

ORIGINAL ARTICLE

Refining a probabilistic model for interpreting verbal autopsy data

PETER BYASS¹, EDWARD FOTTRELL¹, DAO LAN HUONG², YEMANE BERHANE³,
TUMANI CORRAH⁴, KATHLEEN KAHN⁵, LULU MUHE⁶ & DO DUC VAN⁷

¹Umeå International School of Public Health, Umeå University, Umeå, Sweden, ²Health Strategy and Policy Institute, Ministry of Health, Hanoi, Vietnam, ³Department of Community Health, Addis Ababa University, Addis Ababa, Ethiopia, ⁴Medical Research Council Laboratories, Banjul, The Gambia, ⁵School of Public Health, University of the Witwatersrand, Johannesburg, South Africa, ⁶World Health Organization, Geneva, Switzerland, and ⁷Viet Duc Surgical Hospital, Hanoi, Vietnam

Abstract

Objective: To build on the previously reported development of a Bayesian probabilistic model for interpreting verbal autopsy (VA) data, attempting to improve the model's performance in determining cause of death and to reassess it. **Design:** An expert group of clinicians, coming from a wide range geographically and in terms of specialization, was convened. Over a four-day period the content of the previous probabilistic model was reviewed in detail and adjusted as necessary to reflect the group consensus. The revised model was tested with the same 189 VA cases from Vietnam, assessed by two local clinicians, that were used to test the preliminary model. **Results:** The revised model contained a total of 104 indicators that could be derived from VA data and 34 possible causes of death. When applied to the 189 Vietnamese cases, 142 (75.1%) achieved concordance between the model's output and the previous clinical consensus. The remaining 47 cases (24.9%) were presented to a further independent clinician for reassessment. As a result, consensus between clinical reassessment and the model's output was achieved in 28 cases (14.8%); clinical reassessment and the original clinical opinion agreed in 8 cases (4.2%), and in the remaining 11 cases (5.8%) clinical reassessment, the model, and the original clinical opinion all differed. Thus overall the model was considered to have performed well in 170 cases (89.9%). **Conclusions:** This approach to interpreting VA data continues to show promise. The next steps will be to evaluate it against other sources of VA data. The expert group approach to determining the required probability base seems to have been a productive one in improving the performance of the model.

Key Words:

Introduction

Verbal autopsy (VA) is the process of eliciting information about the circumstances of a death from family or friends of the recently deceased person in cases where medical certification of death is incomplete or absent [1–4]. It is a useful surrogate for routine death registration in resource-poor settings and has been used to estimate cause-specific mortality [5]. Physician review of VA data, whereby data are assessed by one or more physicians who assign probable cause of death, has been shown to be a reliable method for VA interpretation [2].

However, issues regarding standardization between different physicians and the risk of having to rely on physicians over time hinder reliable temporal and regional comparisons of mortality [6]. In addition, the time that physicians must devote to assessing large numbers of VAs is far from ideal in areas with insufficient medical personnel. Algorithms have the potential to address these concerns [7] but raise others, such as reliability and the difficulty of considering parallel possibilities along the lines of classic clinical differential diagnoses.

A preliminary model for VA interpretation based on Bayes's theorem was developed in an attempt to

Correspondence: Peter Byass, Umeå International School of Public Health, Umeå University, SE-901 85 Umeå, Sweden. E-mail: peterbyass@aol.com

(Received ??; accepted ??)

ISSN 1403-4948 print/ISSN 1651-1905 online/05/010001-6 © 2005 Taylor & Francis Ltd
DOI: 10.1080/14034940510032202

overcome the weaknesses of physician review and algorithmic approaches. Bayes's theorem seeks to define the probability of a cause (C) given the presence of a particular indicator (I), and can be represented as:

$$P(C/I) = \frac{P(I/C) \times P(C)}{P(I/C) \times P(C) + P(I/!C) \times P(!C)}$$

where P(!C) is the probability of not (C).

