Theme: Strengthening vital statistics and cause-of-death data

Symptom recall and the diagnostic accuracy of verbal autopsies

Peter Serina

Ian Riley

Bernardo Hernandez

Michael Freeman

Devarsetty Praveen

Veronica Tallo

Rohina Joshi

Diozele Sanvictores

Christopher JL Murray

Alan Lopez

Working Paper Series • Number 34 • September 2013 • WORKING PAPER

Health Information Systems Knowledge Hub

School of Population Health
University of Queensland

Strengthening health systems in Asia and the Pacific through better evidence and practice

An AusAID funded initiative

For the PDF version of this paper and other related documents, visit

www.uq.edu.au/hishub

About this series

The Health Information Systems Knowledge Hub's Working Paper Series is the principal means to disseminate the knowledge products developed by the hub as easily accessible resources that collectively form a lasting repository of the research findings and knowledge generated by the hub's activities. Working papers are intended to stimulate debate and promote the adoption of best practice for health information systems in the region. The series focuses on a range of knowledge gaps, including new tools, methods and approaches, and raises and debates fundamental issues around the orientation, purpose and functioning of health information systems. Generally, working papers contain more detailed information than a journal article, are written in less-academic language, and are intended to inform health system dialogue in and between countries and a range of development partners.

Many working papers have accompanying products such as summaries, key points and action guides. The full range of documents, as well as other resources and tools, is available on the Health Information Systems Knowledge Hub website at www.uq.edu.au/hishub/publication-tools.

This research has been funded by AusAID. The views represented are not necessarily those of AusAID or the Australian Government.

Published by the Health Information Systems Knowledge Hub,
School of Population Health
The University of Queensland
School of Population Health Building,
Herston Rd, Herston Qld 4006, Australia

Please contact us for additional copies of this publication, or send us feedback:

Email: hishub@sph.uq.edu.au
Tel: +61 7 3365 5400
Fax: +61 7 3365 5442
www.uq.edu.au/hishub

Design by Biotext, Canberra, Australia

Acknowledgments

The Population Health Metrics Research Consortium (PHMRC) study on gold standard verbal autopsy validity was funded as part of a Bill & Melinda Gates Foundation 'Grand Challenges in Global Health' grant to the following PHMRC members: Institute for Health Metrics and Evaluation at the University of Washington; the Johns Hopkins School of Public Health; and the School of Population Health at the University of Queensland.

Repeat verbal autopsy interviews for symptom recall were funded by AusAID through the Health Information Systems Knowledge Hub at the School of Population Health, University of Queensland. The views represented are not necessarily those of AusAID or the Australian Government.

Field staff were trained and supervised through a National Health and Medical Research Council grant for improving methods to measure comparable mortality by cause.

Author details

Peter Serina, Bernardo Hernandez, Michael Freeman and Christopher JL Murray: Institute for Health Metrics and Evaluation, University of Washington

Ian Riley and Alan Lopez: School of Population Health, The University of Queensland, Australia

Devarsetty Praveen: George Institute, Hyderabad, Andhra Pradesh, India

Veronica Tallo and Diozele Sanvictores: Research Institute of Tropical Medicine, Manila, Philippines

Rohina Joshi: George Institute, University of Sydney, Australia

Contents

Acronyms

CCC	chance-corrected concordance
COD	cause of death
CSMF	cause-specific mortality fraction
PHMRC	Population Health Metrics Research Consortium
VA	verbal autopsy
VAI	verbal autopsy instrument
WHO	World Health Organization

Summary

A verbal autopsy (VA) is a questionnaire administered to the caregivers or family members of a person who has died (the decedent), to elicit specific signs and symptoms that occurred in the period before death. The Population Health Metrics Research Consortium (PHMRC) gold standard validation study developed new approaches to and methods for validating VAs. The study collected VAs from decedents' families shortly after death and compared the cause of death (COD) assigned from these VAs against the 'gold standard' COD originally assigned in the hospital. From this, the study established a set of CODs that can be reliably determined through VA. The study left open the question of whether longer recall periods—the length of time between death and administering the VA instrument—affect the ability of the VA to diagnose COD.

In this study, we repeated the VAs of a subset of the deaths from the PHMRC study to:

1. quantify the effect of recall period on the ability of the VA to diagnose COD

2. analyse the relationship between variation in symptoms recalled and consequent variation in the assigned COD.

We repeated VA interviews with the families of 2113 decedents within 3–52 months of death in India (Andhra Pradesh) and the Philippines (Bohol and Manila). We used the tariff method to assign the COD for each VA collected. For each decedent, we had the VA-assigned COD based on two separate interviews and the gold standard diagnosis assigned in the hospital in the PHMRC study. We then compared the two VA-assigned CODs to each other and to the gold standard diagnosis to address our two research questions.

To assess the effect of recall period on VA quality, we created a logistic regression framework that determined if recall period had an effect on the VA COD assignment as compared to the gold standard COD. To assess the relationship between variation in symptoms recalled, we calculated the average per cent change in symptom recall between the first and second VA. We used this symptom variability in two ways. First, we used a logistic regression framework to determine if a decedent's average symptom variability was associated with the matching of the CODs assigned in the first and second VA. Second, we made a proxy for symptom importance and determined if variability for a given symptom was related to that symptom's relative importance in assigning the COD.

