

Assessing the repeatability of verbal autopsy for determining cause of death in Burkina Faso and Indonesia

Peter Byass ¹

Lucia D'Ambruoso ¹

Moctar Ouédraogo ²

S. Nurul Qomariyah ³

1. Impact, University of Aberdeen, Aberdeen, AB25 2ZD, United Kingdom

2. GREFSaD, Bobo-Dioulasso, Burkina Faso

3. Impact, Center for Family Welfare, Faculty of Public Health, University of Indonesia, Jakarta, Indonesia

Abstract

Background

Verbal autopsy (VA) is an established tool for assessing cause-specific mortality patterns in communities where deaths are not routinely medically certified, and is an important source of data on deaths among the poorer half of the world's population. However, the repeatability of the VA process has never been investigated, even though it is an important factor in its overall validity. This study analyses repeatability in terms of the overall VA process (from interview to cause-specific mortality fractions (CSMF)), as well as specifically for interview material and individual causes of death, using data from Burkina Faso and Indonesia.

Methods

Two series of repeated VA interviews relating to women of reproductive age in Burkina Faso (n=91) and Indonesia (n=116) were analysed for repeatability in terms of interview material, individual causes of death and CSMFs. All the VA data were interpreted using the InterVA-M model, which provides 100% intrinsic repeatability for interpretation, and thus eliminated the need to consider variations or repeatability in physician coding.

Results

The repeatability of the overall VA process from interview to CSMFs was good in both countries. Repeatability was moderate in the interview material, and lower in terms of individual causes of death. Burkinabé data were less repeatable than

Indonesian, and repeatability also declined with longer recall periods between the death and interview, particularly after two years.

Conclusions

While these analyses do not address the validity of the VA process in absolute terms, repeatability is a prerequisite for intrinsic validity. This study thus adds new understanding to the quest for reliable cause of death assessment in communities lacking routine medical certification of deaths, and confirms the status of VA as an important and reliable tool at the community level.

Background

Garenne and Faveau [1] recently set out a brief history of verbal enquiries into cause of death, including identifying some of the possible limitations, but without mentioning any aspect of repeatability of the process. Verbal autopsy (VA) has become an increasingly well-established approach for determining cause of death in populations lacking universal cause of death registration over the past two decades, and is a very important source of data on deaths among the poorer half of the world's people. Following early VA work in West Africa [2], there have been a number of efforts towards standardisation of VA procedures. Much of this has concentrated on the standardisation of interview questionnaires, culminating in WHO's recently published standards [3]. There has also been work on objective approaches to interpreting material from VA interviews [4-10], albeit with a possible trade-off between standardisation and subtlety of interpretation. Other studies have made comparisons between VA-derived cause of death and arguably "harder" evidence, such as hospital records for deaths occurring in institutions [9-13]. Some of these have been described as "validation" studies for VA, although they have generally only considered selected parts of the overall VA process. However, none of this work has objectively assessed the repeatability of the overall VA process (from individual interviews to aggregated mortality patterns), nor of its constituent parts (interviews, interpretations, individual causes of death, aggregated mortality patterns).

In this paper, we report the results of a study designed specifically to examine the repeatability of the VA process. The study was run in parallel in Burkina Faso and Indonesia, to allow comparison between two very different settings, both of which

involved repeated VA interviews concerning deaths among women of reproductive age. We chose to exclude consideration of repeatability of the interpretation stage of the process, by using the InterVA-M model [14], which gives an intrinsic 100% repeatability, and also allows inter-country comparisons without needing to consider systematic differences between interpreters with different training and backgrounds.

Our major aim was to assess the repeatability of the overall VA process, from interview to aggregated mortality patterns, in the two different settings.

Subsidiary aims were:

1. to assess the repeatability of the interview stage of the VA process (in terms of material gathered in VA interviews);
2. to assess the repeatability of cause of death determination at the individual level;
3. to assess the repeatability of cause-specific mortality fractions (CSMF) determined at the population level

It should be noted that this study did not aim to arrive at conclusions, other than on repeatability, about cause of death patterns in either setting, and not explain differences in mortality patterns between the two countries, evaluate the clinical validity of VA, nor draw conclusions for health planning.

Methods

In both Burkina Faso and Indonesia, large scale community-based surveys of mortality among women of reproductive age, with a particular focus on pregnancy-related mortality, were undertaken in 2005-6 (February to May 2006 in Burkina Faso and December 2005 to June 2006 in Indonesia) [15,16]. These surveys involved identifying deaths that had occurred among women of reproductive age and then undertaking verbal autopsy interviews. These interviews were structured to include the collection of the “indicators” (a range of 75 questions with “yes” or “no/unknown” responses, covering background, pregnancy status, clinical history, signs and symptoms before death and obstetric history) that are needed as the input material for the InterVA-M model [14]. This material from the interviews was interpreted using the InterVA-M model. For the purpose of these investigations into the repeatability of the VA process, subsamples of the originally identified cases were reselected, on a purposeful basis that approximately reflected the overall mortality patterns in the original surveys and were logistically feasible for re-interviewing. The majority of these cases either had no contact with health services around the time of death or case-notes were unavailable. These cases were then revisited in November 2007 (Burkina Faso) and January 2008 (Indonesia) and the VA interviews repeated. Aspects of repeatability within the entire VA process, in both Burkina Faso and Indonesia, have been assessed within the conceptual framework shown in Figure 1.

