

RESEARCH ARTICLE

Open Access



How much do the physician review and InterVA model agree in determining causes of death? a comparative analysis of deaths in rural Ethiopia

Berhe Weldearegawi^{1,2*}, Yohannes Adama Melaku¹, Geert Jan Dinant³ and Mark Spigt^{3,1}

Abstract

Background: Despite it is costly, slow and non-reproducible process, physician review (PR) is a commonly used method to interpret verbal autopsy data. However, there is a growing interest to adapt a new automated and internally consistent method called InterVA. This study evaluated the level of agreement in determining causes of death between PR and the InterVA model.

Methods: Verbal autopsy data for 434 cases collected between September 2009 and November 2012, were interpreted using both PR and the InterVA model. Cohen's kappa statistic (κ) was used to compare the level of chance corrected case-by-case agreement in the diagnosis reached by the PR and InterVA model.

Results: Both methods gave comparable cause specific mortality fractions of communicable diseases (36.6 % by PR and 36.2 % by the model), non-communicable diseases (31.1 % by PR and 38.2 % by the model) and accidents/injuries (12.9 % by PR and 10.1 % by the model). The level of case-by-case chance corrected concordance between the two methods was 0.33 (95 % CI for $\kappa = 0.29$ – 0.34). The highest and lowest agreements were seen for accidents/injuries and non-communicable diseases; with $\kappa = 0.75$ and $\kappa = 0.37$, respectively.

Conclusion: If the InterVA were used in place of the existing PR process, the overall diagnosis would be fairly similar. The methods had better agreement in important public health diseases like; TB, perinatal causes, and pneumonia/sepsis; and lower in cardiovascular diseases and neoplasms. Therefore, both methods need to be validated against a gold-standard diagnosis of death.

Keywords: Verbal autopsy, InterVA, Agreement, Cause of death

Background

Information that comes from accurate and complete records of deaths – who died of what – is a huge resource for evidence-based health planning and development [1]. However, most deaths in low-income countries remain unregistered due to the absence of civil registration system [1–3]. In countries, where deaths are not routinely recorded and classified by cause, verbal autopsy (VA) has become an alternative technique [4–6]. Verbal autopsy

covers the entire process of interviewing close caregivers about the circumstances proceeding to death [4, 7, 8]. Verbal autopsy data are collected by trained lay interviewers, and then interpreted into a probable cause of death [2, 4, 7, 8]. Either of the following methods can be applied to derive the cause of death from VA data: physician review, physician review using an algorithm, computer algorithms and InterVA model [2, 9].

Physician review (PR) is the most commonly used method of interpreting VA data [4, 9]. According to this method, two independent physicians review VA questionnaires to assign probable cause of death and corresponding International Classification of Diseases (ICD) code [9, 10]. However, the diagnosis reached by physicians can vary

* Correspondence: berheph@gmail.com

¹Department of Public Health, Mekelle University, Mekelle, Ethiopia

²Centre of Cardiovascular Research and Education in Therapeutics, Department of Epidemiology and Preventive Medicine, Monash University, 99 Commercial Road, Melbourne, VIC 3004, Australia

Full list of author information is available at the end of the article

depending on their training, experience and knowledge of local epidemiology; which in turn limit the internal consistency and comparability of findings [9, 10]. Physician review is also costly and time-taking. A single verbal autopsy review may take up to half an hour of physicians' time; competing the time required for patient care [3, 9, 10]. Despite all its shortcomings, numerous validation studies have demonstrated PR to be capable of producing reasonably valid cause of death information, and is used in many surveillance systems [2, 5, 11].

Recently, an automated method of interpreting verbal autopsy, called InterVA, has been developed and used [1, 4, 7, 10]. It calculates the probability of a set of causes of death, given the presence of indicators (circumstances, signs, and symptoms) reported in VA interviews [10, 12]. Although statistical modeling of this sort may not reflect the subjective subtleties of physicians' review; the InterVA is faster, cheaper and internally consistent [4, 9, 10]. Therefore, there is an increasing interest to shift from PR, which is widely used in several research centers and surveillance systems, in to the automated InterVA approach [2, 4, 5, 9, 12]. However, the reliability of the diagnosis that can be reached by using these methods is not sufficiently studied. According to prior studies which compared PR and InterVA, the level of agreement vary from low ($\kappa = 0.27$) to moderate ($\kappa = 0.42$ – 0.48) [10, 13–15].