We found a decrease over time in the ability of the VA to assign the correct COD, but the association was weak and not statistically significant. Based on these results, we recommend that the maximum duration between the time of death and the time of administering a VA be extended from one to four years. Having said that, the ideal time to administer the VA is within a few months of death, after an appropriate period of mourning.

We conclude, first, that when symptom recall varies, the VA is less consistent in assigning the correct COD. Second, those variables that interviewees remembered best are also the most important for correctly diagnosing COD for adults and children. This finding raises questions about the mechanisms people use to retain memories of circumstances surrounding a death and supports current work to remove symptoms of lesser importance in VA instruments used for routine surveillance.

Introduction

Background

A verbal autopsy (VA) is a questionnaire administered to the caregivers or family members of a person who has died (the decedent) to elicit specific signs and symptoms that occurred in the period before death. VAs have been used in demographic surveillance sites and in studies of the epidemiology of disease for over 40 years. Modern VA instruments (VAIs) belong to a family which includes instruments used by the World Health Organization (WHO) and the Population Health Metrics Consortium (PHMRC).

VAs are increasingly becoming part of routine surveillance of cause of death (COD) through sample and civil registration systems (Setel et al. 2005). Hence it is increasingly important to understand empirically the effects of a given context on the predictive validity of the method. One key contextual feature is the time between death and administration of the instrument: the recall period. The WHO recommends a recall period of not greater than one year. In the PHMRC gold standard validation study described below, a three-month recall period was required, although this was extended to as long as five months in some cases.

Currently there is little or no evidence to inform recall period requirements. This study aimed to quantify the effect of recall period on VA diagnostic performance by using a paired dataset in which two VAs were administered for a single decedent. Preliminary results from the analysis of these data based on the analysis of cause-specific mortality fractions (CSMFs) were presented in Riley et al. (2010). This paper reported little change in CSMFs between first and second interviews but we were unable to explain why this would be so. The development of new methods for assigning COD to VAs enabled the analysis of results at the level of the individual and has allowed us to explain results reported in the earlier paper.

Population Health Metrics Research Consortium gold standard validation study of verbal autopsies

The PHMRC gold standard validation study (Murray et al. 2011) developed new approaches to and methods for validating verbal autopsies. The study was of 12,539 deaths drawn from Bohol and Manila, Philippines; Andhra Pradesh and Uttar Pradesh, India; Dar es Salaam and Pemba Island, Tanzania; and Mexico City, Mexico.

CODs assigned to VAs were validated by comparing them against 'gold standard' hospital diagnoses of the same deaths. Gold standard diagnoses were those made at the highest possible levels of certainty practicable in hospitals in the countries listed above. A target list of causes was based on the global burden of disease (Lopez et al. 2006) and on the practicability of collecting high quality data. Diagnostic criteria with high levels of certainty were developed for each cause. Hospital deaths that met these criteria were known as gold standard deaths. A VA was administered in the home to the families of each of these decedents, with the interviewers blinded to the hospital diagnosis.

The study design made it possible to evaluate different methods of assigning individual COD and to evaluate them against each other. These methods can be classified in two broad groups: i) those based on expert judgement, and ii) computer-driven analyses. The first group includes physician certification of the COD based directly on the VA interview and certification based on expert algorithms. In the second group, the computer-driven analyses make no prior assumptions about the relationship between symptoms and their causes: computer-learning and its validation depend on empirical data analysis. Performance of a method of analysis varies according to the distribution of CODs in a particular sample. The dataset is split randomly into two sets, called Train and Test. The computer algorithm 'learns' from deaths in the Train set; its performance is evaluated by predicting the COD in cases in the Test set. Final accuracy of estimated cause-specific mortality fractions (CSMFs)—performance at the level of the population—and the chance-corrected concordance (CCC) of CODs—performance at the level of the individual—are based on the performance of a particular method against the COD from (gold standard) medical records over some hundreds of test runs. In this symptom recall study, due to the small sample size and lack of data for certain causes, we used simple concordance for the evaluation of the method.

The computer-driven methods outperformed methods based on expert judgement. Tariff 2.0, an updated version of the Tariff Method published by the PHMRC (James et al. 2011), calculates a score, or tariff, for each symptom–cause pair to measure the strength of association between the two. This updated version of

Tariff included the removal of non-significant tariffs due to random noise in the dataset, logic constraints to disallow biologically impossible COD predictions, and the addition of an indeterminate category when there is insufficient symptom information to assign COD. In Tariff 2.0, CSMF accuracy for adults is 0.772 (95% confidence interval [CI]: 0.765, 0.770), for children is 0.785 (95% CI: 0.776, 0.789) and for neonates is 0.827 (95% CI: 0.818, 0.835). CCC for adults is 50.5% (95% CI: 50.3, 50.6), for children is 52.3% (95% CI: 51.5, 52.5) and for neonates is 48.2% (95% CI: 47.9, 48.5).

The PHMRC study developed computer-based methods for predicting 78 different causes of death and opened the way for widespread acceptance of verbal autopsies as a basis for official COD statistics. The database is likely to provide a foundation for evaluating verbal autopsy methods for years to come.

