Tropical Medicine and International Health

VOLUME 5 NO 1 PP 33–39 JANUARY 2000

Validity of data-derived algorithms for ascertaining causes of adult death in two African sites using verbal autopsy

Maria A. Quigley[1], Daniel Chandramohan[1], Philip Setel[2], Fred Binka[3] and Laura C. Rodrigues[1]

1 *London School of Hygiene and Tropical Medicine, London, UK*
2 *Adult Morbidity and Mortality Project, Ministry of Health, Dar es Salaam, Tanzania*
3 *Navrongo Health Research Centre, Navrongo, Ghana*

Summary

BACKGROUND Verbal autopsy (VA) is used to ascertain causes of death using information obtained from bereaved relatives. Causes of death can be ascertained from VA questionnaires by a panel of physicians or from predefined algorithms. In a previous study, we developed data-derived algorithms using VA data from 796 adult deaths in hospitals in Tanzania, Ethiopia, and Ghana (primary sites). These computerized algorithms accurately estimated the cause-specific mortality fractions (CSMFs) for deaths due to injuries, meningitis, TB/AIDS and diarrhoeal diseases in the primary sites. Since the same data were used to generate and to validate the algorithms, the accuracy of our algorithms may have been overestimated. We report here on the validity of the algorithms when they were applied to VA data from two secondary sites in Ghana and Tanzania. Here, 'validity' is taken to mean the degree to which the algorithms replicated the physician-generated CSMF for major causes of death, when applied to the same VA data.

METHODS VA interviews were conducted in two secondary sites: in Navrongo, Ghana, on 406 adult deaths, where three local physicians independently reviewed the questionnaires and assigned a cause of death. In Morogoro, Tanzania, VA interviews were conducted on 209 adult deaths, and a panel of physicians independently reviewed the VA questionnaires together with the hospital death certificates or hospital records to determine the cause of death. The CSMF obtained using each algorithm was compared with the CSMF obtained using physician review.

RESULTS For injuries and meningitis, the algorithms and physician review estimated a similar CSMF in the Morogoro and Navrongo data. For TB/AIDS, the algorithm estimated a similar CSMF as the physicians in Morogoro. The algorithm for diarrhoeal diseases did not agree closely with the physicians in Morogoro or Navrongo.

CONCLUSIONS In general, our data-derived algorithms for assigning causes of death due to injuries, meningitis, and TB/AIDS estimated a similar CSMF as the physicians in the secondary sites. Recommendations for further validation and refinement are discussed. Computerized algorithms offer a potentially quick, affordable, and feasible method for assigning causes of death in mortality surveillance or studies using VA.

keywords verbal autopsy, mortality, validity, algorithm, data-derived, Africa

correspondence Maria Quigley, MRC Tropical Epidemiology Group, London School of Hygiene and Tropical Medicine, Keppel Street, London WC1E 7HT, UK. E-mail: maria.quigley@lshtm.ac.uk

Introduction

In many developing countries, the majority of deaths occur at home, and consequently, accurate information on cause-specific mortality is scarce. The lack of such basic population health data, particularly in sub-Saharan Africa, places health planners and policy makers at a great disadvantage (Cooper *et al*. 1998). In such circumstances, a cause of death may be ascertained using information obtained from bereaved relatives; this approach is called the 'verbal autopsy' (VA). A cause of death may be assigned either by physician review of the VA questionnaires, or by following a set of predefined diagnostic criteria given in an algorithm (Figure 1). In addition to their application in community-based research (Todd *et al*. 1997), the VA technique is being explored as part of a model sentinel surveillance system for the burden of disease

