WHO Technical Consultation on Verbal Autopsy Tools

TALLOIRES, FRANCE, 2-3 NOVEMBER 2004

Final Report

Review of the literature and currently-used verbal autopsy tools

Department of Measurement and Health Information Systems
Evidence and Information for Policy
World Health Organization, Geneva
April 2005

Acknowledgements

This document was prepared by Nadia Soleman, Daniel Chandramohan and Kenji Shibuya on the basis of a technical meeting. Valuable inputs and assistance were provided by Frank Baiden, Peter Byass, Greet Dieltiens, Rajesh Dikshit, Vendhan Gajalakshmi, Yusuf Hemed, Prabhat Jha, Kathleen Kahn, Henry Kalter, Rajesh Kumar, Osamu Kunii, Rosalind Parkes, Chalapati Rao, Philip Setel, James Whitworth, Gonghuan Yang, Stan Yoder, Rajiv Bahl, Somnath Chatterji, Mie Inoue, Paul Kowal, André L'Hours Doris Ma Fat, Sue Piccolo, Ian Scott, Brian Williams, Lara Wolfson, Hongyi Xu, as well as Fred Binka and Sidu Biai. Particular thanks to Ties Boerma for his guidance and advice and to the Ministry of Health, Labour and Welfare of Japan for the grant in support of this work.

LIST OF ABBREVIATIONS

AMMP	Adult Mortality and Morbidity Project
CSMFs	Case specific mortality fractions
CDC	Centers for Disease Control
COD	Cause(s) of Death
DSS	Demographic Surveillance Site(s)
EMRO	WHO Regional Office for the Eastern Mediterranean
HIS	Health Information Systems
ICD	International Statistical Classification of Diseases and Related Health Problems
LSHTM	London School of Hygiene and Tropical Medicine
MESH	Medical Subjects Head
MSF	Médecins Sans Frontières
PAHO	Pan American Health Organization
VA	Verbal Autopsy
WB	World Bank
WHO	World Health Organization
SRS	Sample Registration System

I. Introduction

In many low-and middle-income countries there are insufficient vital registration systems and many factors render an immediate implementation of functioning vital registration systems impossible. But the need for good quality data, especially on causes of death for public health planning and resource allocation, has led to a renewed interest in verbal autopsy (VA), a method known and used for many decades. Other terms used interchangeably for VA are verbal postmortem, verbal autopsy process and tools, and describe a method in which relatives are asked about signs and symptoms of the disease that led to death. The questionnaire is based on a mortality classification system (= causes of death tabulation list) that systematically categorizes the main causes of death (COD) and serves as an aid for both the development of the questionnaire and deriving (coding) causes of death from the answers. In most cases physicians review the questionnaires and assign the diagnoses.

There are currently 36 demographic surveillance sites (DSS) worldwide regularly using verbal autopsy (25 sites are found in Africa, nine sites in Asia, one in Oceania and one in Central America). Many other organizations and researchers in other parts of the world have used this tool to obtain information on mortality. (*1*) The verbal autopsy tool has not only proven to be useful for deriving COD at demographic surveillance sites and to investigate causes of child and maternal mortality but also for short-term use to assess the burden of disease in refugee camps.(*2-4*)

In the early 1990s, discussions about the validity of estimates from verbal autopsy instruments and comparability of data obtained by verbal autopsy started. Expert Committees have emphasized the necessity of a standardized VA tool and pointed out many issues regarding the use of VA for deriving causes of death. However, many sites still use different VA tools and methodologies to derive COD and many of the questions raised at the start of the VA era still remain unanswered.

In November 2004 WHO organized a two-day technical meeting attended by many representatives of demographic surveillance field sites and researchers in the field of VA with the aim of reaching a consensus on core requirements for a standard VA tool. (For the programme and list of participants, please refer to Annex V.)

This paper gives an overview of the different verbal autopsy tools currently used worldwide, summarizes validating study results and methods to derive COD and reports consensus and major discussion points from the Talloires meeting.

II. Methods

A. Collection of verbal autopsy tools

- Thirty-six Demographic Surveillance Sites, researchers or research groups that have worked with verbal autopsy and/or published on this topic were contacted and asked for their questionnaires, mortality classification systems and the way they derived the diagnosis. Furthermore, the websites of NGOs and international

organizations (WHO, CDC, WB, INDEPTH, AMMP) were searched for VA tools.

- Fifteen international experts in the field of VA were contacted and asked for the tools they/their organizations have used and to point out other field sites using VA.
- With the majority of the material coming from Sub-Saharan Africa and Asia we also requested all the country representatives of the Pan American Health Organization (PAHO) and the Eastern Mediterranean Regional Office (EMRO) of WHO whether the method of VA was applied in their respective countries.

