 International Journal of Epidemiology 2001;**30**:509–514

THEORY AND METHODS

Effect of misclassification of causes of death in verbal autopsy: can it be adjusted?

Daniel Chandramohan,[a] Philip Setel[b] and Maria Quigley[a]

Background Verbal autopsy (VA) is an indirect method of ascertaining cause of death from information about symptoms and signs obtained from bereaved relatives. This method has been used in several settings to assess cause-specific mortality. However, cause-specific mortality estimates obtained by VA are susceptible to bias due to misclassification of causes of death. One way of overcoming this limitation of VA is to adjust the crude VA estimate of cause-specific mortality fractions (CSMF) using the sensitivity and specificity of the VA tool. This paper explores the application of sensitivity and specificity of VA data obtained from a hospital-based validation study for adjusting the effect of misclassification error in VA data obtained from a demographic surveillance system.

Method Data from a multi-centre validation study of 796 adult VA, conducted in Tanzania, Ethiopia and Ghana, were used to explore the effect of distribution of causes of death in the validation study population and the pattern of misclassification on the sensitivity and specificity of VA. VA estimates of CSMF for six causes (acute febrile illness, diarrhoeal diseases, TB/AIDS, cardiovascular disorders, direct maternal causes and injures) were obtained from a demographic surveillance system in Morogoro Rural District in Tanzania. These were adjusted for misclassification error by using sensitivity and specificity values of VA obtained from the validation study in a model proposed for correcting the effect of misclassification error in morbidity prevalence surveys.

Results Sensitivity and specificity of VA differed between the three validation study sites depending on the distribution of causes of death. These differences were explained by variations in the level and pattern of misclassification between sites. When these estimates of sensitivity and specificity were applied to data from the demographic surveillance system with a comparable structure of causes of death the difference between crude and adjusted VA estimates of CSMF ranged from 3 to 83%.

Conclusion Estimates of sensitivity and specificity obtained from hospital-based validation studies must be used cautiously as a *de facto* ‘gold standard’ for adjusting the misclassification error in CSMF derived from VA. It is not possible to use sensitivity and specificity estimates derived from a location-specific validation study to adjust for misclassification in VA data from populations with substantially different patterns of cause-specific mortality.

Keywords Verbal autopsy, sensitivity, specificity, adjusting, misclassification

Accepted 24 April 2000

[a] Department of Infectious and Tropical Diseases, London School of Hygiene & Tropical Medicine, Keppel Street, London WC1E 7HT, UK. E-mail: daniel.chandramohan@lshtm.ac.uk

[b] Adult Mortality and Morbidity Project *Tanzanian Ministry of Health and University of Newcastle upon Tyne Medical School, PO Box 65243, Dar es Salaam, Tanzania.

* This publication is, in part, an output of a project funded by the UK Department for International Development for the benefit of developing countries. The views expressed are not necessarily those of DFID.

The verbal autopsy (VA) is an indirect method of ascertaining biomedical causes of death (COD) from information on symptoms, signs and the circumstances preceding death, obtained from a caretaker of a deceased. This method has been used in several settings to estimate cause-specific mortality of childhood, maternal and adult deaths. It is often the only source of cause-specific mortality in settings lacking functioning vital registration systems. The fact that the VA is used in situations where better information is unavailable raises several related questions including: (1) how to assess the validity and reliability of cause-specific mortality estimates obtained from the VA; (2) whether misclassification error can be estimated and adjusted for in VA data; and (3) how to carry out this adjustment. This article concerns the issue of misclassification in the VA and whether and how it can be adjusted.