The Kilite Awlaelo Health and Demographic Surveillance System (KA-HDSS), located in northern Ethiopia, has been using PR method to determine cause of death since September 2009. We used data from the KA-HDSS to measure the agreement in diagnosis between physician review and the computer-based InterVA-4 model.

Methods

Study setting and population

This study was undertaken in the KA-HDSS, which is a longitudinal population-based surveillance system established in 2009. It operates in ten contiguous Kebeles (smallest administrative unit with an average population of 5000). The KA-HDSS is a member of the INDEPTH Network [16]. Details of the surveillance system, study population, and operating procedures have been published previously [2, 17–19]. The surveillance started with a baseline population of 66,453 individuals living in 14,453 households. Data on core health and demographic events (pregnancy status, birth, death, marital status change, and migrations) and verbal autopsy are collected during regular house-to-house visits.

Verbal autopsy tool and the interview process

Death, to any member of the KA-HDSS cohort, was identified by trained full time data collectors during a regular household visit. Adult relative who was a

caregiver during the terminal illness was interviewed using the standard VA questionnaire adapted from the World Health Organization (WHO), INDEPTH Network, and Sample Vital Registration with Verbal Autopsy (SAVVY) [16, 20, 21]. The VA tool was translated into Tigrigna, the local language of the study area, and re-checked again for its consistency with the original version. Three separate questionnaires for the three age groups: neonate, post-neonate and children (29 days to 15 years) and adults (>15 years) were used.

Interpretation of VA data into probable cause of death

The same verbal autopsy data were interpreted in to probable cause of death (CoD) using physician review and the InterVA-4 model. In the case of PR, two physicians independently reviewed the completed VA questionnaires to assign underlying CoD using the International Classification of Diseases (ICD) coding [22]. Agreement in diagnosis between the two physicians was checked by the research team. Physicians listed up to three underlying CoD, ranked according to their chance to be a CoD in a specific subject. Cause of death was established only if the two physicians coded similar diagnosis for the case. When disagreement in diagnosis existed, the case was reviewed by a third tie-breaker physician. Final diagnosis was then assigned based on the agreement between the third and any of the two physicians. Yet, if three of the physicians assigned different diagnosis for the same case, the diagnosis was considered as “undetermined”.

The InterVA-4 model (version 4.02) was used to interpret VA data into probable cause(s) of death. As described by Byass et al. [7], the model is based on Bayes' theorem, which calculates the probability of a set of CoD given the presence of indicators (circumstances, signs, and symptoms), reported in VA interviews [7, 10]. The InterVA, which is a freely available package (<http://www.interva.net/>), requires extraction of the defined set of indicators from the VA questionnaire of each case in to a text files in the comma separated variable (.csv) format [23]. This dataset is then processed to generate a summary of as many as three possible causes of death with their corresponding likelihood [2, 12].

Before running the model, the InterVA requires adjusting for the incidence of malaria and HIV/AIDS in the study population as “high” or “low”. This would conceptually mean a physician's knowledge of whether his/her current case comes from a setting where malaria is more or less likely, irrespective of particular symptoms [23]. National morbidity data from Ethiopia were widely available in the form of prevalence rate, which was used to set the level of indicators in this study. The prevalence of malaria and HIV/AIDS in Ethiopia are estimated to be 1 % and 1–1.9 %, respectively [24, 25]. Therefore, in this study, we set the prevalence of both malaria and

HIV/AIDS to be “low”. The current report is based on VA data for deaths from September 2009 to November 2012. In the present study, the comparison between the two methods was made based on the most probable CoD per case, assigned by both methods rather than all three possible causes.

Sample size was calculated using the sample size for the kappa-statistic of interrater agreement, as implemented for Stata [26]. Based on previous studies, we assumed that physicians would identify the most common outcome (infectious diseases) in 58 % of the cases and the InterVA model in 61 % of the cases [10]. Besides, existing estimates on the concordance between the two methods range between 0.27 and 0.80 [10, 27]. Thus, in the present study, on average we expected a concordance of 0.5. With an absolute precision of 0.1 and significance level 0.05, this study requires a minimum sample size of 332 cases [27]. However, there were 434 cases that have both PR and InterVA based COD, and all of them were included in the present study.