The probability of occurrence of each indicator ($I_1 \dots I_n$) and each possible cause of death ($C_1 \dots C_m$) can be determined at the population level, which in this case means among all deaths. Thus, for a particular case, the probability of C_k is initially the value found among all deaths in general. However, for each case and each applicable indicator, the probability of C_k can be modified by the above theorem. The VA interpretation model adjusts the probability of each possible cause according to a matrix of $P((I_1 \dots I_n)/(C_1 \dots C_m))$ and lists up to three likely causes. In the preliminary model the set of indicators and causes was influenced by Indepth's proposed VA questionnaire [8], and the associated probabilities were estimates based on accumulated personal experience.

Initial validation of the preliminary model was carried out on a set of 189 VAs from rural Vietnam, which had previously been assessed by two physicians leading to a consensus on a single cause of death for each case. Over 70% of individual causes of death corresponded with those determined by the physicians, increasing to over 80% when cases ascribed to "old age" or "indeterminate" by the physicians were excluded. A more detailed background to the preliminary model and its initial validation are described elsewhere [6].

Following validation of the preliminary model it was deemed appropriate to refine the probabilities used in the model and address underlying conceptual issues of VA data collection and interpretation. This paper describes development of this probabilistic approach to VA interpretation using an expert Delphi technique. Validation of the updated and refined model on the same 189 cases from Vietnam and a comparison between the performance of the preliminary and updated models is also described.

Material and methods

The Delphi technique is an approach used to gain consensus among a panel of experts in order to address a lack of agreement or incomplete state of knowledge [9,10]. The technique was adopted here

to develop consensus on probabilities of different causes of death occurring at the population level and probabilities of specific signs and symptoms presenting themselves at the population level and in specific causes of death. The technique was also utilized to develop consensus on key conceptual issues of cause of death classification and VA usage.

An expert group convened over four consecutive days. The group comprised five physicians (YB, TC, KK, LM, DDV) with extensive clinical experience in resource-poor settings. They represented a range of important disciplines of medicine: surgery; maternal and reproductive health; paediatrics; and internal medicine. The experts came from a range of settings in developing or transitional countries where routine death registration is often absent (South Africa, Ethiopia, The Gambia, and Vietnam). It was felt that the range of backgrounds and geographical spread of the expert group would lead to a generalized consensus not specific to any one region or medical discipline. Each member of the expert group was either experienced in or very familiar with the process, importance and limitations of VAs and all were briefed on the probabilistic approach to VA interpretation.

The researchers facilitated discussions in which the experts were requested to consider the inclusion of indicators and causes of death in the model, bearing in mind that friends or relatives of the deceased person must be able to notice and report indicators to the lay fieldworkers [5]. A list of 34 possible causes of death and 104 indicators was developed (Table I). Probabilities were agreed upon and assigned to each indicator and cause of death at the population level and for each specific cause of death using on a semi-qualitative scale following work by Kong et al. (1986) [11] (Table II). A higher degree of precision was not sought since previous work suggests that this is not essential in order to build a workable model [12].

There was strong consensus among the physicians that probabilities of causes of death with large variations in prevalence at the population level between regions, such as HIV/AIDS and malaria, should have the possibility of being adjusted in the model to reflect the local burden of these diseases at the population level. To warrant adjustment of the database it was felt that regional variations of disease prevalence should be at least tenfold. It was not felt necessary to adjust the database to reflect regional variations in causes of death with very specific indicators, such as meningitis or transport accidents. The revised model therefore included a facility to reflect either high or low prevalence for HIV/AIDS and malaria.

Table I. Verbal autopsy indicators and causes of death used in the refined model.