Symptom recall

Recall bias may affect the validity of VAs, but there is little published work on the optimal recall period. A recent review (Soleman et al. 2006) says simply that recall periods ranging from one to 12 months are thought to be acceptable. It refers to a single validation study (Chandromohan 2001) that showed no difference in sensitivity and specificity based on differences in recall period length of one to 21 months; Soleman and colleagues conclude that further work is needed to define the acceptable recall period (Soleman et al. 2006). A second study examined a restricted list of CODs in women of reproductive age in Burkina Faso and Indonesia and reported high levels of agreement at the population but not at the individual level (Byass et al. 2009).

The PHMRC study protocol required VAs to be collected within three months of death. This left open the question of whether longer recall periods would still provide valid and reliable information, when using the same instrument. Most VAs are derived from the demographic surveillance of small populations through regular household rounds. If the time interval between death and VA could be lengthened, the greater flexibility in administering VAs would allow survey costs to be proportionally reduced.

This report addresses two key questions about symptom recall:

1. How long after death can VAs be conducted and still capture key symptoms/signs and medical record recall?
2. How much will cause assignment be affected by symptom-level reliability?

Methods

This study repeated VAs, using the PHMRC VAI, within 3–52 months of death in PHMRC study sites in Bohol and Manila, Philippines, and Andhra Pradesh, India. The PHMRC study aimed to analyse a set number of cases for each COD. The Manila study site was added to the Bohol site in the Philippines to increase the number of deaths in adults from chronic conditions, especially cancers.

PHMRC verbal autopsy instrument

The VAI was a modified version of the WHO instrument (WHO 2012). It comprised:

- general information module: administered for all deaths; collects information such as education of the decedent, household characteristics, and a household roster
- adult module (12y and older): collects a history of chronic conditions, symptoms of the decedent, women's health questions if decedent is female, alcohol and tobacco use, and injury or accident information; it transcribes any available medical and death records; and it has an open-ended response field
- child and neonatal modules (under 12y): both modules collect a history of the birth of the child, transcribe any available medical and death records, and contain an open-ended response field
- child module (28d to 11y): collects symptoms of the deceased and information about injuries and accidents
- neonatal module (under 28d): collects maternal history, neonatal behaviour and symptoms of the decedent.

Collection and analysis of data

We studied deaths that occurred during 2007 and 2008 in Bohol, Manila and Andhra Pradesh, with two VAs being collected from the family of each decedent. The first (validation) VAs had been collected during the PHMRC study between six days and five months after death. We visited the same families and collected a subset of these (recall VAs) between three and twenty months after death in the three sites. We then took the opportunity afforded by the ongoing collection of VA data in Bohol (Lopez n.d.) to collect more recall VAs from a different subset of the validation dataset 18–52 months after death. We refer to these two recall data subsets from Bohol as Bohol (1) and Bohol (2).

COD was assigned by Tariff 2.0 (Serina, unpublished data). The Tariff method calculates a score for each symptom–cause pair to measure strength of association between the two and uses these scores to assign COD (James et al. 2011). This COD was then compared to the gold standard COD assigned in hospital. The effect of time on obtaining a correct diagnosis (against gold standard) was measured using a logistic regression framework that controlled for site of data collection and survey module. The data were clustered on the individual to control for the paired (i.e. non-independent) observations.

$$P(\text{correct diagnosis}) = \text{logit}(\beta_0 + \beta_1 \text{time} + \beta_2 \text{time}^2 + \beta_3 \text{site} + \beta_4 \text{module})$$

where

Correct diagnosis: defined as 1 when Tariff assigns the same COD as the gold standard for a particular VA administration

Time: the time in months between death and VA administration

Site: location where VA was administered. Andhra Pradesh was set as the reference category. The other sites were Bohol (1), Bohol (2) and Manila

Module: survey module of the VA questionnaire that was used for a given death. Adult (12 years and older) was the reference category for child (1 month to 12 years) and neonate (0–28 days).

Symptoms reported for an individual death varied between the two survey rounds, as did the Tariff-assigned COD. Such variation can be simply expressed in a 2 x 2 table:

		Survey 2	
		Yes	No
Survey 1	Yes	A	C
	No	B	D

Percentage change (or variability) of symptoms was calculated by:

$$\text{Per cent Gain} = \frac{\#\text{ (No survey 1, Yes survey 2)}}{\text{Total \# Yes survey 1}} = \frac{B}{A+B}$$

$$\text{Per cent Loss} = \frac{\#\text{ (Yes survey 1, No survey 2)}}{\text{Total \# Yes survey 1}} = \frac{C}{A+B}$$

$$\text{Variability (per cent change)} = \frac{\#\text{ (Yes survey 1, No survey 2)} + \#\text{ (Yes survey 2, No survey 3)}}{\text{Total \# questions}}$$

$$= \frac{B + C}{A + B + C + D}$$

In a second analysis, also controlling for site of data collection, module and respondent characteristics, average per cent change in symptom recall between the PHMRC validation VA and the recall VA was used to predict whether the assigned COD would vary for an individual. A proxy was also created, ranking symptoms in order of their relative importance for diagnosing COD using VA. This proxy was based on ranking the standard deviation of tariffs across symptoms.