Data analysis

Probable CoD for the 434 deaths was attributed to 47 and 67 possible causes by the InterVA model and PR, respectively. To make direct case-by-case comparison, causes of death which were specific to either of the methods were re-categorized into broader categories common to both methods. For example, physicians coded more specific causes like “eclampsia” and “antepartum” and “post-partum hemorrhage”, while the InterVA has only one category of “maternity-related deaths”. In this case they were grouped in to “maternal causes”. Finally, 19 CoD common to both methods were prepared for the analysis (see Additional file 1). The PR process use ICD-10 coding, while the InterVA use “WHO VA cause of death code”. Thus, to facilitate the comparison, outputs of the InterVA model were re-coded in to equivalent ICD-10 code.

Cause specific mortality fraction (CSMF) was used to describe the proportion of cases attributed to probable CoD. Chance corrected concordance between the physician review and InterVA was measured using Cohen's kappa coefficient with corresponding 95 % confidence interval (CI). Data analyses were carried out using Stata version 11.2 for Windows.

Ethical statement

The study protocol was approved by the Ethiopian Science and Technology Agency (IERC-0030). Ethical approval, with a reference number ERC 0432/2014, was also obtained from the Health Research Ethics Review Committee (HRERC) of Mekelle University. Informed verbal consent was obtained from an eligible respondent. This consent procedure was stated in the proposal which was approved by the ethical review

committee. To keep confidentiality, data containing personal identifiers were not shared with third party.

Results

Verbal autopsy data for a total of 434 deaths were interpreted using physician review and InterVA. The median age at death was 52 years (SD = 32.5, IQR = 62). People aged 65 years and above, and children accounted for the largest and smallest proportions accounting for 43.3 and 3.9 % of the total cases, respectively (Table 1).

Causes of death by physicians and InterVA model

Table 2 presents the number of cases (and CSMF) assigned to broad categories of CoD. Out of the total cases, physicians assigned at least one cause of death for 369 cases (85.0 %), while the InterVA did it for 405 cases (93.3 %). The rest, 15 and 6.7 % have undetermined cause of death in the PR and InterVA, respectively. Both methods attributed nearly the same proportion to communicable causes; 159 cases (36.6 %) by physicians and 157 cases (36.2 %) by the InterVA model. Comparable proportions were also attributed to chronic non-communicable diseases (NCDs); 166 cases (38.2 %) by the InterVA and 135 cases (31.1 %) by physicians.

We also compared the frequency of causes of death attributed to specific causes by each method. Figure 1 compares CSMF generated for both methods. Tuberculosis (TB) was the leading cause of death according to the InterVA (13.8 %), and the third commonest cause by PR (12.7 %). Similarly, comparable proportions of deaths were attributed to accidents/injuries, and cardiovascular causes

Table 1 Socio-demographic characteristics of verbal autopsy cases, KA-HDSS, 2009–2012

Characteristics	Frequency (%)
Age	
Neonates	38 (8.8)
Infants	25 (5.8)
Children (1–4)	17 (3.9)
5–14	24 (5.5)
15–49	91 (21.0)
50–64	51 (11.8)
65 years and above	188 (43.3)
Mean age (95 % CI)	47.3 (42.2–50.4)
Sex	
Male	235 (54.1)
Female	199 (45.8)
Place of death	
Home	352 (81.1)
Health facility	46 (10.6)
Other	36 (8.3)

Table 2 Comparison of broad causes of death by PR and InterVA model, KA-HDSS, 2014

Probable cause of death	Physician review		InterVA	
	Number	CSMF	Number	CSMF
Communicable causes	159	36.6	157	36.2
Non-communicable causes	135	31.1	166	38.2
Undetermined	65	15.0	29	6.7
Accidents/injuries	56	12.9	44	10.1
Perinatal causes	19	4.4	38	8.8
Total	434	100.0	434	100.0

CSMF: cause specific mortality fraction

in both methods. However, there was notable difference in the frequency of deaths assigned to neoplasms. The InterVA model assigned twice more deaths to neoplasm (11.8 %) than the PR (5.3 %). Similarly, the model attributed twice as much to acute lower respiratory tract infections (ALRTI) (10.6 %) than PR (5.3 %).