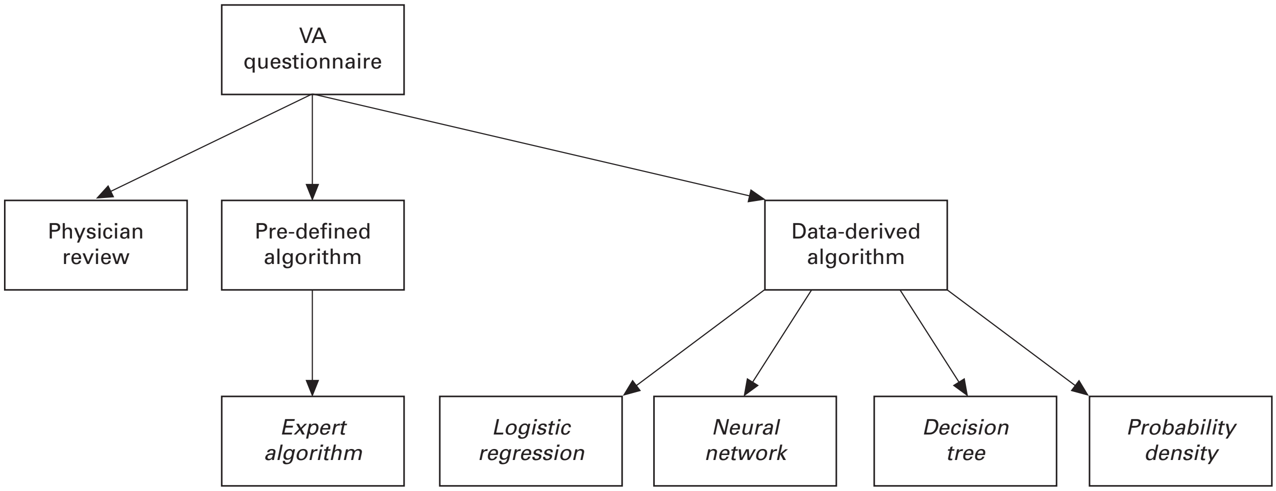

Figure 1 Methods for assigning causes of death using VA data.

in Africa (Ministry of Health & AMMP Team 1997). Such a system could present an alternative to conventional vital registration of death in developing countries. The large number of deaths that must be monitored to make such a system representative means that physician review becomes time-consuming and expensive. Computer algorithm diagnosis of cause of death may offer a cost-effective and sustainable alternative.

In a multicentre VA validation study in Tanzania, Ethiopia, and Ghana, the proportion of deaths due to a particular cause of death (cause-specific mortality fraction; CSMF) obtained using physician review tended to be more accurate than the CSMF estimated using an expert algorithm (Chandramohan *et al*. 1998). Using the data from these primary sites, we employed logistic regression to construct algorithms that most accurately estimated the CSMF for several causes of death (Quigley *et al*. 1999). Our proposed data-derived algorithms accurately estimated the CSMF for deaths due to injuries, meningitis, TB/AIDS, and diarrhoeal diseases. However, the same data were used to derive and validate the algorithms, so the accuracy of our algorithms may be overestimated. We report here on the validity of the data-derived algorithms when they were applied to VA data from two secondary sites in Ghana and Tanzania not included in the original multicentre study.

Methods

Data-derived algorithms

The data-derived algorithms were obtained using data from a multicentre VA validation study conducted in Tanzania, Ethiopia and Ghana (Chandramohan *et al*. 1998). Data from these primary sites included a single cause of death (reached by a review of hospital records and death certificates) and the VA interview data for 796 adult deaths. The methodology for obtaining the data-derived algorithms has been described elsewhere (Quigley *et al*. 1999). Briefly, about half the deaths in the study were used to construct the algorithms (train dataset) and the remaining deaths were used to validate the algorithms (test dataset). The data-derived algorithms that most accurately estimated the CSMF for each cause of death in the train dataset were identified using logistic regression.