The validity and reliability of VA estimates of cause-specific mortality depend on several factors such as the 'true' underlying distribution of COD in the population, the age and sex of the deceased, the specific VA tools used, and the data collection process.[1] VA estimates of cause-specific mortality fractions (CSMF) can be inaccurate if the sensitivity and specificity of the VA are <100%.[2] One way of overcoming this problem is by adjusting the VA estimate of CSMF using the sensitivity and specificity of the VA tool. Assuming that all sources of error in measures of sensitivity and specificity are negligible, then VA estimates of CSMF can be adjusted for the effect of misclassification error using the following model that was suggested for adjusting the effect of misclassifcation error of cause-specific disease prevalence in a morbidity survey[3]:

$$Pt = \frac{(Pe + \text{specificity} - 1)}{\text{sensitivity} + \text{specificity} - 1)} \quad (1)$$

where Pt is the adjusted CSMF and Pe is the crude VA estimate of CSMF. However, it is not clear whether the sensitivity and specificity of VA tools obtained from particular validation studies can be extrapolated to data obtained from demographic surveys or surveillance systems.

The sensitivity and specificity of child VA have been evaluated in Kenya,[4] The Philippines[5] and Namibia,[6] and that of adult VA in Tanzania,[7] Ghana[7] and Ethiopia.[7] All of these studies were hospital-based and they assessed the validity of the VA using the hospital diagnosis as a 'gold standard'. In this paper we explore the limitations of applying sensitivity and specificity of the VA obtained from a hospital-based validation study for adjusting misclassification error in the VA. We also examine the implication of using sensitivity and specificity values obtained in settings where validation studies have been performed to adjust for presumed misclassification in a setting where validation has either not been performed or is impractical.

Methods

We used the data of a multi-centre adult VA validation study to explore the effect of distribution of COD and the patterns of misclassification of the VA on the sensitivity and specificity of VA. The data collection process and the methods of analysis of this study have been reported elsewhere.[7] The study included 796 adult deaths (15+ years) within a 60 km radius of the study hospitals in Ifakara (Tanzania), Jimma (Ethiopia) and Bawku (Ghana). A standard VA tool (mortality classification, VA questionnaire and procedures to derive diagnoses from VA) was used in all three sites and COD from VA questionnaires were reached by the same panel of three physicians in all three sites. The sensitivity and specificity of VA were estimated by comparing the VA diagnosis with the hospital diagnosis. The χ^2 test of association was used to compare sensitivities and specificities between sites. Using the above-mentioned adjustment model, we applied sensitivity and specificity values for six categories of COD in adults (15+ years) obtained from the validation study to VA data collected by the Adult Morbidity and Mortality Project (AMMP) demographic surveillance system in Morogoro Rural District, Tanzania. The VA data collection methods in AMMP have been described elsewhere.[8] Briefly, in the AMMP system, VA interviews are conducted for all incident deaths in a geographically defined population of approximately 100 000. VA interviews are normally conducted within a month after death by clinical officers using a questionnaire similar to the one used in the multi-centre validation study, and COD is determined by a panel of physicians on the project team.

Results

Effects of CSMF on sensitivity and specificity of VA

Since the same VA tool and data collection method were applied in all three sites, the effects of these factors on the validity of the VA would not vary by site. However, the sensitivity and specificity of VA and CSMF did differ between the three sites (Table 1). The sensitivity for acute febrile illness (AFI) was lower and specificity was higher in Ifakara than in Bawku (60% versus 74%; $P = 0.04$ and 94% versus 75%; $P < 0.0001$, respectively) and the CSMF of AFI also differed between these two sites (26% versus 40%; $P = 0.0004$). Similarly, the specificity differed significantly for TB/AIDS and direct maternal causes between Ifakara and Bawku (93% versus 99.5%; $P = 0.0003$ and 99% versus 96%; $P = 0.03$, respectively) and so did their CSMF (24% versus 8%; $P < 0.0001$ and 3% versus 8%; $P = 0.005$, respectively). Between Ifakara and Jimma, the specificity for diarrhoeal diseases (94% versus 98%; $P = 0.02$) differed significantly, and the CSMF of diarrhoea also differed (10% versus 2%; $P < 0.0001$). We used linear regression to estimate the effect of the CSMF on specificity for the 18 CSMF and specificities given in Table 1. There was a significant negative relationship ($\beta = -0.48$; $P < 0.001$); the specificity falls half of one per cent for every one per cent increase in CSMF (Figure 1).

