A probabilistic approach to interpreting verbal autopsies: methodology and preliminary validation in Vietnam

Peter Byass[1], Dao Lan Huong[2] and Hoang Van Minh[3]

[1]Umeå International School of Public Health, Umeå University, Umeå, Sweden, [2]Health Strategy and Policy Institute, Ministry of Health, Hanoi, Vietnam, [3]Faculty of Public Health, Hanoi Medical University, Hanoi, Vietnam

Scand J Public Health 2003; 31(Suppl. 62): 32–37

Aims: Verbal autopsy (VA) has become an important tool in the past 20 years for determining cause of death in communities where there is no routine registration. In many cases, expert physicians have been used to interpret the VA findings and so assign individual causes of death. However, this is time consuming and not always repeatable. Other approaches such as algorithms and neural networks have been developed in some settings. This paper aims to develop a method that is simple, reliable and consistent, which could represent an advance in VA interpretation. *Methods*: This paper describes the development of a Bayesian probability model for VA interpretation as an attempt to find a better approach. This methodology and a preliminary implementation are described, with an evaluation based on VA material from rural Vietnam. *Results*: The new model was tested against a series of 189 VA interviews from a rural community in Vietnam. Using this very basic model, over 70% of individual causes of death corresponded with those determined by two physicians, increasing to over 80% if those cases ascribed to old age or as being indeterminate by the physicians were excluded. *Discussion*: Although there is a clear need to improve the preliminary model and to test more extensively with larger and more varied datasets, these preliminary results suggest that there may be good potential in this probabilistic approach.

Key words: verbal autopsy, Bayesian, probability, cause of death, Vietnam, mortality.

Peter Byass, Umeå International School of Public Health, Umeå University, SE-901 85 Umeå, Sweden. Fax: +46 90 138977, e-mail: peterbyass@aol.com

INTRODUCTION

Verbal autopsy (VA) – the process of interviewing family or friends of recently deceased people to elicit information on the circumstances of death – has emerged over the last 20 years as an important tool for estimating cause-specific mortality in settings where death registration is not routinely undertaken by physicians (1). It provides an opportunity for health planners and policy makers, as well as epidemiologists, to better understand patterns of mortality and their health implications. A lot of work has been put into the process of the VA interview and developing appropriate questionnaires (2), but it is also important to have good methods of interpreting the data collected in order to arrive at reliable causes of death (3). Interpretation has largely relied on either expert assessment of the VA interviews by physicians, or the application of predetermined algorithms, often based on a decision-tree approach (4). Some researchers have reported attempts to use other methods, such as neural networks (5).

Expert assessment has been shown to be a reliable method for VA interpretation (6), but legitimate concerns remain as to standardization between different experts, the risk of having to change experts over time, and the sheer volume of work involved in assessing large numbers of VAs. Algorithms have the potential to address all of these concerns (7) but raise others, such as their reliability, and the difficulty of considering parallel possibilities along the lines of a classic clinical differential diagnosis.

Bayesian probability models, originated in principle by Revd Thomas Bayes in 1763 (8), have been extensively explored in the context of medical decision support systems, and shown in many cases to be relatively effective (9). In this paper, we have developed an application of Bayes' theorem for VA interpretation, which may be able to meet the concerns around expert assessment whilst also overcoming some of the

This paper has been independently peer-reviewed according to the usual *SJPH* practice and accepted as an original article.

 ISSN 1403-4948
DOI: 10.1080/14034950310015086

limitations of algorithmic approaches. This has been evaluated on a preliminary basis using a series of VAs from rural Vietnam (10).

METHODS

The process of VA attempts to collect indications concerning the circumstances of death that can lead to one or more possible causes of death. Bayes' theorem seeks to define the probability of a cause (C) given the presence of a particular indicator (I), represented as P(C|I), and can be stated 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).