The data-derived algorithms for injuries, meningitis, TB/AIDS, and diarrhoeal diseases accurately estimated the CSMF for these causes of death in the test dataset. These algorithms are given in Table 1. The algorithm includes those symptoms that were significantly associated with that cause of death in the logistic model. The algorithm coefficients refer to the natural logarithm of the adjusted odds ratio for each symptom. The algorithms were applied to a given subject in the secondary sites by adding the sum of the coefficients for each symptom that was present in the subject. For example, a subject with a stiff neck and no pallor and a stiff body would have a meningitis score of $2.65 + 1.89 + 1.24 = 5.78$, and would not be classified as meningitis (score < 5.85). However, a subject with a stiff neck and no pallor and continuous fever would have a meningitis score of $2.65 + 1.89 + 1.49 = 6.03$, and would be classified as meningitis (score ≥ 5.85).

We combined TB and AIDS in the same algorithm because these two diseases often occur in the same subject and have

Table 1 Adjusted odds ratio (OR) and coefficient for symptoms in the algorithms for selected causes of death

Symptom	OR	Coefficient
Injuries		
Injury **AND** type of injury **NOT** 'dog bite' **AND** had injury for ≤ 30 days	—	—
Meningitis if NOT suicide/injury **AND** score ≥ 5.85		
Stiff neck	14.09	2.65
No cough	7.19	1.97
No pallor	6.65	1.89
Continuous fever	4.42	1.49
Stiff body	3.44	1.24
Tuberculosis/AIDS if **NOT** suicide/injury **AND** score ≥ 3.59		
Had tuberculosis or HIV/AIDS	23.36	3.15
No vomiting or vomiting for > 1 day	4.19	1.43
Pass stools at least 4 times per day	3.70	1.31
Weight loss	3.13	1.14
Cough for at least 3 weeks	2.77	1.02
[model 1a]		
Diarrhoeal diseases if **NOT** suicide/injury **AND** score ≥ 11.52		
No cough	34.51	3.54
Non-specific* abdominal pain	20.92	3.04
Change in amount of urine for 2 days	12.48	2.52
No distended abdomen	12.42	2.52
Age ≥ 55 years	10.31	2.33
Diarrhoea for 4–30 days	8.76	2.17
Diarrhoea	7.39	2.00
Cramps in abdomen	3.62	1.29
[model 1b]		
Diarrhoeal diseases if **NOT** suicide/injury **AND** score ≥ 7.56		
No cough	17.30	2.85
Diarrhoea	11.22	2.42
Age ≥ 55 years	9.85	2.29
Diarrhoea for 4–30 days	8.13	2.10

The coefficients are the log(OR) for symptoms in the model and are used to obtain a score. For example, looking at meningitis, if the subject has a stiff neck and no pallor and a stiff body then the score is the sum of the coefficients for these symptoms (2.65 + 1.89 + 1.24 = 5.78), which is less than 5.85, so the subject is not classified as a meningitis death. * refers to abdominal pain that is not cramps, dull ache, or burning pain.

many symptoms in common, so it is unlikely that the VA technique will distinguish between them. One of the secondary sites (Morogoro, Tanzania) used a different VA questionnaire to the one used in the primary sites, so data on some of the symptoms in the data-derived algorithms were in a form not amenable to computer algorithm coding. For this reason, the following algorithms were not applied to the Morogoro data: injuries and diarrhoeal diseases (model 1a) (no information on type of abdominal pain, change in amount of urine, abdominal distension, and cramps in abdomen). The following algorithms were applied to the Morogoro data with minor modifications: TB/AIDS (we used presence of diarrhoea rather than frequency of diarrhoea) and meningitis (pallor was dropped from the algorithm, and we used neck pain rather than stiff neck, and paralysis rather than stiff body). We used the train dataset from the multicentre study to obtain an additional algorithm for diarrhoeal diseases (model 1b) that could be applied to the Morogoro data. In model 1b, we tried using presence of abdominal pain instead of type of abdominal pain, change in amount of urine, abdominal distension and cramps in abdomen. However, abdominal pain was not significantly associated with diarrhoeal deaths in the train dataset so it was dropped from the model. The odds ratios and cut-off for model 1b were obtained using the same methodology as for the other algorithms in Table 1. When applied to the test dataset from the multicentre study, the algorithm for model 1b assigned the same number of deaths as diarrhoeal disease (25, CSMF = 6.7%) as the gold standard, and this resulted in a sensitivity of 56% and specificity of 97%.