	Indicators	Causes
Was this an elder 65+ years	Any chronic/recurrent diarrhoea (4+ weeks)	Disease of nervous system
Was this an adult 50–64 years	Any abdominal swelling	Other fatal accident
Was this a female 15–49 years	Any vomiting	Transport-related accident
Was this a male 15–49 years	Any yellowness/jaundice	Accidental drowning
Was this a child 5–14 years	Any abnormality of urine	Accidental poisoning
Was this a child 1–4 years	Any urinary retention	Suicide
Was this an infant 4 weeks–1 yr	Any haematuria	Homicide
Was this a neonate <4 weeks	Any swelling of ankles/legs	Acute cardiac death
Was she pregnant at death	No bilateral swelling of ankle	Chronic cardiac death
Did pregnancy end within 6 weeks	Any skin lesions/ulcers	Stroke
Did final illness last at least 3 weeks	Any rash	Non-bloody diarrhoea
Did final illness last <3 weeks	Any herpes zoster	HIV/AIDS related death
Was death very sudden or unexpected	Any excessive night sweats	Other acute infection
Was death during wet season	Any excessive water intake	Malaria
Was death during dry season	Any excessive urination	Measles
Was she/he in a transport accident	Any excessive food intake	Tetanus
Did she/he drown	Any acute fever	Bloody diarrhoea
Had she/he fallen recently	Any persistent fever (>2 weeks)	Meningitis
Any poisoning, bite, sting	Any enlarged/swollen glands	Other chronic infection
Was she/he a known smoker	Any facial swelling	Sickle cell disease
Any obvious recent injury	Was there a coma >24 h	Other digestive disease
Was she/he known to drink alcohol	Any weight loss	Liver disease
Any suggestion of homicide	Any anaemia/paleness	Malnutrition
Any convulsions or fits	Any drowsiness	Diabetes
Any diagnosis of epilepsy	Any delayed/regressed development	Kidney or urinary disease
Was the fontanelle raised	Any diagnosis of asthma	Acute respiratory disease, not
Was the fontanelle or eyeball sunken	Any diagnosis of diabetes	Pneumonia
Any headache	Any diagnosis of heart disease	Tuberculosis (pulmonary)
Was there paralysis on both sides	Any diagnosis of HIV/AIDS	Pneumonia
Any paralysis/weakness on 1 side	Any diagnosis of hypertension	Chronic respiratory disease
Any stiff neck	Been discharged from hospital very ill	Malignancy
Any oral candidiasis	Any suggestion of suicide	Maternity-related death
Any rigidity/lockjaw	Any surgery just before death	Pre-term/small baby
Any coughing with blood	Any diagnosis of TB	Perinatal asphyxia
Any chest pain	Was she/he adequately vaccinated	Congenital malformation
Was there a cough for >3 weeks	Any diagnosis of liver disease	
Was there a cough for up to 3 weeks	Any diagnosis of cancer	
Any productive cough	Any diagnosis of stroke	
Any rapid breathing	Any diagnosis of measles	
Any breathlessness on exertion	Any diagnosis of kidney disease	
Any breathlessness lying flat	Any diagnosis of sickle-cell disease	
Any chest indrawing	Any diagnosis of malaria	
Any difficulty breathing	Any delivery complications	
Any breast lump or lesion	Any heavy bleeding before/after delivery	
Any wheezing	Was there prolonged labour >24 h	
Any cyanosis	Were there convulsions during delivery	
Any abdominal mass	Was the baby born early <34 weeks	
Any abdominal pain	Was the baby small <2,500 g	
Any diarrhoea with blood	Was there difficulty breathing at birth	
Any vomiting with blood	Any congenital malformations	
Any acute diarrhoea (<2 weeks)	Was this a multiple birth	
Any persistent diarrhoea (2–4 weeks)	Any umbilical infection	

The model was updated using Visual FoxPro database software to make adjustments to probabilities and removal or insertion of various causes and indicators. The revised model's output was modified to show only more than one cause of death if the probability of the additional cause(s) was within

20% of the most likely cause. This is in contrast to the preliminary model, which always gave the three most likely causes irrespective of probabilities. The model was also adjusted so that certain causes of death were extremely unlikely to be diagnosed without the presence of specific indicators. For

Table II. The semi-qualitative scale used for assigning probabilities of indicators and causes in the refined model.