Thus if VAs collect a set of indicators $I_1 \ldots I_n$ which can lead to a set of causes $C_1 \ldots C_m$, then associated with each indicator I_j and each cause C_k is its probability of occurrence at the population level, which in this case means among all cases of death. Furthermore there is an $(n \times m)$ matrix of probabilities $I_j|C_k$, again among all deaths.

Thus, for a particular case, the probability of C_k is initially the value found among deaths in general, which is the cause-specific mortality fraction (CSMF). However, for each case and for each applicable I_j, C_k can be modified by the above theorem. This is likely to increase the probability of some causes, whilst reducing others.

The issue then arises of how to derive a set of $P(I_1 \ldots I_n)$, $P(C_1 \ldots C_m)$ and the matrix $P((I_1 \ldots I_n)|(C_1 \ldots C_m))$ that can be used in practice to interpret VAs. Although this might seem a difficult task, previous work suggests that a high degree of precision is not essential for these probabilities, in order to build a workable model (11). Thus, for an initial exploration of the method, we compiled a set of probabilities on a semi-qualitative scale as given in Table I, following work by Kong et al. (12). These related to a set of indicators and characteristics more or less representing Indepth's proposed VA questionnaire (13), covering the sets of indicators and causes given in Table II. At this early stage of methodological development, we simply made estimates of probability based on accumulated personal experience, without any attempt to validate or establish internal consistency between estimated values.

A simple program was then prepared using FoxPro software to provide a user interface into which indicators for a particular case could be entered successively, leading to an output of the most likely causes and associated probabilities. For each case, the three most likely causes (provided that their final probabilities exceeded the square root of their initial probabilities (14)) were listed with their associated probabilities. From this output, it was also possible to estimate a certainty factor for each case, which we defined as the sum of the probabilities for the three most likely causes, divided by 3.

For the purposes of initial validation, a set of 189 VAs from rural Vietnam were used, which had previously been assessed by two physicians, leading to consensus on a single cause for each case. These data and the underlying VA process are described in detail elsewhere (10). There was no attempt to standardize the sets of indicators and causes between the probabilistic model and the VA process as implemented in Vietnam.

Table I. *A semi-qualitative scale for assigning probabilities of indicators and causes*

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

RESULTS

Applying the probabilistic model to the VA data from 189 interviews in Vietnam enabled a comparison to be made with the cause of death as previously agreed by two local physicians. In 96/189 cases (50.8%) the most probable cause as determined by the model coincided with the physicians' opinion. In a further 38 cases (20.1%), the physicians' opinion was among the three most likely causes given by the model. In 21 cases (11.1%) the model contradicted the physicians, and in 34 cases (18.0%) the conclusions were indeterminate. Of the latter group, 29 cases were described as "unknown" or "old age" by the physicians; if these cases are excluded from the overall comparison, then 96/160 (60.0%) corresponded directly, 38 (23.8%) were among the first three, 21 (13.1%) were contradictory and five (3.1%) were indeterminate. On this basis, the model gave a satisfactory outcome for 134/160 cases (83.8%). Table III gives cause of death as determined by the physicians compared with the most likely cause from the probabilistic model.

Table II. *Sets of verbal autopsy indicators and causes used in the preliminary model*