The interview material thus consisted of individual sets of responses to the 75 possible InterVA-M indicators, and the repeatability of the interview stage of the overall process was assessed by calculating kappa statistics for each indicator, together with p-values assessing whether agreement between the original and follow-up interviews was significantly greater than that expected by chance.

The InterVA-M model was then run on all the original and follow-up interview material from both countries, to interpret likely pregnancy status and cause-of-death outcomes. The InterVA-M model generated, for each case, the most likely pregnancy status at the time of death (pregnant, delivered within 6 weeks or not pregnant within 6 weeks of death) with an associated likelihood. Then up to three likely causes of death were generated, each with an associated likelihood. As the model provides 100% repeatability between indicator input and pregnancy status or cause-of-death output, its performance was not part of this repeatability assessment. The individual level pregnancy status and cause of death likelihoods were summed over all individuals and the proportions calculated for each status/cause according to whether the same status or cause was or was not represented in the output from both an individual's original and follow-up interviews, within each country.

The individual pregnancy status and cause of death outputs were then combined into overall pregnancy status fractions and CSMFs for the population samples in Burkina Faso and Indonesia respectively, which were then compared in terms of magnitude and rank order.

Ethical approvals for the verbal autopsy studies in Burkina Faso and Indonesia, including follow-up interviews where needed, were granted by the Ministry of Health National Health Research Ethics Committee (Ouagadougou, Burkina Faso) and Centre MURAZ Institutional Review Board (Bobo-Dioulasso, Burkina Faso); by the Faculty of Public Health Research Ethics Committee at the University of Indonesia (Jakarta, Indonesia).

Results

A total of 207 VA interviews were successfully repeated, 91 in Burkina Faso and 116 in Indonesia. The basic characteristics of these cases (both the deceased women and the interview respondents) are summarised in Table 1. Educational levels among both the deceased women and the VA respondents were much lower in Burkina Faso than in Indonesia, as were the availability of amenities such as piped water and television. Respondents in Indonesia were generally younger and more likely to be female than those in Burkina Faso. In Indonesia, in 86/116 interviews (74%), the same respondent was re-interviewed at follow-up. In Burkina Faso, the identities of the original respondents were not known and so comparison was not possible. The mean recall period from the death to the original interview was 26 months in Burkina Faso and 8 months in Indonesia, and similarly to the follow-up interview 46 months and 29 months respectively.

Kappa statistics were calculated for each InterVA-M indicator in each country to assess the repeatability of the verbal autopsy interview process. The distribution of kappa statistics for all measurable InterVA-M indicators in relation to their positive response rate in each country are shown in Figure 2. Although the InterVA-M model captures a total of 75 indicators, some were either locally non-applicable or received no positive responses in one of the countries. Thus repeatability was measurable for 68 indicators in Burkina Faso and 63 in Indonesia. The mean κ was 0.24 (range -0.09 to 1.00) for Burkina Faso and 0.45 (range -0.03 to 0.92) for Indonesia. In Burkina Faso 30 out of 68 measurable indicators (44%) showed repeatability better than

expected by chance at the $p < 0.01$ level. Full details are shown in Annex A. In Indonesia 52/63 indicators (83%) showed repeatability better than expected by chance at the $p < 0.01$ level. Full details are shown in Annex B.

Since recall period is obviously a potentially important factor in the repeatability of VA interviews, mean kappa statistics were also calculated separately for original interviews falling before or after the median recall time, for each country. In Burkina Faso the median time to first interview was 26 months, and for 51 indicators represented in interviews on both sides of the median, the mean κ was 0.27 for interviews up to and including the median, compared with 0.17 for interviews after the median. Of the 51 indicators, 76% had a lower κ value for interviews beyond the median recall time. In Indonesia, the median time to first interview was 7 months, and for 60 indicators the mean κ was 0.44 for interviews up to and including the median, compared with 0.45 for interviews after the median time.

Table 2 shows individual level agreement on pregnancy status and likely cause(s) of death for the Burkina Faso and Indonesian cases, after interpreting the VA interview material via the InterVA-M model. In Burkina Faso, 18.0% of the cause of death output was concordant between the original and follow-up interviews, while in Indonesia 25.0% was concordant. For pregnancy status, 67.2% and 85.0% respectively were concordant at the individual level.

Table 3 shows aggregated mortality as cause-specific mortality fractions (CSMF), together with ranked causes of death, in Burkina Faso and Indonesia, from the original and follow-up surveys. These results are presented at the population level, for

each of the four series of VA interviews. Applying the Wilcoxon signed-rank test to the results from Burkina Faso and Indonesia gave results of $z = 0.52$ and 0.50 respectively, $p > 0.6$ in both cases. Thus there was no evidence of significant differences between original and follow-up CSMF patterns.