Effects of patterns of misclassification on sensitivity and specificity

The patterns of misclassification of COD in the three sites are shown in Tables 2, 3 and 4. There was no clear pattern of misclassification in COD in all three sites. Misclassification was observed among all COD, with the exception of injuries and to some extent, direct maternal causes. However, the false negative and false positive rates varied among COD and for the same COD between sites. For example, the false negative rate of AFI (the proportion of AFI cases misclassified as other COD) ranged from 0% (0/82) as direct maternal causes to 9% (7/82) as TB/AIDS or cardiovascular system (CVS) disorders in Ifakara; 0% (0/93) as injuries to 5% (5/95) as direct maternal causes in Bawku; and 0% as injuries (0/61) to 16% (10/61) as TB/AIDS in Jimma. Furthermore, the proportion of false negatives of a

Table 1 Frequency distribution of common causes of death (COD) and the sensitivity and specificity of verbal autopsy in Ifakara, Bawku and Jimma

	Distribution of COD[a] (%-CSMF[b])			Sensitivity of verbal autopsy			Specificity of verbal autopsy		
Cause of death	Ifakara	Jimma	Bawku	Ifakara	Jimma	Bawku	Ifakara	Jimma	Bawku
Acute febrile illness	82 (26)	61 (25)	93 (40)[c]	60	67	74[c]	94	93	75[c]
Diarrhoeal diseases	33 (10)	6 (2)[c]	12 (5)[c]	73	67	25	94	98[c]	97
TB/AIDS	75 (24)	55 (22)	18 (8)[c]	76	82	56	93	89	99.5[c]
CVS[d] disorders	25 (8)	16 (6)	24 (10)	40	50	54	95	98	97
Direct maternal causes	10 (3)	22 (9)[c]	18 (8)[c]	90	77	83	99	98	96[c]
Injuries	9 (3)	17 (7)[c]	7 (3)	89	100	100	99	100	99.6
All other and undetermined causes	81 (26)	72 (29)	60 (26)	73	74	65	88	91	90
Total	315	249	232	–	–	–	–	–	–

[a] COD: causes of death in the hospital considered as 'gold standard'.
[b] cause-specific mortality fraction.
[c] $P < 0.05$ compared to Ifakara (baseline group).
[d] Cardiovascular system.

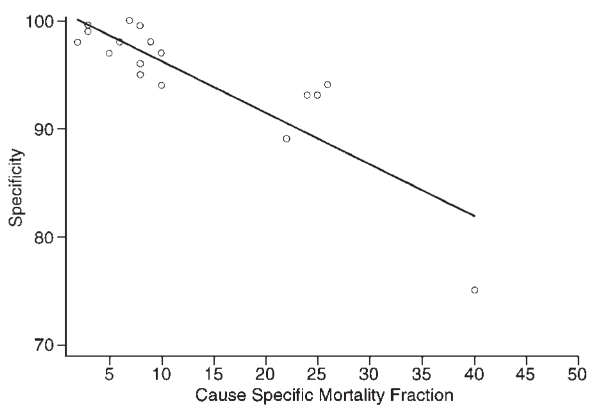

Figure 1 Linear regression of specificity on cause-specific mortality fraction

COD contributed by the same COD differed between the sites. For instance, the proportion of false negatives of AFI as TB/AIDS was 9% (7/82) in Ifakara while it was only 1% (1/93) in Bawku. Thus, the false negative rates (the sum of cause-specific false negative rates) and the sensitivity of VA differed between sites.

The influence of the distribution of COD on the specificity of VA can be explained by the following expression of specificity. Given N possible COD ($COD_1 \ldots COD_N$) then the specificity for a given COD, denoted COD_1, is

$$1 - (M_2P_2 + M_3P_3 + \ldots M_NP_N) \quad (2)$$

where M_i = proportion of true COD_i misclassified as COD_1; and P_i = proportion of true negative COD_1 that are true COD_i. For example, in Ifakara (Table 2) the specificity for AFI equals $1 - \{(3/33)(33/233) + (3/75)(75/233) + (3/25)(25/233) + (0/10)(10/233) + (0/9)(9/233) + (6/81)(81/233)\}$. The misclassification errors associated with diarrhoea, TB/AIDS and CVS disorders each reduce the specificity by 1.3% (i.e. 3/233).