Indicators		Causes
Was this an elder 50+ years	Any abdominal mass	*Accident (not transport related)*
Was this a female 15–49 years	Any abdominal pain	*Transport-related accident*
Was this a male 15–49 years	Any coughing with blood	*Acute cardiac*
Was this a child 1–14 years	Any diarrhoea with blood	*Chronic cardiac*
Was this an infant 6 wks–1 yr	Any vomiting with blood	*Other cardiovascular*
Was this a neonate <6 wks	Any chest pain	*Disease of digestive system*
Was she pregnant at death	Any delivery/neonatal problems	*Drowning*
Pregnancy ended within 6 wks	Any convulsions or fits	*HIV/Aids related*
Final illness lasted >1 week	Any cough	*Infection*
Final illness lasted few days	Any diarrhoea	*Kidney or urinary disease*
Was death very sudden	Any difficulty breathing	*Liver disease*
Was death during wet season	Any history of epilepsy	*Disease of respiratory system*
Was death during dry season	Any fever	*Malaria*
Was s/he in transport accident	Was the fontanelle raised	*Malignancy*
Did s/he drown	Was the fontanelle sunken	*Malnutrition*
Had s/he fallen recently	Any headache	*Maternity related*
Any poisoning, bite, sting	Was there paralysis both sides	*Measles*
Did s/he smoke regularly	Any paralysis on 1 side only	*Neonatal*
Any obvious recent injury	Any skin problems	*Disease of nervous system*
Did s/he drink alcohol	Any stiff neck	*Non-specific old age*
Was this a multiple birth	Any general stiffness	*Poisoning*
Adequately vaccinated	Any swelling of ankles/legs	*Stroke*
Any history of heart disease	Any unusual lumps	*Suicide*
Any history of asthma	Any other swelling	*Tetanus*
Any history of diabetes	Any diagnosis of tb	*Tuberculosis*
Any history of hypertension	Any umbilical infection	
Any history of liver disease	Any abnormality of urine	
Any history of cancer	Any vomiting	
Any history of stroke	Any wheezing	
Any diagnosis of HIV/Aids	Any weight loss	
Any suggestion of suicide	Any yellowness/jaundice	
Any surgery just before death	Any fluid on the lungs	
Unconscious before death		

Figure 1 shows cause-specific mortality fractions (for the 160 cases with specific causes determined by the physicians), separately derived from the physicians' opinions, the most likely causes from the probabilistic model, and weighted averages of the three most likely causes from the probabilistic model.

A certainty factor was calculated for each case, as described above. For the cases corresponding directly, the mean certainty factor was 72%, for those among the first three likely causes 88%, for the contradictory cases 74% and for the indeterminate cases 47%. For the cases coded by the physicians as "old age" the mean certainty factor was 52%, and for those coded "unknown" 25%.

There were common factors among the 21 contradictory and five indeterminate cases among the 160 with specific causes. For example, cases where there was a history of accident or injury leading to death were assigned by the physicians to be accidental deaths, although in some cases medical symptoms such as difficult breathing or paralysis were also noted, possibly arising between the time of injury and death. Thus in eight contradictory cases described as "accident" by the physicians, the model arrived at probable causes related to other noted symptoms, and thus did not include "accident" among the three likely causes. A number of other contradictory cases related to long-standing conditions where the relationship between the patient history and the circumstances of death were not entirely clear.

DISCUSSION

Given that there was no direct linkage between the development of the probabilistic model and the VA process as applied in Vietnam, and that the model's initial database was somewhat arbitrary and not related specifically to the Vietnamese setting, the results from the comparison of the model's output with that of the local physicians are encouraging. Whilst there are clearly lessons to be learned from this preliminary validation, which should serve to improve the performance of the probabilistic model, it seems

Table III. *Comparison of cause of death assessment by physicians and most likely cause by probabilistic modelling, for 189 verbal autopsies from Vietnam*

Physicians' cause	Accident	Childbirth	Congenital	Cardiovascular	Diarrhoea	Digestive	Drowning	Infant	Malaria	Neoplasms	Nervous	Old age	Respiratory	Suicide	Unknown	Urinary	Total
Model most likely cause																	
Accident	4																4
Acute cardiac				5									2				7
Chronic cardiac	1			9		1						5	14		1		31
Digestive						1											1
Drowning							8										8
HIV/Aids			1													3	4
Infection				1		2									1	1	5
Kidney/urinary																1	1
Liver disease	1					3											4
Malaria	1		1				1	1	1	1			1				7
Malignancy	1			2	1					21		10			2	1	38
Malnutrition												2	1				3
Neonatal		4						3									7
Respiratory disease	1	1	1	1		1				3		1	11		3		23
Stroke	8			24		1				1	3	2	4		2		45
Suicide														1			1
Total	17	5	3	42	1	9	9	4	1	26	3	20	33	1	9	6	189

Note: The 96/189 directly corresponding cases are shown in the shaded cells.