Table 4 shows the top five ranking causes of death and associated CSMFs from the original and follow-up surveys in both Burkina Faso and Indonesia, at the population level. In all four surveys, the top five causes of death accounted for approximately 60% of overall mortality.

Discussion

The concept of VA, from the interview, through interpretation, to population-level results, is a complex one. This study has investigated the repeatability of three different stages of this overall process, which is entirely novel, and capitalised on the use of the InterVA-M interpretative model in order to eliminate any subjective or inter-country variation in the process of interpreting the VA interview material.

Methodological issues

It is difficult to discuss these findings extensively in relation to other work, since very little attention has previously been given to the repeatability of VA. All of our interviews were conducted in real field conditions, typical of settings in which VA is an important tool, and where deaths in hospital are rare. We happened to use material

relating to women of reproductive age, although we have no reason to believe our findings would not apply to other population groups. While it is possible that we may have influenced the conduct of the VA process by, exceptionally, undertaking follow-up interviews, we believe this is unlikely. In all cases the interval between the original and follow-up interviews was between one and two years, which probably minimised any effect associated with recalling the previous interview, even where the same respondent was involved. On the other hand, the additional accumulation of recall time since the death itself may have influenced the way in which original events were remembered, and thus reduced repeatability.

VA respondents, recall and repeatability

In terms of the repeatability of the VA interview stage, it is clear that there were major differences between the interviews undertaken in Burkina Faso and Indonesia, with a markedly higher mean kappa statistic and proportion of indicators with non-chance agreement in Indonesia (Figure 2). We cannot say definitively what factors lie behind this major difference, although the marked differences observed in the characteristics of those who had died and of the interview respondents (Table 1) are putative factors. In particular, the lower educational levels, greater age and higher proportion of men among Burkinabé respondents could have resulted in a higher proportion of interviews in which the respondent did not clearly understand the processes that had led to death, with this lack of understanding reflected in inconsistencies between the original and follow-up interview material. However, the difference between the two countries was also confounded by different recall periods, and the analyses of kappa by recall time suggest that the effect of recall bias was

much more pronounced in the longer recall periods experienced in Burkina Faso. From these findings it might be reasonable to conclude that VA interviews should, where possible, be undertaken within two years of death.

When the individual-level pregnancy status and cause of death outputs were compared between the original and follow-up interviews, there was a disappointing lack of concordance, though this was perhaps not surprising given the extent of non-agreement between the interview material, particularly from Burkina Faso (Table 2). This finding supports the view that VA material may not be particularly well suited to individual-level cause of death determination [10].

However, when the output data were considered at the population level, there was a much clearer sense of agreement between the original and follow-up material (Table 3). As might be expected from the repeatability results at the first stage, there were still greater discrepancies in the Burkina Faso results compared with those from Indonesia. There were however, as would be expected, marked differences persisting in the overall patterns between the two countries, counteracting any suggestion that the whole VA process might amount to some kind of reduction to lowest common factors.

Implications of VA repeatability for health policy

Taking a public health perspective, Table 4 considers repeatability of the VA process in the context of the often-asked question “what are the major causes of mortality?”. It is clear that, for women both in Burkina Faso and Indonesia, there was very good

repeatability between the original and follow-up VAs in terms of generating summary information for health policy and planning, even though the specific country findings were, as expected, different.

The relative levels of repeatability, both between the two country settings involved and between the different stages of the VA process, are interesting. The obvious differences between the two countries make clear that the whole VA process is context-dependent, since it might reasonably be inferred that in a context where repeatability is lower, then the intrinsic validity of one-time interview material would also be lower. The effects of non-repeatability at different stages of the VA process are also interesting – our results suggest that, given moderate repeatability within the original interviews, the repeatability of cause of death at the individual level is seriously compromised. However, when a reasonable group of individual VAs are taken as aggregate entities (around 100 per setting in these results) then our findings suggest that there is some recovery of repeatability, certainly to levels that appear to be acceptable in terms of generating aggregate data for health planning. Exactly what factors lie behind this effect is not easy to understand, but it seems likely that some of the non-repeatability of details at the interview stage may have the effect of tipping the balance between possible causes at the individual level, but with many such differences then cancelling out on aggregating causes to the population level.

Why repeatability?

Repeatability is often not explicitly assessed for health measurement tools – so why is it important in relation to VA and what do the results tell us? The origins of VA – as a

proxy source of cause of death data in the absence of medical certification – has sometimes led to confusion about its fundamental nature. While it is not generally argued that VA should be used as a direct replacement for medical certification at the individual level, this principle has not always been made explicit in VA work. Consequently, attempts to validate VA [9-13] have tended to struggle for lack of clear, appropriate gold standards and methods, and the role of VA has to some extent been left in an uncertain position. At the same time, the rigour of medical certification of death has rarely been critically evaluated [17].

We therefore offer these analyses of VA repeatability as a fresh viewpoint on the overall process, to shed some light on the practice of VA and the value of the ensuing results. This partly follows from our previous consideration of the question “Who needs cause of death data?” [18], since VA cannot be considered as reducible to a single one-size-fits-all tool, and must be contextualised.

