the InterVA model and the physicians in establishing the CODs in northern Ethiopia from 01 January 2010–15 February 2012.

COD Categories	Kappa(95%CI)
1. Injuries/accidents	0.83 (0.76, 0.90)
2. Maternity-related death	0.76 (0.60, 0.92)
3. Diabetes	0.52 (0.36, 0.68)
4. Pulmonary TB	0.50 (0.40, 0.60)
5. Other infectious diseases	0.46 (0.34, 0.58)
6. Cardiovascular	0.42 (0.33, 0.51)
7. HIV/AIDS-related death	0.40 (0.30, 0.50)
8. Other non-communicable diseases	0.31 (0.23, 0.40)
9. Digestive diseases	0.17 (0.01, 0.33)

doi: 10.1371/journal.pone.0073463.t001

The high discordance observed between the two approaches in assigning pulmonary TB and HIV/AIDS as CODs in this study is supported by other investigations [7,20,25]. This might be due to a great deal of overlap between both disease conditions in terms of clinical symptoms and signs. Furthermore, the re-emergence of pulmonary TB in several countries of the world is spurred by the HIV/AIDS pandemic. This underlies the high level of interconnectedness between both diseases. Moreover, from a public health perspective, control and prevention of either disease cannot be considered

Table 2. Validating the InterVA model against the physician in diagnosing COD categories in northern Ethiopia from 01 January 2010–15 February 2012.

Five COD categories	Sensitivity (%) (%)	PPV (%)	Specificity (%)	NPV
1. Pulmonary TB	81.9	52.4	77.7	93.5
2. Injuries/accidents	76.5	95.1	99.4	96.7
3. Maternity-related death	72.7	80.0	99.5	99.2
4. Diabetes	61.5	47.0	97.7	98.7
5. Cardiovascular	45.8	51.2	94.2	92.9
6. Other infectious diseases	41.4	60.0	97.9	95.6
7. Other non-communicable diseases	36.4	49.1	91.2	86.0
8. HIV/AIDS-related death	33.3	51.6	95.8	91.5
9. Digestive diseases	13.3	25.0	98.5	96.8

doi: 10.1371/journal.pone.0073463.t002

without regard to the other [40-42]. So what is critical is that the collective burden of both diseases in any population is clear, and the InterVA model demonstrated this as successfully as the physicians did.

The observed level of agreement for varied COD categories indicated a fairly good diagnostic performance of the InterVA model. A nearly similar level of overall agreement was observed between the InterVA model and physicians [kappa = 0.42 (0.37-0.48)] in a validation study conducted in Kenya [6]. This confirmed the temporal and spatial consistency of the InterVA model for establishing cause-specific mortalities. However, a higher level of agreement between both approaches was observed in studies which utilized data collected by a demographic surveillance system [19,43]. The reason for the poor level of agreement observed for digestive diseases could be the overlapping nature of the clinical signs and symptoms with other diseases, especially HIV/AIDS.

Studies indicated that the validation of VA is considered to have an acceptable level of diagnostic accuracy at the population level, if sensitivity and specificity are at least 50% and 90%, respectively [22]. In this study, the observed sensitivity values were above 60% for pulmonary TB, injuries/ accidents, maternity-related death and diabetes. However, lower sensitivity values were observed for deaths related with cardiovascular diseases, other infectious and non-communicable diseases, HIV/AIDS and digestive diseases. In previous studies, the sensitivity value for cardiovascular-related COD varied from 25% to 87% [4-6,22,23,44-46]. The observed specificity values were good, except for pulmonary TB. These criteria of validation (sensitivity at least 50% and specificity at least 90%) are not uniformly regarded as acceptable [47] because low sensitivity and specificity does not necessarily imply low level of accuracy, or relatively high sensitivity and specificity may result in serious misclassification errors. In the case of low sensitivity and specificity, the false positives and false negatives may counterbalance, and may not affect the VA accuracy [44,48].