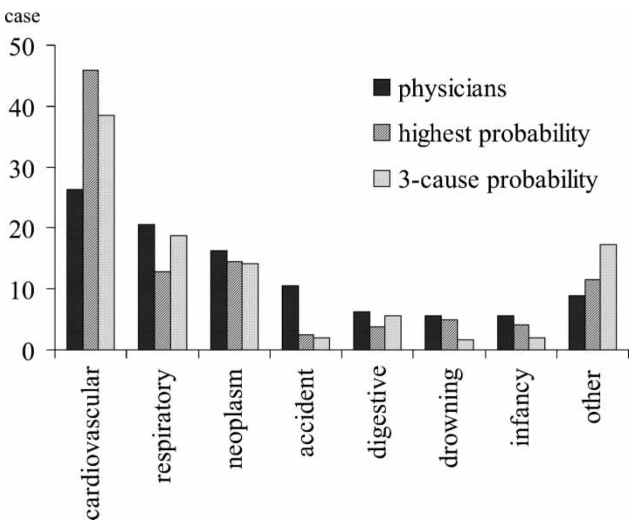

Fig. 1. Cause-specific mortality fractions (CSMF) for major causes, derived from 160 verbal autopsies in Vietnam, according to physicians' opinions, most likely causes from a probabilistic model and weighted averages of the three most likely causes for each case from a probabilistic model.

that this new approach to VA interpretation is relatively robust and has potential.

This validation also raised issues of how to gather and handle VA data. Some indicators included in the probabilistic model were not available routinely in the data from Vietnam (pregnancy, duration of illness, falling, smoking, alcohol consumption, vaccination, hypertension, history of liver disease, history of cancer, history of stroke, diagnosis of HIV/AIDS, recent surgery, cough, raised or lowered fontanelle, skin conditions, stiff neck, fluid on lungs), although in some cases these were noted in free-text comment fields. Similarly some indicators were available in the data (shivering, cyanosis, low back pain, burns) but not built into the model. It is likely that these mismatches reduced the model's performance.

Although physician interpretation of VA data is often considered to be some kind of "gold standard", there is of course scope for misclassification and misinterpretation (15). It may or may not be possible to allow for this in interpreting results (16). In this series, in one case "infant" was ascribed by the physicians as a cause for a 3-year-old child, for example. In some of the contradictory and indeterminate cases, it was not immediately obvious why the physicians' conclusion was more appropriate than that of the model – though in some cases it was clear that the model arrived at the wrong result.

These data also raise the important question of how to describe the death of old people. Vietnam is a rapidly ageing society, and in these data 111/189 (58.7%) of the deaths occurred at the age of 70 and over, with 58/189 (30.7%) at 80 and over. Although some of these deaths had clearly determinable causes, others seemed to be due to a rather imprecise notion of "old age". Whilst the probabilistic model did include "non-specific old age" as a possible cause, its likelihood did not generally exceed more specific causes in the modelling. This may partly be due to the age indicator in the model having 50 years as a cut-off for the oldest group, and is something that needs further thought. At the same time, perhaps it would be appropriate to revisit the conceptual issues around the causes of elderly deaths from a public health perspective, since there is little notion or possibility of preventing many deaths in the very elderly.

The data in the preliminary model tended to overestimate the role of infectious diseases in this Vietnamese population – perhaps not surprising given that it was formulated on the basis of extensive personal experience in Africa. However, this also raises the question as to what extent the database for such a model should reflect local conditions (for example, malaria endemicity, HIV prevalence, etc.). One of the potential strengths of the probabilistic model is the possibility of enabling a standard VA interpretation across a range of settings, but it may also be possible or necessary to refine it in future by including some key local characteristics in the modelling.