Agreement in diagnosis between physician review and InterVA model

In this study, the overall agreement between the InterVA and PR on the broad causes of death was 56.5 % ($\kappa = 0.40$, 95 % CI: 0.39–0.45) [Table 3]. However, on specific causes

of death level (19 causes), the two methods agreed on 165 cases (38.0 %) ($\kappa = 0.33$, 95 % CI: 0.29–0.34). The level of agreement has notably varied by cause of death. The highest and lowest agreement were shown for accidents/injuries and NCDs; with $\kappa = 0.75$ and $\kappa = 0.37$, respectively. Similarly, there was a moderate agreement for perinatal causes ($\kappa = 0.59$), TB ($\kappa = 0.59$); but not for cardiovascular causes ($\kappa = 0.23$) and neoplasms ($\kappa = 0.27$). The two methods had a better agreement in children than adults and the old age group.

Discussion

This study compared the level of case-by-case agreement in diagnosis between the InterVA model and PR methods. In general, the CSMF for the major causes of death were comparable in both methods and also consistent with previous findings [2, 11, 13]. However, the overall case-by-case agreement in diagnosis lies within the fair range of agreement [28]. The level of agreement has varied by causes of death, age and sex of the deceased, ranging from fair ($\kappa = 0.23$ for cardiovascular diseases) to substantial level agreement ($\kappa = 0.75$ for accidents/injuries).

The proportions of deaths attributed to communicable causes by both methods were similar and consistent to