In order to give the most up-to-date account, we excluded VA tools that are currently not used from the review. The 18 questionnaires which are included and causes of death tabulation lists come from the DSS sites:[1] of the 25 sites that provided their tools only 18 were analysed because four sites use the INDEPTH standard questionnaire and three sites in Tanzania use a standard tool designed for the Adult Morbidity and Mortality Project (AMMP).[2]

B. Literature search strategy

- This review includes all VA literature published in peer-reviewed journals after 1992 found in Medline by free-text and MESH searches[3] and literature found by a search of the Popline database.
- Workshop reports and discussion notes from a hand search of verbal autopsy archives of the London School of Hygiene and Tropical Medicine (LSHTM).
- Material from a search of the Internet, including a free text Google search, a search of the websites of NGOs and international organizations that have worked with VA (WHO, CDC, WB, INDEPTH, AMMP, HIS, MSF and UNICEF).

III. Verbal autopsy tools - Overview

A. Questionnaires: Adults

This section gives a brief overview of the questionnaires that are currently in use, compares the contents, and points out similarities and differences. Of the 18 evaluated VA questionnaires, 14 have questions on adults but also separate sections for both neonates and post-neonates; one has a section for children (questions on neonatal and post-neonatal COD are combined on one sheet) and four questionnaires contain questions on all age categories mixed on one form. One tool (the WHO standard verbal autopsy questionnaire for children (*5*)) is designed only for child verbal autopsies and is discussed in section III G.

Two thirds of the 18 reviewed questionnaires share questions on the same major signs and symptoms. Only questions about neck pain, stiff neck, abdominal masses and

[1] All tools are available on request from the author.
[2] The three AMMP sites in Tanzania are also part of the INDEPTH network and will not be mentioned separately in this paper.
[3] Search terms available on request.

operations are found on fewer forms. (Table 1, and Annex II show more detailed analysis of questions and subquestions.)

Symptom	Percentage of questionnaires with question on symptoms
Fever	94%
Rash	78%
Weight loss	94%
Pallor/Jaundice	78%
Oedema/Swelling	94%
Cough	94%
Chest Pain	89%
Diarrhoea	94%
Vomiting	94%
Abdominal Pain	83%
Abdominal Distension	78%
Swallowing	67%
Mass	61%
Consciousness	83%
Fits	89%
Paralysis	89%
Headache	83%
Neck Pain	28%
Stiff Neck	61%
Urine Colour	67%
Urine Amount	83%
Accident	67%
Operation	44%

Table 1: Main signs and symptoms: evaluation of 18 adult questionnaires

The questionnaires do not only show minor differences in content but some of them also have similar sequences and wording of questions. This may be attributable to the fact that most of these sites collaborate under the same umbrella organization (INDEPTH),

which introduced a standardized VA questionnaire for optional use and some sites modified this standardized questionnaire.

The review showed that the questionnaires in French and one questionnaire from South Africa must have used another template as they have questions and sequence patterns in common that distinguish them from most of the other questionnaires.[4]

A few questionnaires are a listing of more or less important symptoms and diseases without any structure or pattern. However, it may be the case that non-field-tested translations of originals used in the respective countries were submitted for review. Even considering this fact these questionnaires are incomplete but measures of accuracy of the VA tools from these areas are not available.

The limitations of our attempted quantification of the questions is that the summary of signs and symptoms asked by the reviewed questionnaires does not necessarily inform about the necessity and importance to ask for a specific symptom and about the relevance of the subquestion. A more evidence-based approach is to develop questions based on a compilation of validity study results for diagnostic criteria/algorithms. However, as both validation studies and algorithms are scarce, especially for adults, we decided to use the approach described above to analyse the questionnaires.

SRS India has a very different approach than most other sites and puts an emphasis on the narrative of the questionnaire, which narrows the diagnostic possibilities to physician review In contrast a few of the other field sites rely mainly on a check-list style closed format questionnaire. While this helps to reduce costs, as physicians are not necessarily needed to derive the cause of death, some important (extra) information from the verbatim is lost.

In a recent paper, Marsh et al. separately validated the verbatim, module (= closed-ended) and the combined verbatim plus module section of their questionnaire.(*9*) Sensitivities of various neonatal diseases were very low for the module option alone while verbatim or the verbatim and module combination yielded much higher and similar results for most disease categories which emphasized the need for an open-ended section if COD from a VA form is derived by physician review.

B. Questionnaire design

Open vs closed-ended section

In order to answer open- and closed-ended questions the respondent uses different types of memory: open questions require the respondent to recall something whereas closed questions require recognitions. (*37*) Psychological experiments (*38*) show that more information will be recognized than recalled which, in contradiction to Marsh's findings,(*9*) emphasizes the importance of sections with closed-ended questions. However, the accuracy of the additional information may be lower as respondents may also recognize symptoms that sound familiar even if they were not part of the disease. Other advantages of checklist-type questionnaires are that they reduce bias as

[4] For example, the exact duration of symptoms with start and end dates has to be reported for each symptom - a feature that is not found in any of the other questionnaires.

interviewers are forced to ask all of the questions even if they have a certain diagnosis in mind during the interview. While the administration of open-ended VA questionnaires requires medical training this is not necessary for prestructured, closed-ended forms.(*6*) The participants debated the advantages and disadvantages of verbatim versus closed-ended section and, based on field experience with the different models, there was a general trend to include both in a core standard tool. One of the arguments against the combination of closed- and open-ended sections was that interviewee compliance may be reduced if the respondent is asked about a symptom that he had already reported in the narrative. It was agreed that this issue could be solved if the two parts of the questionnaire do not duplicate but complement each other (e.g. by the use of filter questions, see below), the interviewers are well trained and the questionnaire has obvious instructions on how and when to skip questions. Others argued asking for signs and symptoms twice in the open and closed part serves as probing and increases accuracy.