Qualitative descriptor	Description	Approximate quantitative equivalent (%)
1	Almost always	100
A	Frequently	50
A-		20
B+	Moderately often	10
B		5
B-		2
C+	Uncommon	1
C		0.5
C-		0.2
0	Virtually never	<0.1
N	Absolutely never	0

example, it is thus highly unlikely that the model will conclude that death resulted from diarrhoeal disease without the symptom of diarrhoea being reported. Each member of the expert group was provided with a working prototype of the model and given the opportunity to test it on hypothetical cases to highlight any inconsistencies and anomalies.

The updated probabilistic model was applied to the VA data from the same 189 Vietnamese cases used to validate the preliminary model. Indicators were gathered from the original VA questionnaires and included open-ended, free-text information. These data and the underlying VA process used in Vietnam are described in detail elsewhere [3]. Comparisons were made with the cause of death as previously agreed by the two local physicians in Vietnam and with the results from the preliminary model.

Many studies aiming to validate VA interpretation methodologies against hospital records or physician review describe sensitivities, specificities, and positive predictive values (PPV) [1,13,14]. However, the calculation of such statistics assumes that the referent diagnosis gives the right answer and is an absolute gold standard. This assumption is flawed due to the inconsistencies of physician review and studies describing sensitivity, specificity, and PPV of VA methods often discuss the possibility that in certain cases the VA diagnosis may be more accurate than the diagnosis provided by physician review or hospital records [13,14]. As such, it was considered inappropriate to calculate the sensitivities, specificities, or PPV for the probabilistic model in this validation study. Instead, kappa (κ) values are calculated since they simply reflect the level of agreement between the two methods and do not imply superiority of one method over the other.

Results

In 142/189 cases (75.1%) the cause of death as determined by the refined model agreed with the consensus of the two original assessing physicians ($\kappa=0.50$ (95% CI 0.42–0.59)). In a number of the indeterminate and contradictory cases it was not always clear why the physicians' conclusion was more appropriate than that of the model. Therefore, the remaining 47 cases (24.9%) were presented to a further experienced clinician, who was involved in neither the original assessment nor the model's development, for reassessment. As a result, 28 cases (59.6%) arrived at consensus between clinical reassessment and the model's output ($\kappa=0.8$ (95% CI 0.74–0.86)); 8 cases (4.2%) arrived at consensus between clinical reassessment and the original clinical opinion ($\kappa=0.66$ (95% CI 0.51–0.81)), and in the remaining 11 cases (5.8%) clinical reassessment, the model and the original clinical opinion all differed. Thus overall the model was considered to have performed well in 170 cases (89.9%). This shows a substantial improvement compared with the preliminary model where 134/189 cases (70.9%) were in agreement ($\kappa=0.42$ (95% CI 0.33–0.51)), 34 cases (18.0%) were indeterminate, and 21 cases (11.1%) were contradictory.

Figure 1 shows cause-specific mortality fractions for the 189 deaths separately derived from the most likely causes from the refined probabilistic model, weighted multiple causes from the refined model (e.g. assigning one-third of a death to each of three likely causes), the original physicians' verdict, the physicians' verdict following reassessment of 47 cases by a third independent physician, and the most likely cause from the preliminary, unrefined model.

Discussion

The results from comparisons of the model's output with that of the physicians is very encouraging given that development of the model and the underlying probabilities were not specifically linked to the Vietnamese VA process or setting. The improved performance of the refined model when compared with the preliminary model illustrates the effectiveness of the Delphi approach utilized.