At least in contexts emphasising community-level cause-specific mortality, the findings of these repeatability analyses are encouraging. Although good repeatability does not guarantee good validity, at least it suggests that intrinsic validity is not compromised by random effects in the overall process. It also became clear that the longer recall periods associated with some of the Burkinabé interviews were detrimental to repeatability – and so presumably to validity. At the same time, overall repeatability was lower in Burkina Faso than in Indonesia, possibly because of different respondent profiles, which again emphasises the importance of considering VA material contextually, rather than simply in terms of standardised methods.

Conclusions

The overall process of VA, from interview to CSMF, has been shown to be highly repeatable in two very different settings. However, VA was less repeatable at the individual level, and recall periods beyond two years compromised repeatability. Although repeatability does not demonstrate validity, it is a prerequisite, and so this study adds new understanding to the quest for reliable cause of death assessment in communities lacking routine medical certification of deaths.

Authors' contributions

PB devised the study; all authors participated in field work; MO and SNQ coordinated work in Burkina Faso and Indonesia respectively and prepared datasets; PB & LD carried out analysis and drafting.

Acknowledgements

The authors would like to thank the families who participated and the Burkina Faso and Indonesian study groups for their assistance in the conduct of this evaluation. This work was undertaken as part of Immpact, funded by the Bill & Melinda Gates Foundation, the UK Department for International Development, the European Commission and USAID. Immpact is an international research programme which also provides technical assistance through its affiliate organisation, Ipact. The funders have no responsibility for the information provided or views expressed in this paper. The

views expressed herein are solely those of the authors. We are grateful to Professor Wendy Graham for comments on an earlier draft.

References

1. Garenne M, Faveau V. (2006) Potential and limits of verbal autopsies. *Bull World Health Organ* 84:164.
2. Garenne M, Fontaine O. (1986) Assessing probable causes of deaths using a standardized questionnaire: a study in rural Senegal. Reprinted (2006) in *Bull World Health Organ* 84:248-253.
3. Baiden F, Bawah A, Biai S, Binka F, Boerma T, Byass P et al. (2007) Setting international standards for verbal autopsy. *Bull World Health Organ* 85: 569–648.
4. Reeves BC, Quigley M (1997) A review of data-derived methods for assigning causes of death from verbal autopsy data. *Int J Epidemiol* 26: 1080–1089.
5. Quigley MA, Chandramohan D, Rodrigues LC (1999) Diagnostic accuracy of physician review, expert algorithms and data-derived algorithms in adult verbal autopsies. *Int J Epidemiol* 28: 1081–1087.
6. Boulle A, Chandramohan D, Weller P (2001) A case study of using artificial neural networks for classifying cause of death from verbal autopsy. *Int J Epidemiol* 30: 515–520.
7. Byass P, Lan Huong D, Van Minh H (2003) A probabilistic approach to interpreting verbal autopsies: methodology and preliminary validation in Vietnam. *Scand J Public Health* 31: 32–37.
8. Byass P, Fottrell E, Huong DL, Berhane Y, Corrah T, et al. (2006) Refining a probabilistic model for interpreting verbal autopsy data. *Scand J Public Health* 34: 26–31.

9. Fantahun M, Fottrell E, Berhane Y, Wall S, Hogberg U, Byass P (2006) Assessing a new approach to verbal autopsy interpretation in a rural Ethiopian community: the InterVA model, *Bull World Health Organ* 84: 204-210.
10. Murray CJL, Lopez AD, Feehan DM, Peter ST, Yang G (2007) Validation of the symptom pattern method for analyzing verbal autopsy data. *PLoS Med* 4(11): e327.
11. Quigley MA, Chandramohan D, Setel P, Binka F, Rodrigues LC (2000) Validity of data-derived algorithms for ascertaining causes of adult death in two African sites using verbal autopsy. *Trop Med Int Health* 5: 33–39.
12. Kahn K, Tollman SM, Garenne M, Gear JSS (2000) Validation and application of verbal autopsies in a rural area of South Africa. *Trop Med Int Health* 5: 824–831.
13. Setel PW, Rao C, Hemed Y, Whiting DR, Yang G, Chandramohan D, Alberti K, Lopez AD (2006) Core verbal autopsy procedures with comparative validation results from two countries. *PLoS Med* 3: e268.
14. Fottrell E, Byass P, Ouedraogo TW, Tamini C, Gbangou A, Sombie I, Hogberg U, Witten KH, Bhattacharya S, Desta T, Deganus S, Tornui J, Fitzmaurice AE, Meda N, Graham WJ. (2007) Revealing the burden of maternal mortality: a probabilistic model for determining pregnancy-related causes of death from verbal autopsies, *Pop Health Metrics* 5:1.
15. Bell JS, Ouedraogo M, Ganaba R, Sombie I, Byass P, Baggaley RF, Filippi V, Fitzmaurice AE, Graham WJ. The epidemiology of pregnancy outcomes in rural Burkina Faso. *Tropical Medicine and International Health* 2008, 13 suppl 1:31-43.