Data collection in secondary sites

Navrongo

As part of an ongoing demographic surveillance system, VA interviews have been conducted for all adult deaths since January 1993 in the Kassena-Nankana District of Ghana. The interviews were performed by five trained fieldworkers who were not medically qualified. Information was collected using the VA questionnaire from our multicentre validation study. VA data are available on 5855 deaths. Three local physicians have independently reviewed the VA questionnaires for 406 subjects (7% of the total) who died between January 1994 and February 1998. A cause of death was assigned when at least two of the physicians agreed on the cause of death.

Morogoro

As part of an ongoing demographic surveillance system, VA interviews have been performed for all adult deaths that have occurred since June 1992 in a sample population of more than 120 000 in the Morogoro Rural District of Tanzania (Kitange *et al*. 1996). The interviews were performed by four trained fieldworkers who were medical assistants. Information was collected using the study's own VA questionnaire. VA data are available on 209 adult deaths that occurred in hospital between 1992 and 1995. Two local physicians independently reviewed the VA questionnaires together with the hospital records. In cases of disagreement over cause of death, VA forms were reviewed by a third physician. A cause of death was assigned when two physicians agreed on a cause of death.

Validation in the secondary sites

The data-derived algorithms were applied to the VA data from the secondary sites and causes of death were assigned. Each algorithm was applied independently, without removing subjects that had already been assigned a cause of death, so that it was possible for subjects to have more than one cause of death. The causes of death reached by the algorithms were compared with the physician diagnoses. In particular, we compared the CSMF obtained using the data-derived algorithm with the CSMF obtained using physician review. We also tested the implications of modifying the algorithm cut-offs where the cut-offs given in Table 1 did not perform well.

Results

Table 2 compares the VA data that were used to derive the algorithms in the multicentre study (primary sites) and the data from the two secondary sites. The subjects in Navrongo tended to be older and more likely to be female compared to the other sites. Only a small proportion of subjects died in hospital in Navrongo (11%). Although over 80% of adult deaths in Morogoro occurred at home (Ministry of Health & AMMP Team 1997), analysis was restricted to those that took place in hospital. The median time between death and the VA interview was approximately 3 weeks in Morogoro, 6 months in Navrongo, and 10 months in the multicentre study.

Navrongo

For injuries and meningitis, there was close agreement between the CSMF as estimated by the physicians and the algorithms (Table 3). The physicians assigned only 23 deaths as TB/AIDS, whereas the algorithm assigned 88. The symptoms in the algorithm were, however, strongly associated with TB/AIDS in the Navrongo data. In particular, the three highest scores for the algorithm were assigned to 16 of the 23 physician-diagnosed cases of TB/AIDS (Table 4). There were 54 subjects who had no vomiting or vomiting for > 1 day and passed stools at least 4 times per day and had weight loss, who were assigned as TB/AIDS by the algorithm (score = 3.88) but only one of these was assigned as TB/AIDS by the physicians. By changing the cut-off from 3.59 to 4.29, the algorithm matched the CSMF reached by physician diagnoses. The respondent's report alone that the subject had TB and/or HIV/AIDS agreed closely with the physician diagnoses

Table 2 Comparison of the VA data in the primary and secondary sites

	Primary sites	Secondary sites	
	Multi-centre	Navrongo	Morogoro
Deaths (*n*)	796	406	209
Age (mean)	41	52	38
Female (%)	45	53	43
Died in hospital (%)	100	11	100
Days between death & VA (median)	288	170	21