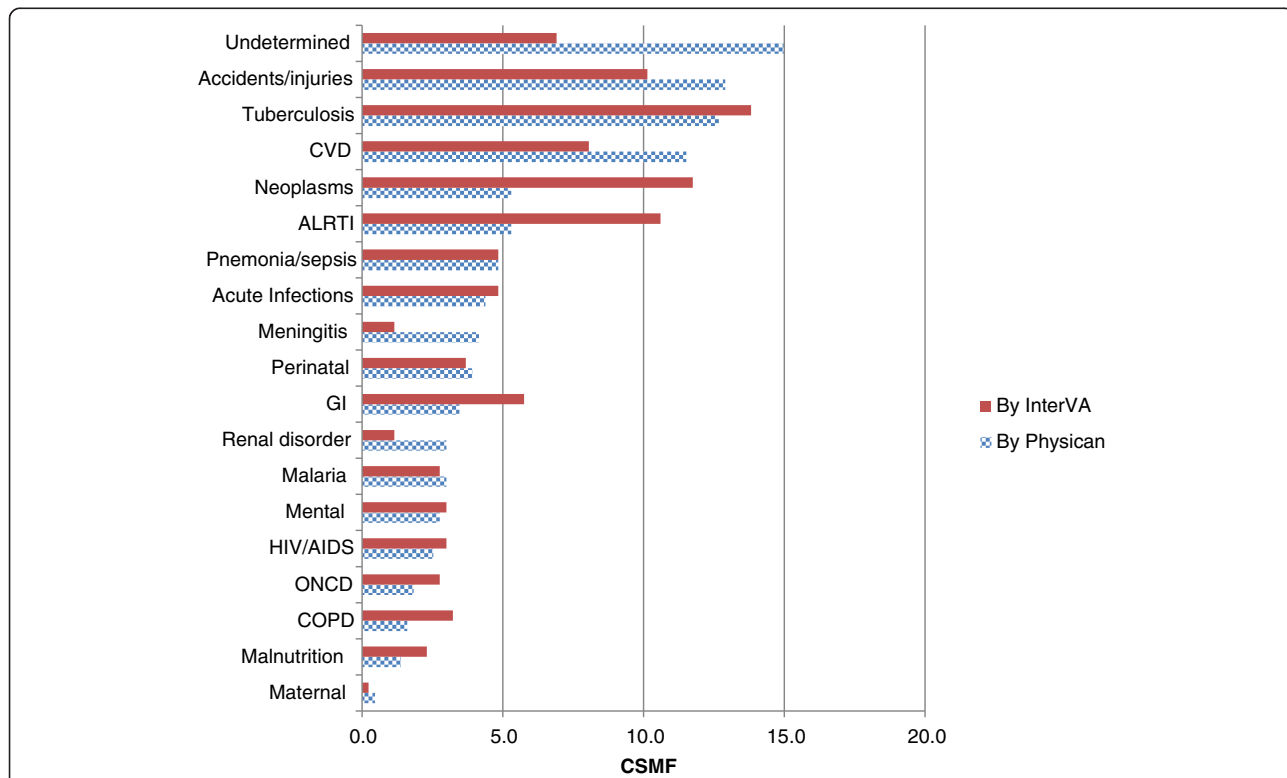


Fig. 1 Comparison of specific causes of death assigned by PR and InterVA model, KA-HDSS, 2014. UND-undetermined, CVD-cardiovascular diseases, COPD- chronic obstructive pulmonary disease, ONCD-other non-communicable diseases, Gastro Intestinal disorder, ALRTI-acute lower tract infections

Table 3 Level of chance corrected case-by-case agreement between PR and InterVA, KA-HDSS, 2014

Probable COD	Agreement (%)	Kappa [95 % CI]
Broad causes of death		
Accidents/injuries	95.0	0.75 [0.65–0.85]
Communicable diseases	71.7	0.39 [0.30–0.48]
Perinatal causes	94.0	0.51 [0.35–0.67]
Non-communicable diseases	71.5	0.37 [0.28–0.46]
Overall	56.5	0.40 [0.39–0.45]
Specific causes of death		
Tuberculosis	90.6	0.59 [0.47–0.70]
Cardiovascular diseases	86.4	0.23 [0.1–0.37]
Neoplasms	88.5	0.27 [0.13–0.41]
Acute lower respiratory tract infections	91.0	0.39 [0.24–0.54]
Pneumonia/sepsis	95.9	0.55 [0.36–0.74]
Perinatal causes	97.0	0.59 [0.39–0.79]
Agreement by age group and sex		
Age		
Under five	44.4	0.35 [0.28–0.41]
15–45 years	38.0	0.33 [0.31–0.34]
45 years and older	32.4	0.26 [0.22–0.29]
Sex		
Male	39.6	0.34 [0.34–0.37]
Female	36.2	0.31 [0.20–0.35]

PR: physician review, COD: cause of death

the existing knowledge of the burden of communicable diseases in Ethiopia [2, 11]. Both methods were similar in attributing the proportion of TB, pneumonia/sepsis, acute infections, malaria and HIV/AIDS. In a similar study in Kenya, the two methods comparably attributed pneumonia/sepsis, TB, malaria and meningitis [10]. However, according to other studies, the InterVA overestimated TB than the physician review [10, 13, 15]. The comparability of both methods in the number of times they diagnosed HIV/AIDS was inconsistent. In some studies, the InterVA diagnosed HIV/AIDS more frequently than physician review [10], while less frequently than PR [13] in other study. This discrepancy may be related to misclassification of HIV/AIDS and TB, which is reported in several studies [10, 22, 29, 30].

In our study, both methods attributed NCDs comparably and the magnitude of the estimate accords with previous findings [31–33]. Similarly, cardiovascular causes of death were comparably estimated by both methods, which was also reported in another study [15]. The consistency of both methods in estimating deaths attributed to accidents/injuries shown in our study concurs with that of other studies [10, 15]. This may be related to the clarity of the indicators (signs

and symptoms) reported for accidents and injuries than other causes.

In our study, the overall chance corrected agreement, at broad and specific causes of death categories, falls between 0.21 and 0.40, which is considered as a fair agreement [28]. The case-by-case agreement at specific cause of death level was higher than a similar study in Kenya ($\kappa = 0.27$) and lower than another findings from Ethiopia ($\kappa = 0.49$) and Kenya ($\kappa = 0.42$) [10, 13, 15]. As reported in a similar study, the level of agreement was better in younger ages than the older age groups [34]. This could be explained in terms of the difference in epidemiology of causes of diseases across age groups. Older age groups experience multiple illness conditions with overlapping symptomatic nature than younger groups [34].