16. Bell J, Qomariyah SN. Immpact - Tools & Methods: Selected findings on maternal mortality. Presentation at Immpact International Symposium 'Delivering Safer Motherhood: Sharing the Evidence' February 2007; London. www.immpact-international.org/uploads/files/Session19_section3_Immpact_JBell_MM.pdf
17. Lahti RA, Penttilä A (2003) Cause-of-death query in validation of death certification by expert panel; effects on mortality statistics in Finland, 1995. *Forensic Sci Int* 131: 113–124.
18. Byass P (2007) Who needs cause-of-death data? *PLoS Med* 4(11): e333.

Table 1: Background characteristics for verbal autopsy interviews in Burkina Faso (n=91) and Indonesia (n=116). Selected factors recorded at follow-up interviews, for women who had died and primary verbal autopsy interview respondents, are shown.

subject	characteristic	level	%	
			Burkina Faso	Indonesia
deceased woman	education	primary	2.2	62.9
		secondary/higher	0	29.3
	work for income	yes	45.1	37.9
	place of death	health facility	51.7	45.7
	water supply	piped	1.1	30.2
	television	yes	2.2	50.0
respondent	sex	female	26.4	57.8
	age	under 40 yrs	30.8	44.0
	education	primary	8.8	62.1
		secondary/higher	2.2	11.3
	work for income	yes	90.1	62.1

Table 2: Individual repeatability of pregnancy status and cause-of-death assessment from repeated verbal autopsy interviews in Burkina Faso (n=91) and Indonesia (n=116)

Likely cause of death	Burkina Faso		Indonesia	
	% of interviews leading to this likely cause/status	of which, % with cause/status found in both interviews	% of interviews leading to this likely cause/status	of which, % with cause/status found in both interviews
Abortion related death	2.7	5.8	2.7	10.8
Anaemia	5.2	0	5.2	14.9
Cancer	2.9	20.3	2.9	20.5
Cardiovascular disease	5.7	5	5.7	24.8
Diabetes	4.1	19	4.1	35.2
Ectopic pregnancy	0.8	0	0.8	10.4
HIV/AIDS related death	5.0	9.9	5.0	0
Haemorrhage	8.8	17.2	8.8	28.4
Indeterminate	1.2	100.0	1.2	100.0
Kidney disease	2.0	13.6	2.0	18.7
Liver disease	1.9	0	1.9	23.3
Malaria	24.1	24.1	24.1	30.2
Non-pregnancy related infection	0	0	0	0
Obstructed labour	2.1	6.6	2.1	12.1
Other maternal cause	0.8	0	0.8	23.9
Pregnancy-induced hypertension	3.5	0	3.5	11.5
Pregnancy-related infection	20.8	29.5	20.8	22.1
Respiratory disease	0.6	0	0.6	0
Ruptured uterus	1.1	0	1.1	0
Suicide	0.9	0	0.9	35.3
Tuberculosis (pulmonary)	5.7	10.5	5.7	27.0
OVERALL	100.0	18.0	100.0	25.0
Pregnancy status				
not pregnant within 6 wks of death	6.7	66.0	8.0	82.2
pregnant within 6 wks of death	70.8	78.0	51.5	79.1
pregnant at death	22.7	34.0	40.5	90.1
OVERALL	100.0	67.2	100.0	85.0

Table 3: Overall cause-specific mortality fractions and ranked causes, together with pregnancy status assessments, from verbal autopsies in Burkina Faso (n=91) and Indonesia (n=116), in original and follow-up interviews

	Burkina Faso				Indonesia			
	original interview		follow-up interview		original interview		follow-up interview	
cause of death	%	rank	%	rank	%	rank	%	rank
Abortion related death	3.6	10	1.8	11	1.3	17	1.9	15
Anaemia	6.8	5	3.6	7	5.4	7	4.7	8
Cancer	2.2	13	3.6	7	2.9	11	5.2	7
Cardiovascular disease	8.4	3	3.0	9	7.0	6	5.5	5
Diabetes	4.1	9	4.1	5	4.5	9	5.3	6
Ectopic pregnancy	0.3	19	1.4	14	1.8	15	2.9	13
HIV/AIDS related death	5.4	7	4.6	5	0.3	20	1.8	16
Haemorrhage	8.2	4	9.4	3	17.2	1	11.5	2
Indeterminate	1.2	17	0	18	2.8	12	0.9	18
Kidney disease	2.4	12	1.5	13	3.3	10	3.0	12
Liver disease	1.6	15	2.1	10	5.2	8	3.3	11
Malaria	19.6	1	30.0	1	8.8	4	7.8	4
Non-pregnancy related infection	0	20	0	18	0	21	0.9	18
Obstructed labour	2.6	11	1.7	12	2.6	13	2.0	14
Other maternal cause	1.6	16	0	18	2.1	14	3.6	10
Pregnancy-induced hypertension	5.8	6	1.2	15	8.3	5	3.7	9
Pregnancy-related infection	18.5	2	23.2	2	11.8	2	24.2	1
Respiratory disease	0	20	1.2	17	0.3	19	0	21
Ruptured uterus	2.2	14	0	18	1.6	16	1.0	17
Suicide	0.6	18	1.2	15	1.0	18	0.4	20
Tuberculosis (pulmonary)	5.0	8	6.4	4	11.8	3	10.6	3
Pregnancy status								
not pregnant within 6 wks of death	6.8	3	21.5	2	8.2	3	7.8	3
pregnant within 6 wks of death	70.9	1	67.0	1	51.4	1	51.6	1
pregnant at death	22.3	2	11.5	3	40.4	2	40.6	2

Table 4: Top five causes of death from original and follow-up interviews in series of verbal autopsies in Burkina Faso (n=91) and Indonesia (n=116).