Literatures reveal that there are robust validation metrics other than Cohen's kappa, sensitivity and specificity to assess how well a VA method estimates CSMFs. These are: chance-corrected concordance, absolute CSMF errors, relative CSMFs error, and CSMF accuracy [5,7,20-22,45,49-54]. An average chance-corrected concordance across causes is recommended for assessing how well a method does at individual COD assignment. This metric is insensitive to the CSMF composition of the test sets and corrects for the degree to which a method will get the cause correct due strictly to chance. For the evaluation of CSMF estimation, CSMF accuracy is proposed. CSMF accuracy is defined as one minus the sum of all absolute CSMF errors across causes divided by the maximum total error. It is scaled from zero to one and can generalize a method's CSMF estimation capability, regardless of the number of causes where a value of one means no error in the predicted CSMFs, and a value of zero means the method is equivalent to the worst possible method of assigning cause fractions [54].

A validation study of VA often faces the question as to how to obtain a true gold standard. Several studies have used CODs based on hospital diagnoses as the gold standard [5,6,9,44,50]. However, hospital diagnoses have limitations as gold standard since the composition and distribution of hospital CODs may not be representative of deaths occurring in the community. Moreover, in resource-constrained healthcare settings, hospital diagnoses which are often unavailable are of low quality when available and are limited by inadequate clinical data and record keeping. Furthermore, the ability to recognize, recall, and report signs of illnesses may be different among hospital users and nonhospital users. In this study, physician review was used as a reference standard to examine InterVA. The use of physician review was the only alternative source of COD assessment for this study population. This choice however has limitations. Physicians are influenced by their experience, perception, and interpretation of local epidemiology that may lead to inconsistencies in COD data, hindering reliable temporal and spatial comparisons of COD. Moreover, they often use open history to reach decisions and may not account consistently for all the indicators. They may also be influenced by their own biases, particularly for less obvious CODs for which decisions had to be made between equally likely diagnoses. These inherent limitations of physicians could lead them to misinterpret some of the VA data and finally reach a wrong conclusion of COD. Previous VA literature has also suggested that the physician review is not a robust method to interpret VA data [55]. Therefore, considering a physician review as a true gold standard to validate the InterVA model could influence the true diagnostic accuracy of the InterVA model.

The current study can only provide evidence on how the COD estimates derived from InterVA compared to those ascertained by the physician review. It cannot infer the performance of the InterVA compared to other existing methods which have been shown to perform better than InterVA previously. Studies proved that other automated options such as the Tariff Method, Simplified Symptom Pattern, Random Forests, and Machine Learning for the analysis of VA data have validated performance equal to or better than physician review [53,56-59]. Given the widespread use of VA for understanding the burden of disease and setting health intervention priorities in areas that lack reliable vital registrations systems, accurate analysis of VAs is essential. Therefore, users should be aware of the suboptimal performance of the InterVA in relation to other methods.

The other possible limitation of this study could be the cross-sectional study design which might not be appropriate for establishing cause-specific mortalities accurately. Using data from a well-established longitudinal demographic surveillance system may reduce the effect of recall biases associated with a long recall period. The absence of some variables in the WHO adult VA questionnaire is a factor challenging the diagnostic accuracy of the InterVA model. The model does not employ open-ended questions which are more relevant in a society with poor knowledge of symptoms of certain diseases and where more local terms may be used in this case. Even though the data collectors in this study had a long experience in field data collection processes, none of them had academic expertise in medical diagnosis of diseases which might adversely affect the quality of the data collected. This could in turn result in misleading interpretations by both the InterVA model and the physicians and finally lead to a wrong conclusion of COD. This study applied 9 broad COD categories

which clearly increased the possibility that the two methods would agree. Therefore, an additional sensitivity analysis should be performed to see the impact of any change in the COD categories chosen on the level of agreement between the two methods. Another limitation could be the relatively small sample size of the study which might also contribute to the underestimation of the sensitivity and specificity values. Besides, the indeterminate probability of the COD would decrease if more than two physicians reviewed the data, but this was not done due to the inadequacy of the budget.