The misclassification error associated with TB/AIDS reduced the specificity of VA for AFI by 1.3% in Ifakara compared to 3% (4/139) in Bawku. In Ifakara, only 4% (3/75) of TB/AIDS deaths were misclassified as AFI, but since 32% (75/233) of true negative cases were TB/AIDS this misclassification error reduced the specificity of VA of AFI by 1.3%. Although 22% (4/18) of TB/AIDS were misclassified as AFI in Bawku this error reduced the specificity by only 3% since TB/AIDS contributed just 13% (18/139) of the true negatives. Thus even if there are no differences in the false positive rates contributed by each COD between sites, the specificity will differ if the CSMF vary

Table 2 Patterns of misclassification error of verbal autopsy in Ifakara: comparison of verbal autopsy diagnosis versus hospital diagnosis

	Cause of death (reference diagnosis)							
Cause of death (VA diagnosis)	Acute febrile illness	Diarrhoeal diseases	TB/AIDS	CVS[a] disorders	Direct maternal causes	Injuries	All other and undetermined causes	Total (VA diagnosis)
Acute febrile illness	**49**	3	3	3	0	0	6	64
Diarrhoeal diseases	6	**24**	3	3	0	0	4	40
TB/AIDS	7	2	**57**	1	0	0	7	74
CVS[a] disorders	7	0	4	**10**	0	0	4	25
Direct maternal causes	0	0	2	0	**9**	1	1	13
Injuries	1	0	1	1	0	**8**	0	11
All other and undetermined causes	12	4	5	7	1	0	**59**	88
Total (reference diagnosis)	82	33	75	25	10	9	81	315

[a] Cardiovascular system.

Table 3 Patterns of misclassification error of verbal autopsy in Jimma: comparison of verbal autopsy diagnosis a versus hospital diagnosis

	Cause of death (reference diagnosis)							
Cause of death (VA diagnosis)	Acute febrile illness	Diarrhoeal diseases	TB/AIDS	CVS[a] disorders	Direct maternal causes	Injuries	All other and undetermined causes	Total (VA diagnosis)
Acute febrile illness	**41**	1	3	2	3	0	4	54
Diarrhoeal diseases	0	**4**	1	0	2	0	2	9
TB/AIDS	10	0	**45**	3	0	0	8	66
CVS[a] disorders	1	0	1	**8**	0	0	3	13
Direct maternal causes	2	0	0	0	**17**	0	2	21
Injuries	0	0	0	0	0	**17**	0	17
All other and undetermined causes	7	1	5	3	0	0	**53**	69
Total (reference diagnosis)	61	6	55	16	22	17	72	249

[a] Cardiovascular system.

Table 4 Patterns of misclassification error of verbal autopsy in Bawku; comparison of verbal autopsy diagnosis versus hospital diagnosis

	Cause of death (reference diagnosis)							
Cause of death (VA diagnosis)	Acute febrile illness	Diarrhoeal diseases	TB/AIDS	CVS[a] disorders	Direct maternal causes	Injuries	All other and undetermined causes	Total (VA diagnosis)
Acute febrile illness	**69**	7	4	6	2	0	16	104
Diarrhoeal diseases	3	**3**	0	0	0	0	3	9
TB/AIDS	1	0	**10**	0	0	0	0	11
CVS[a] disorders	4	1	0	**13**	0	0	1	19
Direct maternal causes	5	0	1	2	**15**	0	1	24
Injuries	0	0	0	1	0	**7**	0	8
All other and undetermined causes	11	1	3	2	1	0	**39**	57
Total (reference diagnosis)	93	12	18	24	18	7	60	232

[a] Cardiovascular system.

because the total false positive rate is a weighted sum of cause-specific false positive rates, where the weights are the proportion of each COD among true negative cases.