	Burkina Faso				Indonesia			
	original interview		follow-up interview		original interview		follow-up interview	
rank	cause of death	CSMF %	cause of death	CSMF %	cause of death	CSMF %	cause of death	CSMF %
1	malaria	19.6	malaria	30.0	haemorrhage	17.2	pregnancy-related infection	24.2
2	pregnancy-related infection	18.5	pregnancy-related infection	11.8	pregnancy-related infection	11.8	haemorrhage	11.5
3	cardio-vascular disease	8.4	haemorrhage	9.4	tuberculosis (pulmonary)	11.8	tuberculosis (pulmonary)	10.6
4	haemorrhage	8.2	tuberculosis (pulmonary)	6.4	malaria	8.8	malaria	7.8
5	anaemia	6.4	HIV/AIDS related death	4.6	pregnancy-induced hypertension	8.3	cardio-vascular disease	5.5
total		61.1		62.2		57.9		59.6

Legends to Figures

Figure 1: Conceptual framework for assessing the repeatability of verbal autopsy interviews

Figure 2: Distribution of kappa statistics versus positive response rate for indicators obtained from repeated verbal autopsy interviews in Burkina Faso (n=91) and Indonesia (n=116)

Annex A: Repeatability of VA indicators in Burkina Faso (by ascending κ values within each category)

category	indicator	% positive responses	% observed agreement	% expected agreement	κ	p for observed > expected
background	married	91.8	85.7	84.6	0.074	0.209
	20-34 yrs at death	51.1	63.7	49.7	0.104	0.004
	under 20 yrs at death	13.7	83.5	76.2	0.309	0.001
	died in wet season	30.2	76.9	57.7	0.455	0.000
	35+ years at death	35.2	75.8	54.4	0.470	0.000
pregnancy status	had ever been pregnant	56.0	14.3	14.0	0.003	0.348
	early pregnancy ended	6.6	89.0	87.6	0.115	0.121
	not pregnant at death	24.2	71.4	60.9	0.269	0.001
	pregnant at death	20.9	78.0	67.0	0.335	0.001
	breast feeding at death	24.2	80.2	63.3	0.461	0.000
	delivered within 6 wks	47.3	73.6	49.3	0.480	0.000
	first pregnancy	14.8	87.9	74.4	0.527	0.000
> 4 pregnancies	44.0	78.0	50.1	0.559	0.000	
clinical history	HIV/AIDS diagnosis	3.3	93.4	93.6	-0.030	0.620
	suspected suicide	1.6	96.7	96.8	-0.015	0.560
	recent injury	1.6	96.7	96.8	-0.015	0.560
	malaria diagnosis	17.6	71.4	71.0	0.015	0.443
	liver disease diagnosis	9.3	81.3	82.0	0.040	0.737
	CVD diagnosis	12.1	80.2	78.4	0.086	0.185
	had a previous C-sect	2.7	96.7	94.6	0.389	0.000
	TB diagnosis	2.2	97.8	95.7	0.491	0.000
	malignancy diagnosis	1.1	100.0	97.8	1.000	0.000
signs and symptoms before death	persistent fever	15.9	70.3	72.7	-0.088	0.815
	blurred vision	25.8	59.3	61.4	-0.053	0.698
	breathless	53.8	40.7	42.5	-0.032	0.661
	pallor or anaemia	39.6	51.7	51.8	-0.003	0.512
	acute fever	24.7	57.1	56.9	0.004	0.476
	swollen body	10.4	81.3	81.0	0.017	0.432
	abdominal pain	24.2	64.8	63.2	0.043	0.339
	collapse	31.9	53.9	56.5	0.061	0.719
	in bed during day	27.5	62.6	60.1	0.063	0.274
	jaundice	20.3	70.3	67.1	0.098	0.164
	fever with shivering	25.3	64.8	59.3	0.136	0.054
	night sweats	39.0	61.5	52.3	0.193	0.033
	recurrent fever	25.3	69.2	61.8	0.194	0.029
	coma >24h	16.5	78.0	72.3	0.208	0.021
	foul-smelling discharge	4.4	93.4	91.6	0.216	0.020
	stiff neck	11.0	84.6	80.3	0.217	0.017
	swollen feet	20.3	74.7	67.6	0.220	0.018
	swollen glands	4.4	93.4	91.5	0.224	0.006
	weight loss	41.8	62.6	50.5	0.245	0.008
swollen face	12.6	83.5	77.6	0.263	0.004	