Table 3 Validity of data-derived algorithms in secondary sites

Cause of death	Cut-off	O (CSMF%)		E^L(CSMF%)		se^L	sp^L
Navrongo (*n* = 406)							
Injuries		26	(6.4)	22	(5.4)	73	99
Meningitis		15	(3.7)	19	(4.7)	93	99
TB/AIDS		23	(5.7)	88	(21.7)	91	83
TB/AIDS	≥ 4.29	23	(5.7)	23	(5.7)	78	99
Diarrhoeal diseases (model 1a)		68	(16.7)	20	(4.9)	25	99
Diarrhoeal diseases (model 1a)	≥ 10.23	68	(16.7)	57	(14.0)	63	96
Morogoro (*n* = 209)							
Injuries		17	(8.1)		–	–	–
Meningitis*		3	(1.4)	2	(1.0)	67	100
TB/AIDS*		77	(36.8)	70	(33.5)	79	93
Diarrhoeal diseases (model 1b)		12	(5.7)	4	(2.9)	0	98

* algorithm has been modified slightly, compared to Table 1, because some variables were not available. Cut-off is indicated if it differs from that given in Table 1. O, observed number of cause-specific deaths using physician review; E^L, expected number of cause-specific deaths using logistic regression algorithm. Sensitivity and specificity of logistic regression algorithm (se^L and sp^L).

(the algorithm assigned 21 cases whereas the physicians assigned 23). The physicians assigned 68 deaths as diarrhoeal disease whereas the algorithm assigned only 20. The symptoms in the algorithm were strongly associated with diarrhoeal deaths. Moreover, the five highest scores for the algorithm were assigned to 44 of the 68 physician-diagnosed diarrhoea cases. By changing the cut-off from 11.52 to 10.23, as we did in the case of TB/AIDS, the algorithm agreed closely with the physician diagnoses.

Morogoro

According to the physician diagnoses, the CSMF was 8.1% for injuries, 1.4% for meningitis, 36.8% for TB/AIDS, and 5.7% for diarrhoea. There was close agreement between the CSMFs for meningitis, and TB/AIDS as estimated by the physicians and the algorithms (Table 3). The physicians assigned 12 deaths as diarrhoeal disease whereas the algorithm (model 1b) assigned only 4.

Discussion

The proposed data-derived algorithms have been validated using data from two different settings. For injuries and meningitis, the algorithms estimated a similar CSMF as the physicians. For TB/AIDS, the algorithm estimated a similar CSMF as the physicians in Morogoro, and the same algorithm with a different cut-off agreed closely with the physicians in Navrongo. The algorithm for diarrhoeal diseases did not agree closely with the physicians in Morogoro, but the algorithm with a different cut-off agreed closely with the physicians in Navrongo.

There are several factors that need to be considered when assessing the performance of the data-derived algorithms in the secondary sites. First, it is possible that the 'gold standard' of physician diagnosis in the secondary sites is inaccurate. In Navrongo, three physician diagnoses were used as the gold standard, and there was much disagreement in their diagnoses. At least two physicians agreed on the cause of death in 88% of subjects, but all three physicians agreed on the cause of death in only 16% of subjects. Despite a lack of agreement between physicians at the individual level, the physicians tended to estimate a similar CSMF for injuries, meningitis, and diarrhoeal diseases. Compared to the physicians, the data-derived algorithms underestimated the number of deaths due to diarrhoeal diseases, and overestimated the number of deaths due to TB/AIDS. We cannot rule out the possibility that the physicians over-report diarrhoeal diseases and under-report TB/AIDS, and that the data-derived algorithms may be more accurate than the physician diagnoses. In Morogoro, the gold standard was based on physician review of the VA questionnaire and hospital records, and this probably gives a reliable diagnosis. Second, the original algorithms were derived using data from deaths that had occurred in hospital, whereas only a small proportion of subjects in Navrongo (11%) died in hospital. Both the clinical presentation of a disease and CSMFs may vary between subjects who die in hospital and those who die at home. Moreover, a hospital episode may affect what the respondent recalls. For example, some respondents may receive detailed information about the signs and symptoms, any test results, and even the cause of death. Third, an algorithm that accurately estimates the CSMF is useful for determining causes of

Table 4 Sensitivity (sens), specificity (spec), and expected number of deaths due to TB/AIDS (E^L) for different cut-offs of the algorithm score compared to physician review in Navrongo