Conclusions

In understanding the burden of disease and setting health intervention priorities in areas that lack reliable vital registrations systems, an accurate analysis of VAs is essential. Therefore, users should be aware of the suboptimal performance of the InterVA model. Similar validation studies need to be undertaken considering the limitation of the physician review as gold standard since physicians may misinterpret some of the VA data and finally reach a wrong conclusion of the COD.

Supporting Information

File S1. Detailed descriptions for the calculations of kappa, sensitivity, specificity, positive and negative predictive values were available for further reference.
(DOCX)

References

1. Setel PW, Macfarlane SB, Szreter S, Mikkelsen L, Jha P et al. (2007) A scandal of invisibility: making everyone count by counting everyone. Lancet 370: 1569–1577. doi:10.1016/S0140-6736(07)61307-5. PubMed: 17992727.
2. Byass P (2007) Who needs cause-of-death data? PLOS Med 4(11): 333. doi:10.1371/journal.pmed.0040333. PubMed: 18031198.
3. Fottrell E (2009) Dying to count: mortality surveillance in resource-poor settings. Glob Health Action 2: 10. PubMed: 200272693402/ghav2i0. 1926.
4. Kahn K, Tollman SM, Garenne M, Gear JS (2005) Validation and application of verbal autopsies in a rural area of South Africa. Trop Med Int Health 5(11): 824–831.
5. Setel PW, Whiting DR, Hemed Y, Chandramohan D, Wolfson LJ et al. (2006) Validity of verbal autopsy procedures for determining cause of death in Tanzania. Trop Med Int Health 11(5): 681–696. doi:10.1111/j.1365-3156.2006.01603.x. PubMed: 16640621.
6. Bauni E, Ndila E, Mochamah G, Nyutu G, Matata L et al. (2011) Validating Physician-Certified Verbal Autopsy and Probabilistic Modeling (InterVA) Approaches to Verbal Autopsy Interpretation Using Hospital Causes of Adult Deaths. Popul Health Metrics 9: 49. doi:10.1186/1478-7954-9-49. PubMed: 21819603.
7. Oti SO, Kyobutungi C (2010) Verbal autopsy interpretation: a comparative analysis of the InterVA model versus physician review in determining causes of death in the Nairobi DSS. Popul Health Metrics 8: 21. doi:10.1186/1478-7954-8-21. PubMed: 20587026.
8. Fottrell E, Byass P (2010) Verbal autopsy: methods in transition. Epidemiol Rev 32: 38-55. doi:10.1093/epirev/mxq003. PubMed: 20203105.
9. Coldham C, Ross D, Quigley M, Segura Z, Chandramohan D (2000) Prospective validation of a standardized questionnaire for estimating childhood mortality and morbidity due to pneumonia and diarrhoea. Trop Med Int Health 5: 134-144. doi:10.1046/j.1365-3156.2000.00505.x. PubMed: 10747274.
10. Soleman N, Chandramohan D, Shibuya K (2006) Verbal autopsy: current practices and challenges. Bull World Health Organ 84: 239-245. doi:10.2471/BLT.05.027003. PubMed: 16583084.
11. Ronsmans C, Vanneste AM, Chakraborty J, Ginneken JV (1998) A comparison of three verbal autopsy methods to ascertain level and causes of death in Matlab Bangladesh. Int J Epidemiol 27: 660–666. doi:10.1093/ije/27.4.660. PubMed: 9758122.
12. Todd JE, De Francisco A, Dempsey TJO, Greenwood BM (1994) The limitations of verbal autopsy in a malaria endemic region. Ann Trop Paediatr 14: 31–36. PubMed: 7516132.
13. Vergnano S, Fottrell E, Osrin D, Lewycka S, Costello A et al. (2011) Adaptation of a probabilistic method (InterVA) of verbal autopsy to improve the interpretation of cause of stillbirth and neonatal death in Malawi, Nepal, and Zimbabwe. Popul Health Metrics 9: 48. doi:10.1186/1478-7954-9-48.
14. Byass P, Kahn K, Fottrell E, Mee P, Collinson MA (2011) Using verbal autopsy to track epidemic dynamics: the case of HIV-related mortality in South Africa. Popul Health Metrics 9: 46. doi:10.1186/1478-7954-9-46. PubMed: 21819601.
15. WHO (2004) Beyond the Numbers: Reviewing maternal deaths and complications to make pregnancy safer. Geneva.
16. Byass P, Huong DL, Minh HV (2003) A probabilistic approach to interpreting verbal autopsies: methodology and preliminary validation in Vietnam. Scand J Public Health Suppl 62: 32-37. PubMed: 14649636.
17. Tadesse S, Tadesse T (2012) Evaluating the performance of Interpreting Verbal Autopsy 3.2 model for establishing pulmonary tuberculosis as a cause of death in Ethiopia: a population-based cross-sectional study. BMC Public Health 12: 1039. doi:10.1186/1471-2458-12-1039. PubMed: 23190770.
18. Misganaw A, Araya T, Aneneh A, Hailemariam D (2012) Validity of verbal autopsy method to determine causes of death among adults in the urban setting of Ethiopia. BMC Med Res Methodol 12: 130. doi:10.1186/1471-2288-12-130. PubMed: 22928712.
19. Vergnano S, Fottrell E, Osrin D, Lewycka S, Costello A et al. (2011) Adaptation of a probabilistic method (InterVA) of verbal autopsy to improve the interpretation of cause of stillbirth and neonatal death in Malawi, Nepal, and Zimbabwe. Popul Health Metrics 9: 48. doi:10.1186/1478-7954-9-48.
20. Byass P, Kahn K, Collinson MA, Tollman SM, Fottrell E (2010) Moving from Data on Deaths to Public Health Policy in Agincourt, South Africa: Approaches to Analyzing and Understanding Verbal Autopsy Findings. PLOS Med 7(8): e1000325.
21. Freeman JV, Christian P, Khatry SK, Adhikari RK, Clerq SCL et al. (2005) Evaluation of neonatal verbal autopsy using physician review versus algorithm-based cause-of-death assignment in rural Nepal. Paediatr Perinat Epidemiol 19: 323-331. doi:10.1111/j.1365-3016.2005.00652.x. PubMed: 15958155.
22. Quigley MA, Chandramohan D, Rodrigues LC (1999) Diagnostic accuracy of physician review, expert algorithms and data derived algorithms in adult verbal autopsies. Int J Epidemiol 28: 1081–1087. doi:10.1093/ije/28.6.1081. PubMed: 10661651.
23. Boulle A, Chandramohan D, Weller P (2001) A case study of using artificial neural networks for classifying cause of death from verbal autopsy. Int J Epidemiol 30(3): 515-520. doi:10.1093/ije/30.3.515. PubMed: 11416074.
24. Quigley MA, Chandramohan D, Setel P, Binka F, Rodrigues LC (2000) Validity of data-derived algorithms for ascertaining causes of adult