Many of the above issues raise wider questions of what is expected and hoped for as the outcome of VA investigations. Naturally VAs are generally performed where other cause of death data are not available. The assumption tends to follow that the aim of the VA is to mimic as far as possible a process of physician certification, including where necessary post-mortem findings, leading to one or more causes of death for an individual. In many cases, VAs have simplified the process to a single cause of death per individual, and not pursued secondary or underlying causes. This probabilistic approach would seem to lend itself, with further development, to the possibility of multiple causes. But taking this reasoning further, it may also be useful, in determining causes of death at the community level, to allow alternative probable causes of death to contribute proportionately in an overall analysis, rather than forcing an unsubstantiated choice between one and another at the individual level. It is likely to be difficult in some cases to determine, for example, the difference between respiratory disease and chronic heart conditions, or whether or not a fall and injury arose following an acute stroke or heart attack.

The possibility of attaching some indication of certainty to the model's outcome may also be useful in interpreting results. It is clear that the calculated certainty for those cases that the physicians were unable to assess reliably was indeed lower. The highest certainty was associated with the cases where the physicians' interpretation was the second or third likely cause; this generally reflects two or three high probability causes in each of these cases, often more or less equal. Higher probabilities for particular causes, and hence greater certainty, generally arose from larger numbers of contributing indicators. This is probably an appropriate principle, as it means for example that a VA reporting only "stroke" could lead to an outcome of "stroke" with lower certainty than one that described all the associated signs and symptoms of the stroke.

There are thus many unanswered questions around the whole process of collecting VA data appropriately and accurately, plus interpreting it meaningfully and reliably. We believe that this new approach to interpretation, with further development, may help to address some of these issues. The next steps in the development of this system need (a) to address refinements to the probability model, based on data and/or on an expert Delphi approach, and (b) to test the refined model with more extensive data from a wider range of sources.

ACKNOWLEDGEMENTS

Institutional support from the Swedish Council for Research on Work and Social Science for the methodological development work, and support to FilaBavi in Vietnam from Sida/SAREC, Stockholm for the verbal autopsy process are appreciated.

REFERENCES

1. Chandramohan D, Maude GH, Rodrigues LC, Hayes RJ. Verbal autopsies for adult deaths: issues in their development and validation. Int J Epidemiol 1994; 23: 213–22.
2. Chandramohan D, Maude GH, Rodrigues LC, Hayes RJ. Verbal autopsies for adult deaths: their development and validation in a multicentre study. Trop Med Int Health 1998; 3: 436–446.
3. Kahn K, Tollman SM, Garenne M, Gear JSS. Who dies from what? Determining cause of death in South Africa's rural north-east. Trop Med Int Health 1999; 4: 433–41.
4. Quigley MA, Chandramohan D, Setel P, Binka F, Rodrigues LC. Validity of data-derived algorithms for ascertaining causes of adult death in two African sites using verbal autopsy. Trop Med Int Health 2000; 5: 33–9.
5. Boulle A, Chandramohan D, Weller P. A case study of using artificial neural networks for classifying cause of death from verbal autopsy. Int J Epidemiol 2001; 30: 515–20.
6. 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.
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. Bayes T. A problem in the doctrine of chances. Philos Trans R Soc Lond 1763; 53: 370–418.
9. Long WJ. Medical informatics: reasoning methods. Artif Intell Med 2001; 23: 71–87.
10. 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.
11. Byass P, Corrah PT. Assessment of a probabilistic decision support methodology for tropical health care. In: Barber B, Cao D, Qin D, Wagner G, editors. Proceedings of Medinfo-89. Amsterdam: North-Holland: 995–1000.
12. Kong A, Barnett GO, Mosteller F, Youtz C. How medical professionals evaluate expressions of probability. N Engl J Med 1986; 315: 740–4.
13. Indepth Network. Population and health in developing countries, Volume 1: Population, health and survival. Ottawa: Indepth Network, IDRC: 2002.
14. Byass P. A probabilistic approach to health care delivery in tropical Africa. Lecture Notes in Medical Informatics 1988; 36: 136–41.
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. Chandramohan D, Setel P, Quigley M. Effect of misclassification of causes of death in verbal autopsy: can it be adjusted? Int J Epidemiol 2001; 30: 509–14.