Score	Physician diagnosis of TB/AIDS No n (spec)	Yes n (sens)	Total n (E^L)
0*	31 (0)	2 (100)	33 (406)
1.14	8 (8)	0 (91)	8 (373)
1.43	35 (10)	0 (91)	135 (365)
2.45	5 (45)	0 (91)	5 (230)
2.57	130 (47)	0 (91)	130 (225)
2.74	6 (81)	0 (91)	6 (95)
3.47	1 (82)	0 (91)	1 (89)
3.59	8 (83)	2 (91)	10 (88)
3.76	1 (85)	0 (83)	1 (78)
3.88	53 (85)	1 (83)	54 (77)
4.29	0 (99)	1 (78)	1 (23)
4.90	2 (99)	0 (74)	2 (22)
5.60	0 (99)	1 (74)	1 (20)
5.72	2 (99)	3 (70)	5 (19)
6.74	1 (100)	12 (57)	13 (14)
7.03	0 (100)	1 (4)	1 (1)
Total	383	23	406

* score = 0 if the subject has none of the symptoms in the algorithm (n = 14) or has missing values for any of the symptoms in the algorithm (n = 17). Sensitivity of a cut-off = (number of physician diagnoses of TB/AIDS ≥ cut-off)/23. For example, sensitivity of a cut-off of 4.29 is (1 + 12 + 3 + 1 + 0 + 1)/23 = 78%. Specificity of a cut-off = (number of physician diagnoses of non-TB/AIDS < cut-off)/383. For example, specificity of a cut-off of 4.29 is (31 + 8 + 135 + 5 + 130 + 6 + 1 + 8 + 1 + 53)/383 = 99%.

death at the population level, but not necessarily at the individual level. We focused on classification at the population level because, for many causes of death in our study, the VA did not have high enough sensitivity and specificity for use at the individual level (Quigley *et al.* 1999). Finally, some of our proposed algorithms could not be properly tested in the site (Morogoro) that used a different VA questionnaire.

There are some important limitations of the data-derived algorithm approach to assigning causes of death. First, we tested the algorithms for only four causes of death. The proposed algorithms for other causes of death had poor diagnostic accuracy even in the test dataset (Quigley *et al.* 1999). It is unlikely that data-derived algorithms are appropriate for assigning some causes of death, particularly cardiovascular disorders that tend to have nonspecific symptoms and do not usually generate highly sensitive, specific or consistent results in VAs, even using physician review. Secondly, the proposed algorithms may need to be modified in different settings since the most appropriate cut-offs may differ between sites. Therefore, a validation study may need to be conducted as part of an initial stage of research or surveillance, in order to obtain data for validating and modifying the algorithms. Clearly, a gold standard cause of death is necessary for such a study. The practical implications of conducting such a study, together with the validity of the algorithms and cut-offs obtained, should be reported. Finally, algorithms do not employ the open-ended sections and comments that may be included in the VA questionnaire, and so exclude valuable information that physicians would be able to incorporate in determining causes of death. Such information is useful in the physician diagnosis of broad causes of death that comprise distinct subgroups. In our study, the number of subjects in each subgroup of maternal deaths and neoplasms was too small to identify an algorithm. Clearly, small numbers are not an obstacle to physician diagnosis.