	persistent cough	12.6	85.7	77.9	0.355	0.000
	coughing blood	2.7	96.7	94.7	0.384	0.000
	transfusion needed	30.2	81.3	57.8	0.557	0.000
	ever had fits	12.1	91.2	78.4	0.594	0.000
obstetric record	unwanted pregnancy	23.1	56.0	59.6	-0.074	0.882
	died in labour	6.0	87.9	88.5	-0.050	0.708
	abnormal baby position	12.1	78.0	78.7	-0.033	0.624
	assisted delivery	3.8	92.3	92.5	-0.019	0.605
	baby part prolapsed	2.2	95.6	95.7	-0.017	0.573
	hysterectomy	1.6	96.7	96.8	-0.015	0.560
	big baby	11.5	79.1	78.9	0.013	0.442
	fits during pregnancy	14.8	74.7	74.2	0.019	0.423
	BP raised in pregnancy	11.0	82.4	80.1	0.119	0.104
	first trimester death	6.0	90.1	88.5	0.141	0.063
	labour >24h	8.8	86.8	83.9	0.183	0.035
	bleeding at delivery	25.3	69.2	61.8	0.194	0.029
	bleeding in 1 st trimester	7.1	90.1	86.7	0.255	0.007
	delivered at home	20.9	78.0	66.9	0.337	0.001
	attempted termination	2.7	96.7	94.6	0.389	0.000
	delivered in a facility	33.0	78.0	55.7	0.504	0.000
	professional care at del	31.3	81.3	56.9	0.563	0.000
	death within 24h of del	23.6	85.7	63.9	0.605	0.000
	multiple pregnancy	3.3	97.8	93.6	0.657	0.000
	placenta retained	6.0	96.7	88.6	0.710	0.000
	baby delivered alive	36.3	86.8	53.8	0.715	0.000
	delivered by C-section	5.5	100.0	89.6	1.000	0.000

Annex B: Repeatability of VA indicators in Indonesia (by ascending κ values within each category)

category	indicator	% positive responses	% observed agreement	% expected agreement	κ	p for observed > expected
background	died in wet season	34.1	73.3	54.8	0.409	0.000
	under 20 yrs at death	7.3	94.0	86.4	0.556	0.000
	married	95.7	96.6	91.8	0.582	0.000
	20-34 yrs at death	59.9	80.2	52.0	0.587	0.000
	35+ years at death	32.3	85.3	56.2	0.665	0.000
pregnancy status	had ever been pregnant	92.2	86.2	85.0	0.055	0.067
	early pregnancy ended	3.9	95.7	92.5	0.422	0.000
	had a previous C-sect	5.2	96.6	90.1	0.651	0.000
	pregnant at death	39.2	85.3	52.2	0.693	0.000
	> 4 pregnancies	27.2	88.8	60.1	0.719	0.000
	first pregnancy	24.6	92.2	62.9	0.791	0.000
	delivered within 6 wks	46.6	89.7	50.1	0.793	0.000
not pregnant at death	10.3	96.6	81.5	0.814	0.000	
clinical history	recent injury	6.5	88.8	87.8	0.081	0.177
	recent surgery	2.6	96.6	95.0	0.316	0.000
	CVD diagnosis	4.7	94.0	90.9	0.339	0.000
	TB diagnosis	8.6	91.4	84.1	0.457	0.000
	liver disease diagnosis	6.0	94.8	88.6	0.545	0.000
signs and symptoms before death	stiff neck	6.0	87.9	88.3	-0.031	0.686
	foul-smelling discharge	8.2	85.3	84.9	0.031	0.362
	antibiotics needed	25.9	63.8	60.2	0.091	0.138
	collapse	18.5	73.3	69.8	0.116	0.105
	blurred vision	13.8	79.3	76.2	0.132	0.075
	pallor or anaemia	50.4	56.0	49.2	0.135	0.066
	persistent fever	7.3	88.8	86.2	0.186	0.013
	night sweats	36.6	64.7	53.5	0.239	0.005
	in bed during day	37.9	67.2	51.1	0.330	0.000
	abdominal pain	24.1	75.9	63.2	0.343	0.000
	coma >24h	12.1	86.2	78.5	0.357	0.000
	jaundice	16.8	81.9	71.7	0.360	0.000
	acute fever	33.2	73.3	55.7	0.398	0.000
	swollen body	7.8	91.4	85.7	0.398	0.000
	fever with shivering	21.1	80.2	66.7	0.405	0.000
	recurrent fever	22.4	79.3	65.2	0.405	0.000
	weight loss	37.1	72.4	53.0	0.413	0.000
	persistent cough	16.8	83.6	72.0	0.415	0.000
	swollen glands	5.6	94.0	89.4	0.430	0.000
	breathless	24.1	79.3	63.4	0.435	0.000
	swollen feet	25.0	79.3	62.5	0.449	0.000
	transfusion needed	19.0	84.5	69.3	0.495	0.000
ever had fits	13.4	88.8	76.8	0.516	0.000	
coughing blood	7.8	93.1	85.7	0.519	0.000	