We then ran a logistic regression looking at the effect of the variability in symptom recall (per cent change) on the probability of the VA diagnosis matching between the first and second administration of the VA.

$$P(\text{diagnosis match}) = \text{logit}(\beta_0 + \beta_1 \text{variability} + \beta_2 \text{respondent match} + \beta_3 \text{site} + \beta_4 \text{module})$$

where

Diagnosis match: defined as 1 when the first and second VA have the same diagnosis according to Tariff 2.0

Variability: the percentage of symptoms that changed in recall between the first and second interview

Respondent match: indicator referencing whether the respondent is the same in the first and second administration of the VA

Site: location where the VA was administered. Andhra Pradesh was set as the reference category. The other sites were Bohol (1), Bohol (2) and Manila

Module: survey module of the VA questionnaire that was used for a given death. Adult (12 years and older) was the reference category for child (1 month to 12 years) and neonate (0–28 days).

Ethical approval

Ethical approval for these studies was obtained from the Medical Research Ethics Approval Committee of the University of Queensland, Brisbane, Australia; the Research Institute of Tropical Medicine Institutional Review Board, Alabang, Muntinlupa, Philippines; and Gandhi Medical College, Hyderabad, India.

Results

From the PHMRC study (Murray et al. 2011), 2113 VAs were chosen (Table 1) and repeated for symptom recall (Table 2). CCC varied by survey round (validation or recall), VAI module (adults, children or neonates) and by site (Table 3). Mean CCC was 0.497 for the validation round and 0.464 for the recall round.

Table 1 Validation datasets (number of cases)

Site	Bohol	Manila	Andhra Pradesh	Total
Adult	547	190	657	1394
Child	87	59	203	349
Neonate	176	37	157	370
Total	810	286	1017	2113

Table 2 Symptom recall datasets (number of cases)

Site	Bohol (1) (3–18 months after death)	Bohol (2) (18–52 months after death)	Manila	Andhra Pradesh	Total
Adult	235	312	190	657	1394
Child	45	42	59	203	349
Neonate	69	107	37	157	370
Total	349	461	286	1017	2113

Table 3 Concordance by survey round and site

Module	Site	Validation data	Symptom recall data	Number of cases
Adult	Andhra Pradesh	0.406	0.393	657
	Bohol (1)	0.542	0.513	312
	Bohol (2)	0.481	0.417	235
	Manila	0.484	0.474	190
Child	Andhra Pradesh	0.547	0.483	203
	Bohol (1)	0.357	0.357	42
	Bohol (2)	0.356	0.400	45
	Manila	0.407	0.305	59
Neonate	Andhra Pradesh	0.541	0.516	157
	Bohol (1)	0.785	0.729	107
	Bohol (2)	0.696	0.638	69
	Manila	0.703	0.568	37

Figure 1 shows simple concordance (or correct per cent diagnoses) by the period since death in months for the study as a whole; Figures 2–5 show simple concordance by period for the three sites. Although the initial impression from Figure 1 is that percentage concordance has increased over the period since death, this impression is corrected in Figures 2–5, which indicate very little variation over time once limits of confidence have been taken into account.

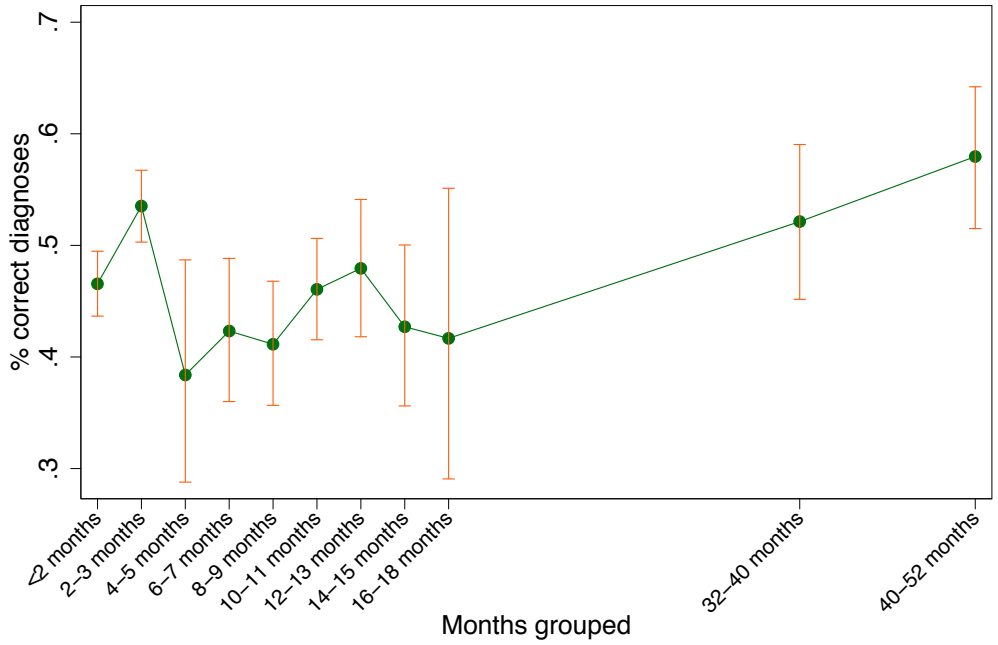

Figure 1 Simple concordance vs. time from death to verbal autopsy administration, ignoring site and module