Findings from several [10, 13–15], but not all [26], studies show that the concordance level between the PR and InterVA is insufficient. A comparative study in Northwest Ethiopia, which included 408 adult deaths, measured a concordance level of 0.49 on broad CoD level [13]. Even much lower levels of agreement were also reported from the African Population and Health Research Center ($k = 0.27$) [10] and Kilifi Health Demographic Surveillance System ($k = 0.32$) [15], both in Kenya, which did similar comparison. On the other hand, finding from a recent multi-center study, which used data from Health and Demographic Surveillance systems, and Health and Demographic Surveys, showed an almost perfect level of agreement, reporting overall concordance correlation coefficient of 0.83 [26].

In the present study, inference about validity of either of the methods can not be made in the absence of a gold-standard diagnosis. However, validation studies which simultaneously evaluated PR and InterVA methods against hospital certified deaths, showed that the PR performs better than the InterVA model [14, 15]. A validation study which compared both the InterVA and PR methods against hospital CoD revealed that the level of agreement between InterVA and hospital CoD ($\kappa = 0.32$) was lower than the agreement between physician review and hospital CoD ($k = 0.52$). In addition, in another study which evaluated the PR and InterVA using clinical diagnostic gold standards in a sample of 12,542 verbal autopsy cases, the PR has shown a better performance than the InterVA, across all age-groups [14].

Discrepancy in the diagnosis between these two methods may not be unexpected, though further investigation is needed to explain the variation. Nevertheless, according to previous studies the discordance in diagnosis was related to a variation on how the two methods process and use the verbal autopsy data. The InterVA uses the data from the closed ended questions only, while the PR involve extensive use of the open ended narrative part of the VA data [3, 9, 10, 14]. In addition,

the InterVA use a probability matrix to process the indicators in the verbal autopsy data, while the PR is based on expert judgment [7, 9].

In addition to the minimal effort it requires, the InterVA has a comparative advantage of being completely internally consistent that enables producing comparable outputs. In contrary, PR is labor intensive, and prone to inter-observer variation. However, it has also some benefits. As a part of their routine clinical practice, reviewer physicians treat patients who come from the same population where the VA cases come from. This gives reviewer physicians a chance to correlate the signs and symptoms used to describe illness in the specific community with the actual illness confirmed through clinical investigations. Although, such prior knowledge can affect the possibility of coding less prevalent causes [9], it may help the PR process to be a robust on CoD which are common in the community.

The present study has the following limitations. The two methods were compared in the absence of a gold-standard diagnosis. As a result, it was possible to conclude about the validity of the methods. Although the study included more cases than the minimum sample size required, it was not sufficient when it comes to comparing sub-groups or rare causes of death.

Conclusion

In summary, this study reported an overall low chance corrected agreement in probable cause of death between PR and InterVA. The level of agreement varies across different categories of causes of death, and age of the deceased. The agreement ranged from moderate to substantial for important public health diseases like TB, perinatal causes, pneumonia/sepsis, and accidents and injuries; while the agreement for NCDs, especially for cardiovascular causes and neoplasms was low. Both methods showed a relatively better agreement in under-five children and adults aged 15-45, while they least agreed for cases aged 45 and above years. Therefore, if the InterVA were used in place of the PR process, the overall diagnosis would be fairly similar.

Additional file

Additional file 1: Categorization of causes of death from the physicians review and InterVA model, KA-HDSS, 2014.

Competing interests

The authors declare that they have no competing interests.

Authors' contributions

BW and YAM involved in design of the surveillance, data collection and supervision. BW was involved in study conception, data processing and analysis. BW wrote the manuscript and interpreted the results. BW, YAM, GJD, and MS made substantial revision to the manuscript. MS and GJD mentored the process of paper writing from preparation to manuscript write up. All authors read and approved the final manuscript.

Acknowledgements

The authors are thankful to the funding organizations, field workers and data management staff. We are grateful to Prof. Peter Byass from WHO Collaborating Centre for Verbal Autopsy, Umeå Centre for Global Health Research, Umeå University, Sweden, for providing training on the application of InterVA to the KA-HDSS research team.