	swollen face	7.8	94.8	85.7	0.639	0.000
obstetric record	unwanted pregnancy	28.0	59.5	59.6	-0.004	0.516
	bleeding in 1 st trimester	6.0	89.7	88.6	0.089	0.166
	labour >24h	6.0	91.4	88.6	0.244	0.003
	died in labour	4.3	94.8	91.7	0.374	0.000
	big baby	7.8	91.4	85.7	0.398	0.000
	death within 24h of del	14.2	85.3	75.5	0.402	0.000
	bleeding at delivery	23.7	78.5	63.8	0.404	0.000
	fits during pregnancy	7.3	92.2	86.4	0.429	0.000
	placenta retained	8.6	91.4	84.2	0.455	0.000
	baby part prolapsed	3.0	97.4	94.1	0.558	0.000
	first trimester death	8.6	93.1	84.2	0.564	0.000
	abnormal baby position	4.3	96.6	91.7	0.585	0.000
	BP raised in pregnancy	19.4	90.5	68.7	0.697	0.000
	baby delivered alive	44.4	87.1	50.6	0.738	0.000
	professional care at del	26.3	90.5	61.2	0.756	0.000
	multiple pregnancy	5.6	97.4	89.4	0.756	0.000
	delivered in a facility	19.0	93.1	69.3	0.776	0.000
	delivered at home	39.7	91.4	52.1	0.820	0.000
	assisted delivery	6.0	98.3	88.7	0.848	0.000
	delivered by C-section	5.6	99.1	89.4	0.919	0.000

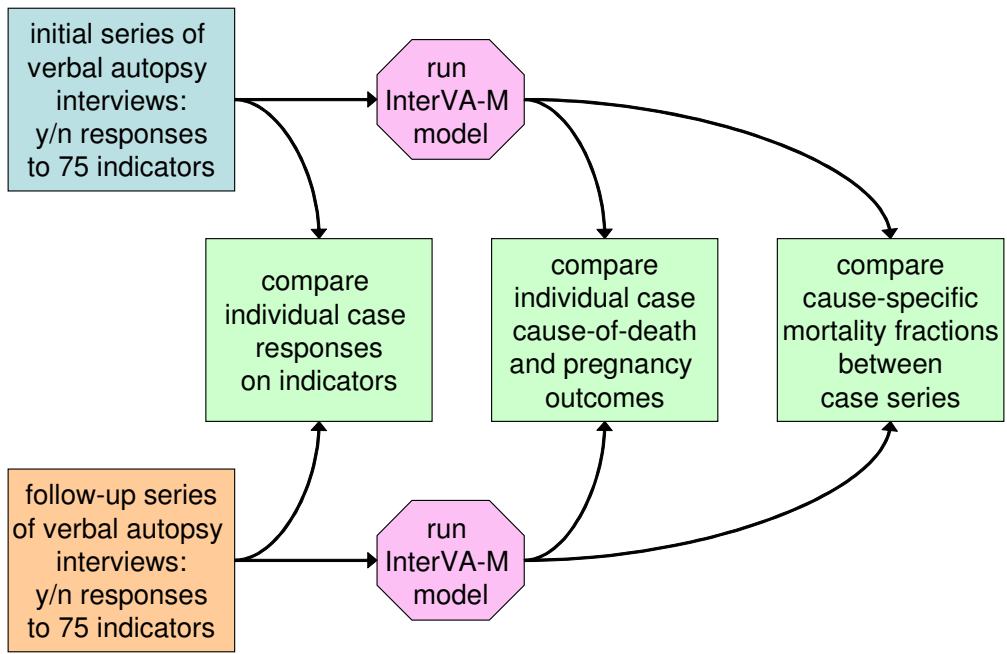
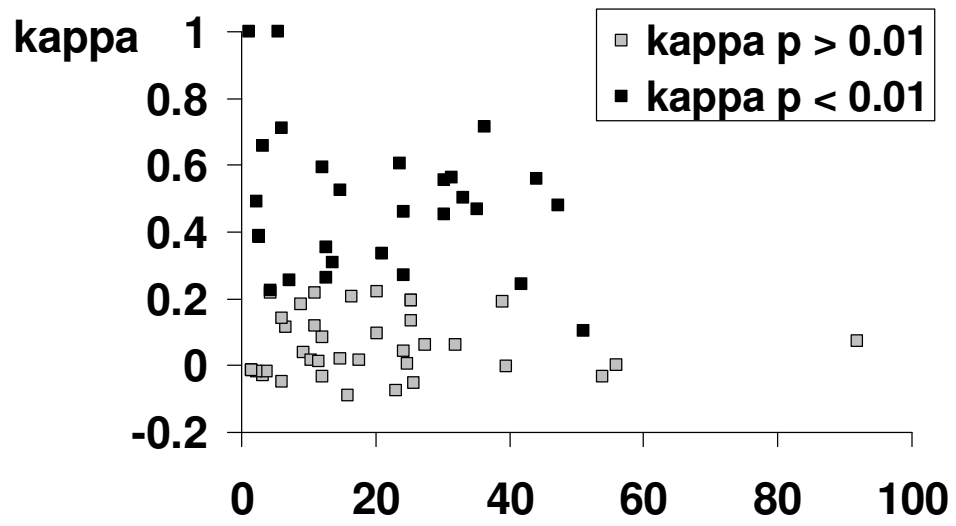


Figure 1

Burkina Faso



Indonesia