The development of this approach to VA interpretation has highlighted many of the unanswered questions around the whole process of VA data collection and interpretation. For example, not all indicators available in the data were built into the model, and not all indicators built into the model are routinely available in the data. Such mismatches

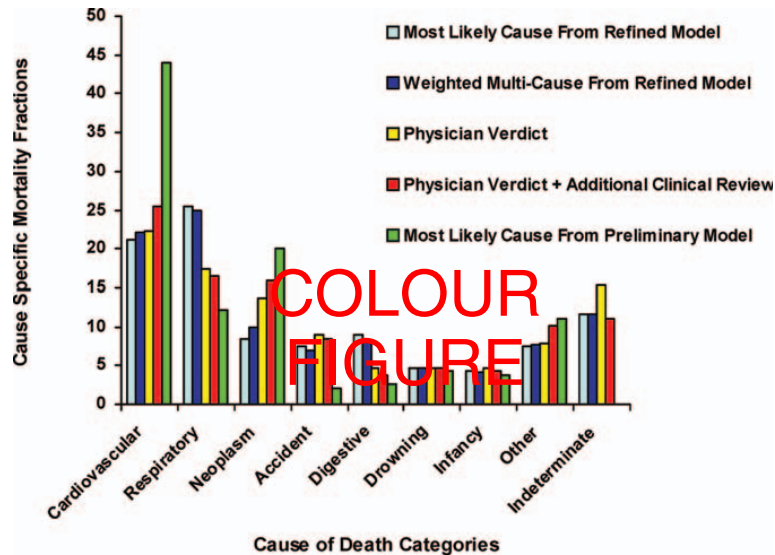


Figure 1. Cause-specific mortality fractions for major causes of death, derived from 189 verbal autopsies in Vietnam, according to the most likely causes from the refined probabilistic model, weighted multiple causes from the refined model, the original physicians’ verdicts, the physicians’ verdicts following additional clinical review, and the preliminary probabilistic model.

may have reduced the model’s overall performance and there is scope for development of more standardized VA data collection tools. Other issues include variations in concepts and definitions of “acute” and “chronic” between physicians and regions, and difficulties in determining the sequence of events and identifying immediate versus antecedent and underlying causes of death from VA data. With regard to the latter point, it was felt that identification of the deceased’s primary complaint or most prominent indicator before death could overcome some of the issues of sequencing events. It may be possible to include questions of this nature in VA questionnaires and introduce composite indicators into the model.

All validations ideally require a suitable “gold standard” for comparison. Although physician interpretation is often considered the gold standard for VA interpretation there is potential for misclassification and misinterpretation [2,15]. This is highlighted by the fact that only 8/47 cases (17%) presented to a third physician for reassessment reached consensus with the original clinical opinion in this study. As such, physician interpretation of VA data must be used cautiously as a gold standard for validations [4]. It was therefore considered more appropriate in this study to compare the probabilistic approach with physician review in terms of agreement rather than sensitivity, specificity or PPV.

The expert group felt that “old age” should not be allowed as an acceptable diagnosis of cause of death and it was decided to eliminate “non-specific old age” as a possible cause from the model. Although

the usefulness of gathering specific cause of death data in elderly people where there is little notion or possibility of implementing interventions is questionable, mortality data for the elderly have the potential to be useful in evaluating interventions implemented when the “old” generation was the “middle aged” generation. Furthermore, there are considerable cultural and regional variations in the concept of old age. As such, cases where the physicians’ verdict was “old age” were considered as indeterminate during the validation process. For the purposes of population health surveillance it would be useful to have standardized age categories across Indepth sites and indeed globally, but since that is not the case the expert group settled on age groups that made sense from a clinical perspective.

Adjusting the model’s database to reflect local conditions of malaria endemicity and HIV/AIDS prevalence worked well in improving the performance of the model. This highlights the potential of using this probabilistic approach to standardized VA interpretation across a range of settings. Further testing of the refined model with more data from a wider range of settings may identify other key local characteristics that can be reflected in the model. Based on these principles it may also be possible to develop databases for the analysis of VA data relating to specific sub-categories, such as specific age groups and maternal deaths. This may provide more specific details of causes of death in populations that have significant potential for public health interventions [16].

Based on the principle of adjusting for prevalence it may also be possible or necessary to adjust

prevalence of diseases based on successful intervention programmes. However, knowing the effectiveness of interventions is difficult given the lack of reliable data in settings where the VA model will be applied. It is therefore unrealistic to make such adjustments at present.