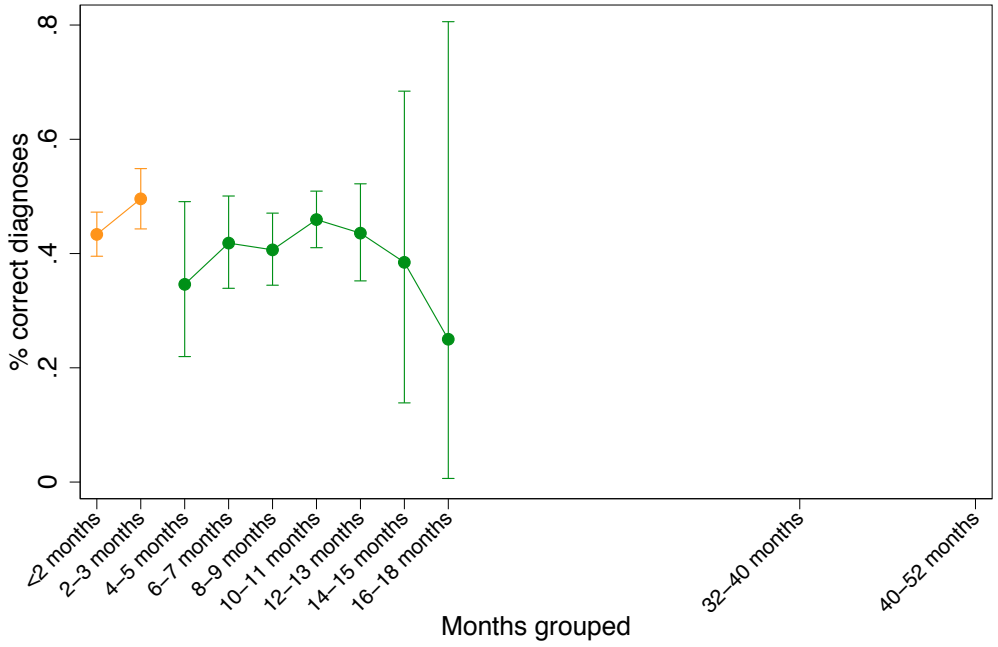

Figure 2 Simple concordance vs. time from death to verbal autopsy administration for Andhra Pradesh

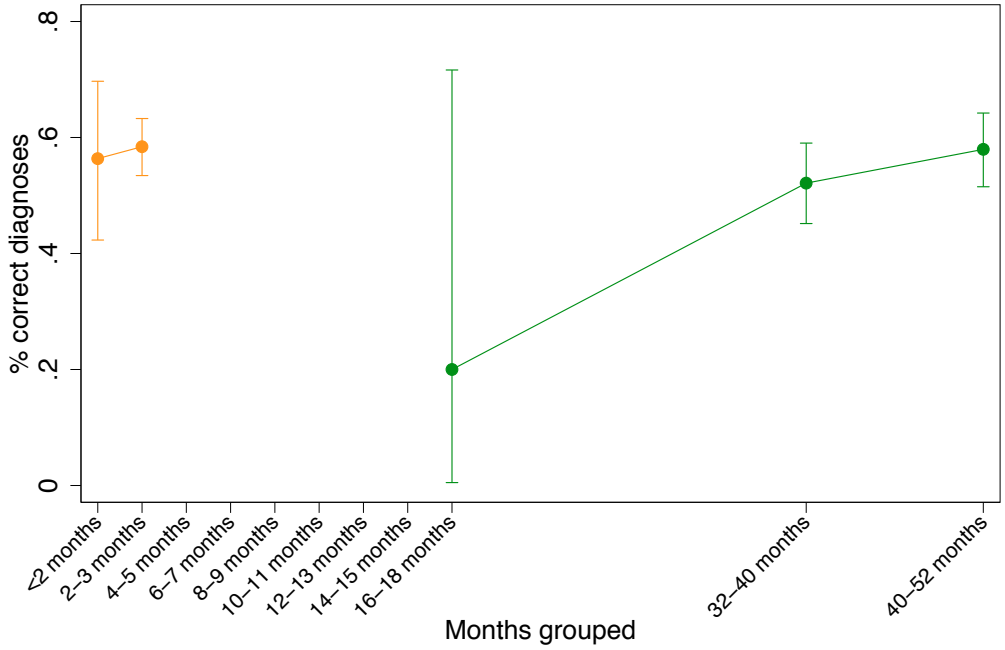

Figure 3 Simple concordance vs. time from death to verbal autopsy administration for Bohol (1)