There are, however, several advantages of the data-derived algorithms. First, the algorithms may be applied quickly, cheaply, and easily. This is in contrast to physician review which is time-consuming, expensive, and subjective. The data-derived approach to assigning causes of death may tie in with other analyses that are performed as part of a demographic surveillance system or cohort study. For example, in Mwanza, Tanzania, data from VA questionnaires were used to classify deaths as AIDS using the WHO clinical case definition (Todd *et al.* 1997). Logistic regression has been used to identify diagnostic algorithms that screen Chinese children for *schistosomiasis japonica* (Zhou *et al.* 1998), diagnose clinical malaria in Gambian children (Olaleye *et al.* 1998), and detect cervical infections in pregnant women in Tanzania (Mayaud *et al*. 1998). It is conceivable, however, that alternative data-derived methods may result in more accurate algorithms. An alternative method for the development of algorithms using neural networks resulted in more accurate algorithms for predicting *Schistosomiasis mansoni* infection in Egyptian children than those obtained using logistic regression (Hammad *et al*. 1996). A potential disadvantage of using logistic regression to identify a separate algorithm for each cause of death is that each subject may be assigned several causes of death. This is in contrast to discriminant analysis, probability density estimation, and decision tree classification methods, which would assign at most one cause of death to each subject (Reeves & Quigley 1997). However, the algorithms arising from probability density estimation methods and neural networks are difficult to interpret, and discriminant analysis is not optimal for categorical data. Logistic regression is more widely used and more commonly available in statistical software than these alternative methods. Using logistic regression will more easily enable algorithms to be refined periodically or adapted to different settings.

In summary, the proposed data-derived algorithms for injuries, meningitis, TB/AIDS and diarrhoeal diseases can estimate a similar CSMF as physician review, though some adjustment of algorithm cut-offs may be necessary. All four algorithms may be used as an alternative to physician review in a demographic surveillance system in Navrongo, and will save an estimated 20–30 physician-minutes per VA. We propose that other sites use a VA questionnaire that is amenable to analysis using data-derived algorithms to estimate the CSMF, though further testing and validation is recommended. In settings where physician review is not feasible or too costly, data-derived algorithms provide an alternative method for estimating the CSMF for some major causes of death.

References

Chandramohan D, Maude GH, Rodrigues LC & Hayes RJ (1998) Verbal autopsies for adult deaths: their development and validation in a multicentre study. *Tropical Medicine and International Health* **3**, 436–446.

Cooper RS, Rotimi C, Kaufman J & Lawoyin T (1998) Mortality data for sub-Saharan Africa. *Lancet* **351**, 1739–1740.

Hammad TA, Abdel-Wahab MF, DeClaris N *et al*. (1996) Comparative evaluation of the use of artificial neural networks for modelling the epidemiology of schistosomiasis mansoni. *Transactions of the Royal Society of Tropical Medicine and Hygiene* **90**, 372–376.

Kitange HM, Machibya H, Black J *et al*. (1996) Outlook for survivors in sub-Saharan Africa: adult mortality in Tanzania. *British Medical Journal* **312**, 216–220.

Mayaud P, Uledi E, Cornelissen J *et al*. (1998) Risk scores to detect cervical infections in urban antenatal clinic attenders in Mwanza, Tanzania. *Sexually Transmitted Infections* **74** (Suppl. 1), S139–S146.

Ministry of Health & AMMP Team (1997) *The Policy Implications of Adult Morbidity and Mortality*. *End of Phase 1 Report*. United Republic of Tanzania, Dar es Salaam.

Olaleye BO, Williams LA, D'Alessandro U *et al*. (1998) Clinical predictors of malaria in Gambian children with fever or a history of fever. *Transactions of the Royal Society of Tropical Medicine and Hygiene* **92**, 300–304.

Quigley MA, Chandramohan D & Rodrigues LC (1999) Diagnostic accuracy of physician review, expert algorithms, and data-derived algorithms in adult verbal autopsies. *International Journal of Epidemiology* **78**, in press.

Reeves BC & Quigley MA (1997) A review of data-derived methods for assigning causes of death from verbal autopsy data. *International Journal of Epidemiology* **76**, 1080–1089.

Todd J, Balira R, Grosskurth H *et al*. (1997) HIV-associated mortality in a rural Tanzanian population. *AIDS* **11**, 801–807.

Zhou H, Ross AGP, Hartel GF *et al*. (1998) Diagnosis of Schistosomiasis japonica in Chinese schoolchildren by administration of questionnaires. *Transactions of the Royal Society of Tropical Medicine and Hygiene* **92**, 245–250.