death in two African sites using verbal autopsy. Trop Med Int Health 5(1): 33–39. doi:10.1046/j.1365-3156.2000.00517.x. PubMed: 10672203.

25. Fantahun M, Berhane Y, Fottrell E, Wall S, Högberg U et al. (2006) Assessing a new approach to verbal autopsy interpretation in a rural Ethiopian community: the InterVA model. Bull World Health Organ 84: 204–210. doi:10.2471/BLT.05.028712. PubMed: 16583079.

26. Reeves BC, Quigley M (1997) A review of data derived methods for assigning cause of death from verbal autopsy data. Int J Epidemiol 26(5): 1080–1089. doi:10.1093/ije/26.5.1080. PubMed: 9363531.

27. Byass P, Fottrell E, Huong DL, Berhane Y, Corrah T et al. (2006) Refining a probabilistic model for interpreting verbal autopsy data. Scand J Public Health 34: 26–31. doi:10.1080/14034940510032202. PubMed: 16449041.

28. Inter VA 3.2 model. Available: http://www.interva.net. Accessed 2012 February 12.

29. Byass P, Tollman SM, Kahn K, Fottrell E, Ambruoso LD et al. (2012) Strengthening standardized interpretation of verbal autopsy data: the new InterVA-4 tool. Glob Health Action 5: 19281.