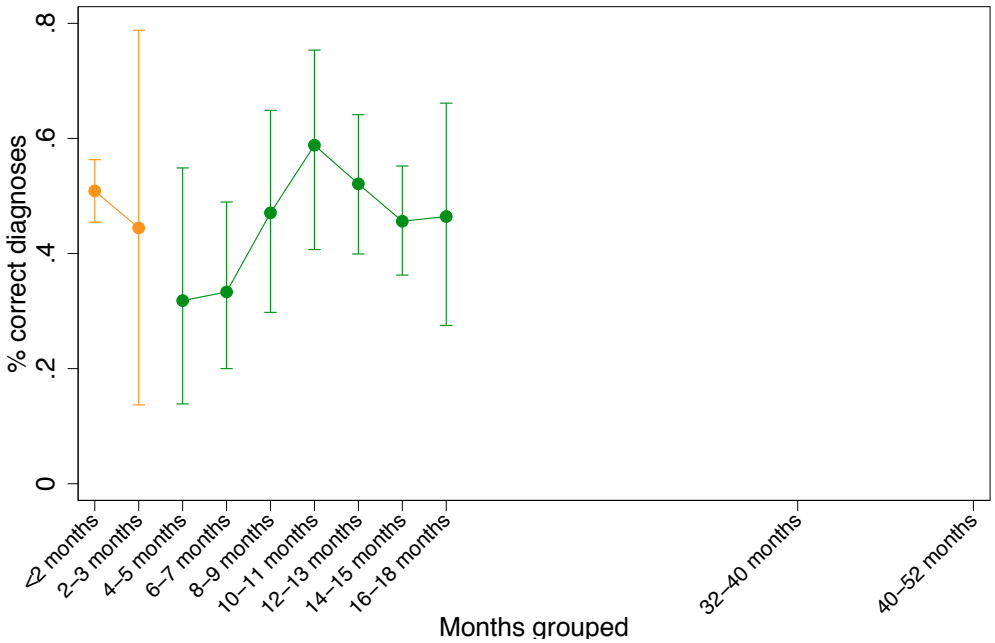

Figure 4 Simple concordance vs. time from death to verbal autopsy administration for Bohol (2)

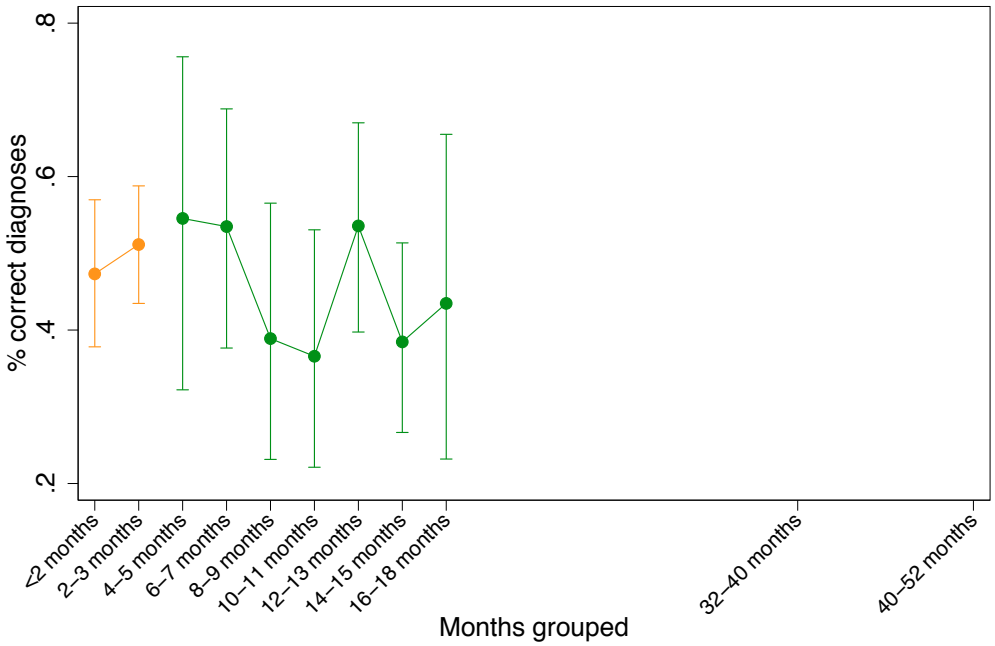

Figure 5 Simple concordance vs. time from death to verbal autopsy administration for Manila

The effect of time in months between death and the VA administration upon the probability of a correct diagnosis was found to have an odds ratio of 0.996 (95% CI 0.991–1.001) between validation and recall VAs (Table 4). This effect was confounded by study site and VAI module. The crude estimate that did not account for site and module had an odds ratio of 1.006 (95% CI 1.001– 1.011); this naïve regression would have led us to the illogical conclusion that an increased amount of time between death and survey administration would have actually increased the probability of getting a correct diagnosis. Therefore, including study site and module and accounting for the paired nature of the dataset were essential to this analysis. This result was not affected by the addition of respondent characteristics (age, gender and level of education) to the framework; therefore, they were excluded from the model.

There was more variation at the individual level than the below results might suggest. Although 49.7% of diagnoses were correct at the first interview and 46.3% were correct at the second interview, 27.3% of correct adult diagnoses, 15.1% of correct child diagnoses and 14.8% of correct neonate diagnoses changed between the two interviews (Table 5). Clearly, this was a consequence of variability in the reporting of symptoms for individual deaths.