Assuming that the VA process is intended to mimic as far as possible the process of physician death certification, the innovative probabilistic approach to VA interpretation described here represents an improvement to current interpretation methods. The identification of multiple causes of death mimics the classical physician assessment (based on differential diagnosis) and has the added potential to facilitate cause of death surveillance at the community level where unsubstantiated choices between possible causes of death at the individual level are often made. Further thought and discussion is needed about how to interpret and analyse multiple causes of death for individual cases. One suggestion is that the death can be divided proportionately between different causes. For example, if the model lists chronic heart failure and pneumonia as two likely causes of death it might be reasonable to assign 50% of that death to each cause. This does not seem to greatly affect CSMFs as illustrated in Figure 1 and may in fact be more useful from an epidemiological perspective than making arbitrary choices between different likely individual causes.

Standardization over time and between regions is a major advantage of this approach to VA interpretation. The benefits of 100% standardization may justify a trade off against 100% accuracy whereby physician review is used as the somewhat flawed gold standard. In addition, the probabilistic model allows the interpretation of VA cases at a rate of approximately two per second (having captured the VA data) and does not require extensive expertise to operate it, thereby greatly increasing efficiency and freeing up the time of physicians. The next step in the development of this system is to test the refined model with more extensive data from a wider range of settings and sources.

References

- [1] Chandramohan D, et al. Verbal autopsies for adult deaths: Their development and validation in a multicentre study. *Trop Med Int Health* 1998;3:436–446.
- [2] Quigley MA, Chandramohan D, Rodrigues LC. Diagnostic accuracy of physician review, expert algorithms and data-derived algorithms in adult verbal autopsies. *Int J Epidemiol* 1999;28:1081–7.
- [3] Huong DL, Minh HV, Byass P. Applying verbal autopsy to determine cause of death in rural Vietnam. *Scand J Public Health* 2003;31(Suppl 62):19–25.
- [4] Chandramohan D, Setel P, Quigley M. Effect of misclassification of cause of death in verbal autopsy: Can it be adjusted? *Int J Epidemiol* 2001;30:509–14.
- [5] Okosun IS, Dever GE. Verbal autopsy: A necessary solution for the paucity of mortality data in the less-developed countries. *Ethnicity Disease* 2001;11:575–7.
- [6] Byass P, Huong DL, Minh HV. A probabilistic approach to interpreting verbal autopsies: Methodology and preliminary validation in Vietnam. *Scand J Public Health* 2003;31(Suppl 62):32–7.
- [7] Reeves BC, Quigley MA. A review of data-derived methods for assigning causes of death from verbal autopsy data. *Int J Epidemiol* 1997;26:1080–9.
- [8] Indeph, Population and health in developing countries, Vol.1: Population, health and survival, editor I. Network. Ottawa: IDRC; 2002.
- [9] Keeney S, Hasson F, McKenna HP. A critical review of the Delphi technique as a research methodology for nursing. *Int J Nurs Stud* 2001;38:195–200.
- [10] Powell C. The Delphi technique: Myths & realities. *J Adv Nurs* 2003;41:376–82.
- [11] Kong A, et al. How medical professionals evaluate expressions of probability. *N Engl J Med* 1986;315:740–4.
- [12] Byass P, Corrah PT. Assessment of a probabilistic decision support methodology for tropical health care. In: *Medinfo-89*. Amsterdam: North Holland, 1989.
- [13] Kahn K, et al. Validation and application of verbal autopsies in a rural area of South Africa. *Trop Med Int Health* 2000;5(11):824–831.
- [14] Chandramohan D, et al. The validity of verbal autopsies for assessing the causes of institutional maternal death. *Stud Fam Plann* 1998;29(4):414–422.
- [15] Anker M. The effect of misclassification error on reported cause-specific mortality fractions from verbal autopsy. *Int J Epidemiol* 1997;26:1090–6.
- [16] Høj L, Stensballe J, Aaby P. Maternal mortality in Guinea-Bissau: The use of verbal autopsy in a multi-ethnic population. *Int J Epidemiol* 1999;28:70–6.