30. Central Statistical Authority (2007) Population and Housing Census of Ethiopia: Results for Amhara Regional State. Ethiopia: Addis Ababa.

31. INDEPTH network: INDEPTH Standardized Verbal Autopsy Questionnaire. Available: www.Indepthnetwork.org/core_documents/indepthtools.htm. Accessed 2012 February 20.

32. WHO International Classification of Diseases (ICD). Available: http://www.who.int/classifications/help/icdfaq/en/index.html. Accessed 2012 February 22.

33. Byass P, Fottrell E, Witten KH, Bhattacharya S, Fitzmaurice AE et al. (2007) Revealing the burden of maternal mortality: a probabilistic model for determining pregnancy-related causes of death from verbal autopsies. Popul Health Metrics 5: 1. doi:10.1186/1478-7954-5-1. PubMed: 17288607.

34. Kyobutungi C, Ziraba AK, Ezeh A, Yé Y (2008) The burden of disease profile of residents of Nairobi's slums: Results from a Demographic Surveillance System. Popul Health Metrics 6: 1. doi: 10.1186/1478-7954-6-1. PubMed: 18331630.

35. Van EAM, Adazu K, Ofware P, Vulule J, Hamel M et al. (2008) Causes of deaths using verbal autopsy among adolescents and adults in rural western Kenya. Trop Med Int Health 13(10): 1314-1324. doi:10.1111/j.1365-3156.2008.02136.x. PubMed: 18721187.

36. Jamison DT, Feachem RG, Makgoba MW, Bos ER, Baingana FK et al. (2006) Disease and mortality in Sub-Saharan Africa. 2nd Edition. World Bank.

37. Lozano R, Naghavi M, Foreman K, Lim S, Shibuya K et al. (2012) Global and regional mortality from 235 causes of death for 20 age groups in 1990 and 2010: a systematic analysis for the Global Burden of Disease Study 2010. Lancet 380(9859): 2095-2128. doi:10.1016/S0140-6736(12)61728-0. PubMed: 23245604.

38. Murray CJL, Lopez AD (2009) Mortality by cause for eight regions of the world: Global Burden of Disease Study. Lancet 349(9061): 1269-1276. PubMed: 9142060.

39. Aikins AG, Unwin N, Agyemang C, Allotey P, Campbell C et al. (2010) Tackling Africa's chronic disease burden: from the local to the global. Globalization Health 6: 5. doi:10.1186/1744-8603-6-5. PubMed: 20403167.

40. Keshinro B, Diul MY (2006) HIV-TB: epidemiology, clinical features and diagnosis of smear negative TB. Trop Doct 36(2): 68-71. doi: 10.1258/004947506776593396. PubMed: 16611435.

41. Myers J, Sepkowitz K (2008) HIV/AIDS and TB. International Encyclopedia Public Health: 421-430.

42. Abdool KSS, Churchyard GJ, Abdool KQ, Lawn SD (2009) HIV infection and tuberculosis in South Africa: an urgent need to escalate the public health response. Lancet 374: 921-933. doi:10.1016/S0140-6736(09)60916-8. PubMed: 19709731.

43. Herbst AJ, Mafojane T, Newell ML (2011) Verbal autopsy-based cause-specific mortality trends in rural KwaZulu-Natal, South Africa, 2000-2009. Popul Health Metrics 9: 47. doi:10.1186/1478-7954-9-47. PubMed: 21819602.

44. Chandramohan D, Maude GH, Rodrigues LC, Hayes RJ (1998) Verbal autopsies for adult deaths: their development and validation in a multicentre study. Trop Med Int Health 3(6): 436–446. doi:10.1046/j.1365-3156.1998.00255.x. PubMed: 9657505.

45. Yang G, Rao C, Ma J, Wang L, Wan X et al. (2006) Validation of verbal autopsy procedures for adult deaths in China. Int J Epidemiol 35(3): 741–748. doi:10.1093/ije/dyi181. PubMed: 16144861.