Table 4 Logistic regression of the effect of time on correct diagnosis for verbal autopsy pairs

	Odds ratio	95% lower-level confidence	95% upper-level confidence	*p*-value
Time in months	0.996	0.991	1.001	0.121
Andhra Pradesh (reference)				
Manila	1.162	0.921	1.465	0.206
Bohol (1)	1.674	1.364	2.032	0.000
Bohol (2)	1.151	0.934	1.418	0.187
Adult (reference)				
Child	1.066	0.861	1.320	0.559
Neonate	2.071	1.694	2.533	0.000

Table 5 Variability in assigned cause of death in the two survey rounds

	Number of correct diagnoses at first interview	Number of correct diagnoses at second interview	Per cent of correct diagnoses lost	Per cent of correct diagnoses gained	Per cent change in correct diagnosis
Adult	641	606	32.8%	27.3%	5.8%
Child	166	149	25.3%	15.1%	11.4%
Neonate	243	224	22.6%	14.8%	8.5%

Table 6 shows the extent of that variability within VA pairs for the one individual (average per cent gain and loss of positive answers) and the level of change of positive answers for the questionnaire as a whole. On average, positive symptoms changed between the first and second administration by 10.7% for adults, 10.0% for children, and 8.2% for neonates. Logistic regression showed that symptom variability, defined as the per cent change in symptom recall for an individual, was negatively associated with consistent COD prediction with an odds ratio of 0.908 (p = .000). We found that the respondent match between the first and second interviews had a non-significant impact on the probability of Tariff 2.0 getting the same diagnosis for each. The proxy for symptom importance was negatively correlated to change in symptom recall for all modules and significantly correlated in the adult and child modules
(r = .461 and r = .297, respectively) but non-significantly correlated in the neonatal module (r = .075).

Table 6 Variability in responses to symptom questions in the two survey rounds

Module	Total questions	Average per cent gain	Average per cent loss	Average per cent change
Adult	170	52.9%	38.8%	10.7%
Child	84	30.1%	19.1%	10.0%
Neonate	117	27.3%	19.1%	8.2%

Discussion

In these studies, although there was a decline in the performance of the VAI between the validation and recall rounds (Table 3), this was only weakly related to the duration of time between death and the administration of the instrument (Table 4). Graphs of concordance versus time after death do not suggest any particular point in time that would indicate the beginning of a marked decline in performance (Figures 1–4).

Performance at second interview was related also to field site and to module (Table 4). Different performance at the different field sites can be explained in part by variation in the cause composition of mortality at the sites which, in turn, is associated with variability in CCC for different diseases. CCC variability occurs for a number of reasons. One is that diseases such as AIDS and malaria are pleomorphic in symptomatology because of the number of organ systems that can be affected: there is a great deal of variation in their presentations. A second is that many symptoms lack specificity: breathlessness may be reported in deaths from pneumonia, from chronic lung disease, from heart failure due to many different causes, in people with high fevers and so on. A third is that certain symptoms, such as cough, are reported more accurately than are less well-defined symptoms, such as fever.

If diseases are ranked by the level of CCC, it is possible to examine the effect of cause composition on the performance of the instrument in the different sites. Thus, 24.5% of the ranked causes were in the first quintile in Bohol but only 14.4% were in Andhra Pradesh: consequently the instrument performed better in Bohol. Variation between modules is partly due to variation in CCC by disease but, more importantly, the good performance of the neonatal module is a result of the small number of conditions that are being diagnosed.

Although correctness of diagnosis varied little with time, there was much more change in correct diagnosis between interviews than was apparent from the final results: the changes tended to cancel each other out (Table 5). We conclude from the second logistic regression analysis, first, that variation in symptom recall causes inconsistency in prediction of COD in VAs. Second, variables that are the most important for correct diagnosis in the adult and child modules have less variance in recall rate as compared to more unimportant variables. If correctness of diagnosis varies little with time, it follows that "important" variables—those that make the greatest contribution to tariff scores—are retained best in memory and, despite variability, that the strength of these memories is sufficient to ensure consistency in prediction of CODs. This finding raises questions about the mechanisms of retention of key memories. Earlier research has found that the salience of an event and its emotional impact can affect memories, and this is likely to affect the pattern of recall of symptoms as family members and friends try to answer very specific questions about signs and symptoms that occurred during a very tumultuous time in their lives (Bradburn et al. 1987). This conclusion also supports work being done to systematically remove symptoms of lesser importance to create shortened form VAIs for routine surveillance.

The major weakness of this study is that two interviews were collected from the same group of respondents, and that the first interview may have reinforced memories of the events surrounding the death. Also, further study is needed to clarify reasons for variation between interviews that is not associated with time.

We feel it safe, however, to recommend that VAs can be collected for up to four years after a person's death. Clearly, it is best to collect such data within a few months of death, after allowing an appropriate period for mourning.

References

Bradburn, NM, Rips, LJ & Shevell, SK 1987, 'Answering autobiographical questions: the impact of memory and inference on surveys', *Science*, vol. 236, pp. 157–61.

Byass, P, D'Ambruoso, L, Ouédraogo, M & Qomariyah, SN 2009, 'Assessing the repeatability of verbal autopsy for determining cause of death: two case studies among women of reproductive age in Burkina Faso and Indonesia', *Population Health Metrics*, vol. 7, pp. 6. http://www.pophealthmetrics.com/content/7/1/6.

Chandramohan, D 2001, 'Verbal autopsy tools for adult deaths', Unpublished PhD Thesis, London School of Hygiene and Tropical Medicine, London.