Disclaimer

The KA-HDSS is financially supported by Mekelle University and the US Centers for Disease Control and Prevention (CDC), in accordance with the EPA-CDC Cooperative Agreement No.5U22/PS022179_10. The INDEPTH Network has also provided technical support. However, contents of this publication are solely the responsibility of the authors and do not necessarily represent the official views of the funding organizations.

Author details

¹Department of Public Health, Mekelle University, Mekelle, Ethiopia. ²Centre of Cardiovascular Research and Education in Therapeutics, Department of Epidemiology and Preventive Medicine, Monash University, 99 Commercial Road, Melbourne, VIC 3004, Australia. ³CAPHRI, School for Public Health and Primary Care, Maastricht University, Maastricht, Netherlands.

Received: 21 March 2015 Accepted: 7 July 2015

Published online: 15 July 2015

References

- Byass P. A new era dawns for death registration. *PLoS Speaking of Medicine*. Available from: [http://blogs.plos.org/speakingofmedicine/2012/08/22/a-new-era-dawns-for-death-registration/] [cited 10 June 2014].
- Weldearegawi B, Ashebir Y, Gebeye E, Gebregziabher T, Yohannes M, Mussa S, et al. Emerging chronic non-communicable diseases in rural communities of Northern Ethiopia: evidence using population-based verbal autopsy method in Kilite Awlaleo surveillance site. *Health Policy Plan*. 2013;28(8):891–8.
- Vergnano S, Fottrell E, Osrin D, Kazembe PN, Mwansambo C, Manandhar DS, et al. 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*. 2011;9:48.
- Byass P, Kahn K, Fottrell E, Collinson MA, Tollman SM. Moving from data on deaths to public health policy in Agincourt, South Africa: approaches to analysing and understanding verbal autopsy findings. *PLoS Med*. 2010;7(8):e1000325. doi:10.1371/journal.pmed.1000325.
- Setel PW, Sankoh O, Rao C, Velkoff VA, Mathers C, Gonghuan Y, et al. Sample registration of vital events with verbal autopsy: a renewed commitment to measuring and monitoring vital statistics. *Bull World Health Organ*. 2005;83:611–7.
- King G, Lu Y, Shibuya K. Designing verbal autopsy studies. *Popul Health Metrics*. 2010;8:19.
- Byass P, Chandramohan D, Clark SJ, D'Ambruoso L, Fottrell E, Graham WJ et al. Strengthening standardized interpretation of verbal autopsy data: the new InterVA-4 tool. *Glob Health Action*. 2012;5:19281.
- Lulu K, Berhane Y. The use of simplified verbal autopsy in identifying causes of adult death in a predominantly rural population in Ethiopia. *BMC Public Health*. 2005;5:58. doi:10.1186/1471-2458-5-58.
- Fottrell E, Byass P. Verbal autopsy: methods in transition. *Epidemiol Rev*. 2010;32:38–55.
- Oti S, Kyobutungi C. 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*. 2010;8:21.
- Misganaw A, Haile Mariam D, Araya T, Aneneh A. Validity of verbal autopsy method to determine causes of death among adults in the urban setting of Ethiopia. *BMC Med Res Methodol*. 2012;12:130.
- Fantahun M, Fottrell E, Berhane Y, Wall S, Hogberg U, Byass P. Assessing a new approach to verbal autopsy interpretation in a rural Ethiopian community: the InterVA model. *Bull World Health Organ*. 2006;84:204–10.
- Tadesse S. Validating the InterVA model to estimate the burden of mortality from verbal autopsy data: a population-based cross-sectional study. *PLoS One*. 2013;8(9):e73463. doi:10.1371/journal.pone.0073463.
- Lozano R, Freeman M, James SL, Campbell B, Lopez AD, Flaxman AD, et al. Performance of InterVA for assigning causes of death to verbal autopsies: multisite validation study using clinical diagnostic gold standards. *Popul Health Metrics*. 2011;9:50.