46. Chandramohan D, Setel P, Quigley M (2001) Effect of misclassification of causes of death in verbal autopsy: can it be adjusted? Int J Epidemiol 30(3): 509–514. doi:10.1093/ije/30.3.509. PubMed: 11416073.

47. Chandrahoman D (2001) Verbal autopsy tools for adults deaths. PhD Thesis, London School of Hygiene and Tropical Medicine.

48. Anker M (1997) The effect of misclassification error on reported cause-specific mortality fractions from verbal autopsy. Int J Epidemiol 26: 1090–1096. doi:10.1093/ije/26.5.1090. PubMed: 9363532.

49. Marsh DR, Sadruddin S, Fikree FF, Krishnan C, Darmstadt GL (2003) Validation of verbal autopsy to determine the cause of 137 neonatal deaths in Karachi, Pakistan. Paediatr Perinat Epidemiol 17: 132–142. doi:10.1046/j.1365-3016.2003.00475.x. PubMed: 12675779.

50. Polprasert W, Rao C, Adair T, Pattaraarchachai J, Porapakkham Y et al. (2010) Cause-of-death ascertainment for deaths that occur outside hospitals in Thailand: application of verbal autopsy methods. Popul Health Metrics 8: 13. doi:10.1186/1478-7954-8-13. PubMed: 20482760.

51. Khademi H, Etemadi A, Kamangar F, Nouraie M, Shakeri R (2010) Verbal Autopsy: Reliability and Validity Estimates for Causes of Death in the Golestan Cohort Study in Iran. PLOS ONE 5: e11183. doi: 10.1371/journal.pone.0011183. PubMed: 20567597.

52. Kumar R, Thakur JS, Rao BT, Singh MMC, Bhatia SPS (2006) Validity of verbal autopsy in determining causes of adult deaths. Indian J Public Health 50: 90-94. PubMed: 17191410.

53. Murray CJL, Lopez AD, Feehan DM, Peter ST, Yang G (2007) Validation of the Symptom Pattern Method for Analyzing Verbal Autopsy Data. PLOS Med 4: e327. doi:10.1371/journal.pmed.0040327. PubMed: 18031196.

54. Murray CJL, Lozano R, Flaxman AD, Vahdatpour A, Lopez AD (2011) Robust metrics for assessing the performance of different verbal autopsy cause assignment methods in validation studies. Popul Health Metrics 9: 28. doi:10.1186/1478-7954-9-28. PubMed: 21816106.

55. Lozano R, Lopez AD, Atkinson C, Naghavi M, Flaxman AD et al. (2011) Performance of physician-certified verbal autopsies: multisite validation study using clinical diagnostic gold standards. Popul Health Metrics 9: 32. doi:10.1186/1478-7954-9-32. PubMed: 21816104.

56. Murray CJL, James SL, Birnbaum JK, Freeman MK, Lozano R et al. (2011) Simplified Symptom Pattern Method for verbal autopsy analysis: multisite validation study using clinical diagnostic gold standards. Popul Health Metrics 9: 30. doi:10.1186/1478-7954-9-30. PubMed: 21816099.

57. James SL, Flaxman AD, Murray CJL (2011) Performance of the Tariff Method: validation of a simple additive algorithm for analysis of verbal autopsies. Popul Health Metrics 9: 31. doi:10.1186/1478-7954-9-31.

58. Flaxman AD, Vahdatpour A, Green S, James SL, Murray CJL (2011) Random forests for verbal autopsy analysis: multisite validation study using clinical diagnostic gold standards. Popul Health Metrics 9: 29. doi:10.1186/1478-7954-9-29. PubMed: 21816105.

59. Lozano R, Freeman MK, James SL, Campbell B, Flaxman AD et al. (2011) Performance of InterVA for assigning causes of death to verbal autopsies: multisite validation study using clinical diagnostic gold standards. Popul Health Metrics 9: 50. doi:10.1186/1478-7954-9-50. PubMed: 21819580.