James, SL, Flaxman, AD, Murray CJL 2011 & the Population Health Metrics Research Consortium (PHMRC) 2011, 'Performance of the Tariff Method: validation of a simple additive algorithm for analysis of verbal autopsies', *Population Health Metrics*, vol. 9, pp. 31. http://www.pophealthmetrics.com/content/9/1/31.

Lopez, AD, Mathers, CD, Ezzati, M, Jamison, DT & Murray CJL 2006, 'Global and regional burden of disease and risk factors, 2001: systematic analysis of population health data', *Lancet*, vol. 367, pp. 1747–1757.

Lopez, AD (lead investigator) n.d., NHMRC Study: Project Grant Application: 631494, Scientific title: Improving methods to measure comparable mortality by cause.

Murray, C, Lopez, A, Black, R, Ahuja, R, Mohd Ali, S, Baqui, A, Dandona, L, Dantzer, E, Das, V, Dhingra, U, Dutta, A, Fawzi, W, Flaxman, A, Gómez, S, Hernández, B, Joshi, R, Kalter, H, Kumar, A, Kumar, V, Lozano, R, Lucero, M, Mehta, S, Neal, B, Lockett Ohno, S, Prasad, R, Praveen, D, Premji, Z, Ramírez-Villalobos, D, Remolador, H, Riley, I, Romero, M, Said, M, Sanvictores, D, Sazawal, S & Tallo, V 2011, 'Population Health Metrics Research Consortium gold standard verbal autopsy validation study: design, implementation, and development of analysis datasets', *Population Health Metrics* vol. 9, pp. 27. http://www.pophealthmetrics.com/content/9/1/27.

Riley, I, Freeman, M, Praveen, D, Tallo, V, Joshi, R, Sanvictores, D, Murray, CJL & Lopez A 2010, 'Symptom recall and the diagnostic accuracy of verbal autopsies', HIS Knowledge Hub Documentation Note No. 11, School of Population Health, University of Queensland, Brisbane.

Setel, PW, Sankoh, O, Rao, C, Velkoff, VA, Mathers, C, Gonghuan, Y, Hemed, Y, Jha, P & Lopez AD 2005, 'Sample registration of vital events with verbal autopsy: a renewed commitment to measuring and monitoring vital statistics', *Bulletin of the World Health Organization*, vol. 83, pp. 611–617.

Soleman, N, Chandramohan, D & Shibuya, K 2006, 'Verbal autopsy: current practices and challenges', *Bulletin of the World Health Organization*, vol. 84, pp. 239–245.

World Health Organization 2012, *Verbal Autopsy Standards: 2012 WHO Verbal Autopsy Instrument*, World Health Organization, Geneva.

The Knowledge Hubs for Health Initiative

The Health Information Systems Knowledge Hub is one of four hubs established by AusAID in 2008 as part of the Australian Government's commitment to meeting the Millennium Development Goals and improving health in the Asia and Pacific regions. All four hubs share the common goal of expanding the expertise and knowledge base to help inform and guide health policy.

The Knowledge Hubs are funded by AusAID's Strategic Partnership for Health Initiative.

Health Information Systems Knowledge Hub

The University of Queensland

Aims to facilitate the development and integration of health information systems into the broader health system strengthening agenda, and increase local capacity to ensure that cost-effective, timely, reliable and relevant information is available. The Health Information Systems Knowledge Hub also aims to better inform health information systems policies across Asia and the Pacific.
www.uq.edu.au/hishub

Human Resources for Health Knowledge Hub

The University of New South Wales

Aims to contribute to the quality and effectiveness of Australia's engagement in the health sector in the Asia–Pacific region by developing innovative policy options for strengthening human resources for health systems. The hub supports regional, national and international partners to develop effective evidence-informed national policy-making in the field of human resources for health.
www.hrhhub.unsw.edu.au

Health Policy and Health Finance Knowledge Hub

The Nossal Institute for Global Health (University of Melbourne)

Aims to support regional, national and international partners to develop effective evidence-informed national policy-making, particularly in the field of health finance and health systems. Key thematic areas for this hub include comparative analysis of health finance interventions and health system outcomes; the role of non-state providers of health care; and health policy development in the Pacific.
www.ni.unimelb.edu.au

Compass: Women's and Children's Health Knowledge Hub

Compass is a partnership between the Centre for International Child Health, The University of Melbourne, Menzies School of Health Research and Burnet Institute's Centre for International Health.

Aims to enhance the quality and effectiveness of women's and children's health interventions and focuses on supporting the Millennium Development Goals 4 and 5—improved maternal and child health, and universal access to reproductive health. Key thematic areas for this hub include regional strategies for child survival; strengthening health systems for maternal and newborn health; adolescent reproductive health; and nutrition.
www.wchknowledgehub.com.au

The Knowledge Hubs for Health are a strategic partnership initiative funded by the Australian Agency for International Development

Australian AID

Health Information Systems Knowledge Hub
School of Population Health
University of Queensland

COMPASS
Women's and Children's Health Knowledge Hub

UNSW
THE UNIVERSITY OF NEW SOUTH WALES
Human Resources for Health Knowledge Hub

THE UNIVERSITY OF MELBOURNE
THE NOSSAL INSTITUTE FOR GLOBAL HEALTH
HEALTH POLICY & HEALTH FINANCE KNOWLEDGE HUB