The inverse relationship between CSMF and specificity is not related to any particular COD or geographical site. This relationship is probably due to the mathematical operation of VA in which a single COD is assigned to each case. If the CSMF increases then the number of true negatives will decrease proportionately. However, the number of true negatives correctly diagnosed by the VA may decrease disproportionately because the number of false positive and false negative diagnoses of the VA tend to balance. For example, the number of true non-AFI cases correctly diagnosed by VA depends not only on the number of false positive diagnosis of non-AFI as AFI but also on the number of false negative diagnoses of AFI as non-AFI.

The number of false positive and negative cases for each COD is small in our data set and thus one has to be cautious in interpreting the patterns of misclassification. Nevertheless these data show that the pattern of misclassification can be influenced by differences in the distribution of COD and thereby the sensitivity and specificity of the VA.

Adjusting the effect of misclassification error on the VA

If the sensitivity and the specificity of the VA are influenced by the distribution of the true COD then the measures obtained from a validation study are unlikely to be useful to estimate the true CSMF from the VA in settings where the underlying distribution of COD differs from that of the validation study population. We examined this proposition by adjusting the CSMF of AMMP data using formula (1). We applied the sensitivity and specificity obtained from Ifakara, the difference between the adjusted and crude CSMF ranged from –83% to +19% (Table 5). When the sensitivity and specificity obtained from Bawku were applied, the adjustment model returned spurious values for some CSMF; the adjusted AFI mortality was –16.9% and the adjusted mortality fraction for direct maternal causes was –1.7%. It is not clear whether the adjusted CSMF are more accurate than the crude estimates. It is worth noting that Ifakara borders the area where the AMMP data were obtained (Morogoro Rural), yet the adjusted CSMF varied markedly even when we used the sensitivity and specificity values from Ifakara to perform the adjustment.

Discussion

We have shown that the sensitivity and specificity of the VA depend on the COD distribution of the validation study population and that if the COD in the general population differ from the validation study population then the application of sensitivity and specificity to adjust for misclassification error can produce spurious results. All validation studies reported to date are hospital based since, community-based validation studies are almost impossible in areas where only a selective proportion of the population contacts health facilities for serious illness, and which, in turn, are the same areas where VA are needed.

Table 5 Mortality data estimated by verbal autopsy in Morogoro Rural District, Tanzania. Comparison of cause-specific mortality fractions (CSMF) unadjusted and adjusted for misclassification error

Cause of death	Total (N)	Unadjusted CSMF[a]	CSMF adjusted to Ifakara	CSMF adjusted to Jimma	CSMF adjusted to Bawku
Acute febrile illness	1193	16.7	19.8 (+19)	14.1 (–15)	–16.9
Diarrhoeal diseases	1065	14.9	13.3 (–11)	19.9 (+33)	54.3 (+260)
TB/AIDS	1486	20.8	20.1 (–3)	13.9 (–33)	36.7 (+76)
CVS[b] disorders	378	5.3	0.9 (–83)	6.9 (+30)	4.5 (–15)
Direct maternal causes	190	2.7	1.9 (–29)	0.9 (–66)	–1.7
Injuries	436	6.1	5.8 (–5)	6.1 (0)	5.7 (–7)
All other and undetermined causes	2380	33.4	35.1 (+5)	37.5 (+13)	42.5 (+28)

[a] CSMF: cause-specific mortality fractions expressed in %.

[b] Cardiovascular disorders.

Figures in parenthesis are the difference between adjusted and crude CSMF expressed as % of crude CSMF; adjusted CSMF do not add up to 100% due to the differences in CSMF between Morogoro district population and the validation study population.

Furthermore, model (1) operates with the following limitation that the VA estimate of CSMF plus specificity will be more than one if sensitivity plus specificity is more than one; conversely it will be less than one if the latter is less than one. These assumptions of preconditions may not always be fitting. For example, the sensitivity and specificity of the VA for AFI are 74% and 75%, respectively, in Bawku where the CSMF of AFI among the hospital population was 40%. Let us assume that a community mortality survey was carried out in Ghana with this VA tool and the VA estimate of the CSMF of AFI was 20%. If we adjust this estimate using the above sensitivity and specificity, the true CSMF will be –10% according to this model. Conversely, VA estimates of the CSMF of AFI should always be >25% if this VA tool is applied to the data from Ghana.