15. Bauni E, Ndila C, Mochamah G, Nyutu G, Matata L, et al. Validating physician-certified verbal autopsy and probabilistic modeling (InterVA) approaches to verbal autopsy interpretation using hospital causes of adult deaths. *Popul Health Metrics*. 2011;9:49.
16. INDEPTH Network. <http://www.indepth-network.org/>.
17. Weldearegawi B, Spigt M, Berhane Y, Dinant G. Mortality Level and Predictors in a Rural Ethiopian Population: Community Based Longitudinal Study. *PLoS One*. 2014;9(3):e93099. doi:10.1371/journal.pone.0093099.
18. Melaku YA, Sahle BW, Tesfay FH, Bezabih AM, Aregay A, Abera SF, et al. Causes of Death among Adults in Northern Ethiopia: Evidence from Verbal Autopsy Data in Health and Demographic Surveillance System. *PLoS One*. 2014;9(9):e106781. doi:10.1371/journal.pone.0106781.
19. Melaku YA, Weldearegawi B, Aregay A, Tesfay FH, Abreha L, Abera SF, et al. Causes of death among females—investigating beyond maternal causes: a community-based longitudinal study. *BMC Res Notes*. 2014;7:629.
20. World Health Organization. Verbal autopsy standards: The 2012 WHO verbal autopsy instrument. Geneva: World Health Organization 2012 [Available from: http://www.who.int/healthinfo/statistics/WHO_VA_2012_RC1_Instrument.pdf] [cited 10 June 2013].
21. SAVVY: Sample Vital Registration with Verbal Autopsy. International Verbal Autopsy Form 1-3. [Available from: <http://www.cpc.unc.edu/measure/tools/monitoring-evaluation-systems/savvy> (cited 10 June 2013)].
22. World Health Organization: International Classification of Diseases (ICD). [Available from: <http://www.who.int/classifications/icd/en/>] [cited 10 June 2013].
23. InterVA. [Available from: <http://www.interva-4.net/>] (cited 29 September 2014).
24. Jima D, Getachew A, Bilak H, Steketee RW, Emerson PM, Graves PM, et al. Malaria indicator survey, Ethiopia: coverage and use of major malaria prevention and control interventions. *Malar J*. 2010;9:58.
25. Central Statistical Agency [Ethiopia] and ICF International. Ethiopia demographic and health survey 2011. Addis Ababa, Ethiopia and Calverton, Maryland, USA: Central Statistical Agency and ICF International; 2012.
26. Byass P, Herbst K, Fottrell E, Ali MM, Odhiambo F, Amek N, et al. Comparing verbal autopsy cause of death findings as determined by physician coding and probabilistic modelling: a public health analysis of 54,000 deaths in Africa and Asia. *J Glob Health*. 2015;5:010402.
27. Reichenheim ME. Sample size for the kappa-statistic of interrater agreement. *Stata Tech Bull*. 2001;10:58.
28. Viera AJ, Garrett JM. Understanding Inter-observer agreement: The Kappa Statistic. *Fam Med*. 2005;37(5):360–3.
29. Keshinro B, Diul MY. HIV-TB: epidemiology, clinical features and diagnosis of smear-negative TB. *Trop Doct*. 2006;36(2):68–71.
30. Abdool Karim SS, Churchyard GJ, Abdool Karim Q, Lawn SD. HIV infection and tuberculosis in South Africa: an urgent need to escalate the public health response. *Lancet*. 2009;374:921–33.
31. Misganaw A, Haile Mariam D, Araya T, Ayele K. Patterns of mortality in public and private hospitals of Addis Ababa, Ethiopia. *BMC Public Health*. 2012;12:1007.
32. Prevett M. Chronic non-communicable diseases in Ethiopia a hidden burden. *Ethiop J Health Sci*. 2012;22:2.
33. Mamo Y, Seid E, Adams SS, Gardiner A, Parry E. A primary healthcare approach to the management of chronic disease in Ethiopia: an example for other countries. *Clin Med*. 2007;7:3.
34. Tadesse S. Agreement between physicians and the InterVA-4 model in assigning causes of death: the role of recall period and characteristics specific to the deceased and the respondent. *Arch Public Health*. 2013;71:28.

Submit your next manuscript to BioMed Central and take full advantage of:

- Convenient online submission
- Thorough peer review
- No space constraints or color figure charges
- Immediate publication on acceptance
- Inclusion in PubMed, CAS, Scopus and Google Scholar
- Research which is freely available for redistribution

Submit your manuscript at
www.biomedcentral.com/submit