The sensitivity and the specificity of the VA depend on the distribution of COD. Thus the use of values of sensitivity and specificity of VA obtained from hospital-based validation studies in the proposed model for adjusting the effect of misclassification error will not be appropriate in settings where the distribution of COD differs markedly from the validation study population.

Sensitivity and specificity measure the accuracy of the VA at the individual level. At the population level, the accuracy of the CSMF depends on a complex relationship between sensitivity, specificity and CSMF. Even with low sensitivity or specificity, the crude CSMF will be accurate if the number of false positive and false negative diagnoses is equal. For example, the VA estimate and true CSMF of TB/AIDS were the same (24%) even though the sensitivity and specificity were 76% and 93%, respectively. We argue that validation studies are useful to understand the patterns of misclassification of COD and to identify COD that are likely to have systematic or unbalanced misclassification. Among adult deaths robust estimates of CSMF of injuries, AFI, TB/AIDS, tetanus, and direct maternal causes (abortion, eclampsia, obstructed labour, ante/postpartum haemorrhage and puerperal sepsis) can be obtained using VA.[7,9,10] However, the sensitivity and specificity of the VA obtained from validation studies are unlikely to be useful to adjust for the effect of misclassification error.

In summary, CSMF can differ dramatically across a region, between geographical regions within a single country, and between hospital user and non-user populations within a single area.[10] Thus, not only is there reason to question the validity of applying adjustment parameters derived in one location to VA data from another, the application of the estimates of sensitivity and specificity obtained from hospital-based validation studies must also be used cautiously as a *de facto* 'gold standard' for adjusting the misclassification error in CSMF derived from VA.

KEY MESSAGES

- Sensitivity and specificity of verbal autopsy varies between populations depending on the distribution of cause of death.
- Hospital-based validation studies of verbal autopsy are helpful to understand the patterns of misclassification error in verbal autopsy.
- However, sensitivity and specificity of verbal autopsy estimated from hospital-based validation studies are unlikely to be useful for adjusting the effect of misclassification error in the estimates of cause-specific mortality fractions.

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 Anker M. The effect of misclassification error on reported cause-specific mortality fractions. *Int J Epidemiol* 1997;**26:**1090–96.

3 Kalter H. The validation of interviews for estimating morbidity. *Health Policy Plann* 1992;**7:**30–39.

4 Snow RW, Armstrong JRM, Forster D *et al*. Childhood deaths in Africa: uses and limitations of verbal autopsies. *Lancet* 1992;**340:** 351–55.

5 Kalter HD, Gray RH, Black RE. Validation of postmortem interviews to ascertain selected causes of death in children. *Int J Epidemiol* 1990; **19:**380–86.

6 Mobley CC, Boerma JT, Titus S *et al*. Validation study of verbal autopsy method for causes of childhood mortality in Namibia. *J Trop Pediatr* 1996;**42:**365–69.

7 Chandramohan D, Maude GH, Rodrigues LC, Hayes RJ. Verbal autopsies for adult deaths: their development and validity in a multi-centre study. *Trop Med Int Health* 1998;**3:**436–46.

8 Kitange H, Machibya H, Black J *et al*. Outlook for survivors of childhood in sub-Saharan Africa: adult mortality in Tanzania. *Br Med J* 1996;**312:**216–20.

9 Chandramohan D, Rodrigues L, Maude G, Hayes R. The validity of verbal autopsies for assessing the causes of institutional maternal death. *Stud Family Plann* 1998;**29:**414–22.

10 Quigley M, Chandramohan D, Rodrigues L. Diagnostic accuracy of physician review, expert algorithms and data derived algorithms in adult verbal autopsies. *Int J Epidemiol* 1999;**28:**1081–87.

11 *Adult Morbidity and Mortality Project*, Phase I report, Ministry of Health & DFID, Dar es Salaam, 1999.