Another issue discussed was whether information from hospital reports such as official cause of death and medications received at the hospital, if available, should be included in the VA or if this would be an additional source of error: first because the quality of hospital document varies (see also VI.B) and second because it may also introduce additional bias as there is often limited availability of medical reports in rural areas. On the other hand, the additional circumstantial information may help to reach diagnoses. For instance, one of the explanations for the open-ended question performing better than the symptom/signs checklist is that the verbatim account of the respondent often includes the treatment and circumstantial information which helps to each a diagnosis.

C. Sequence

In most of the reviewed questionnaires the open section precedes the closed-ended part in order to reduce recall bias and the report of false positive symptoms in the verbatim part. However, validation studies comparing the results of the different questionnaire formats would answer the question of the optimal sequence.

At most field sites that we reviewed, physicians evaluate the questionnaires and assign COD. Financial and logistic constraints may render physician coding impossible if the use of VA for sample registration systems is scaled up and data derived or expert algorithms may become cost-effective alternatives.

For these automated systems only the closed-ended sections can be used as word processing systems are not advanced to process the information from the verbatim account.

The question sequence is also an important factor that has to be taken into account in the design of the closed-ended part of the questionnaire: the items should follow a logical sequence - resembling a sequence the majority of respondents are believed to follow.(*37*) The arrangement of questions can also influence the response if the respondent believes the content to be related to the previous question. In order to avoid this bias, filter questions, which are general questions that are followed by more detailed and specific questions if positively answered, are useful. More than half of the evaluated questionnaires use this technique, which also helps to avoid respondent fatigue.

D. Layout

The outline, besides the content, is one of the most important factors in the development of a questionnaire.(*37*) A clear layout and questions that are listed in a logical sequence decrease confusion, increase compliance of interviewers and reduce misclassification by coders. However, the quality of the layout differs between the 18 VA tools: while in some cases the layout facilitates the use of the questionnaires other forms are complex and often make a correct marking of answers difficult. As validation studies from most of these sites are missing there are no data available to quantify this issue. An emphasis on pretesting the questionnaires will help to optimize layout.

E. Pretesting, language and biomedical concepts

Pretesting is "state of the art" in designing questionnaires. Nevertheless, it seems that some of the forms were never field tested as some of the questions were vague and ambiguous (two in one questions) or liable for misunderstanding due to grammatical errors.

One of the members of the WHO meeting pointed out that more attention to local language should be given when adapting a questionnaire. Several translators should be involved in the translation of the VA tool, ideally medically-trained persons with knowledge of the local language. The importance of field-testing, both of language and biomedical concepts, that differed considerably between cultures, was also highlighted at this meeting. The use of questionnaires for self-reported symptoms has placed much of the analytical work on respondents, therefore tools that work with local disease concepts would yield better validities.

F. Standardization

In some cases efforts to synchronize the use of a standard tool (INDEPTH) at the different DSS field sites have been successful. However due to financial and logistical constraints many other sites have not been able to change their tools yet: for example, one of the field sites has used their questionnaire for over ten years and the invention of a new tool would result in extra costs. Other field sites obtain the major part of their funds tied to research in a specific public health area and therefore prioritize the part of the tool that helps to get information about the specific research-related diseases and there is less emphasis on modules for other diseases as this may increase costs and workload.

In order to avoid both respondent and interviewer fatigue and increase compliance of the different field sites it was agreed that the signs and symptoms section of the questionnaire should be brief and preferably consist of two pages.

G. Questionnaires - Children

In 1999 the WHO, in collaboration with the Johns Hopkins School of Public Health, developed and validated a standard VA questionnaire for children, which is currently

being improved and updated. This standard tool is divided into two separate sections - one for neonates and other for post-neonates - and was used as a template for the INDEPTH standard tool. Some field sites, have not, however, adopted this standard yet and the task to systematize the different questionnaires for children for the review has been difficult as some focus on birth-related events and prematurity while others emphasize signs and symptoms of diseases of older children and some forms mix questions for children and adults.

Of the 18 questionnaires, 15 have separate parts for children and of these, 14 have further divisions into children and neonates sections. The analysis of the various instruments has yielded that it would be advantageous to separate sections for neonates and children but to have them on one form so that questions from the neonate period can also be chosen when needed (e.g. for a child that has just completed the neonatal period).

There are also inconsistencies in the age criteria for children between the questionnaires: some questionnaires are used for children up to 12 years while others are used for children between 29 days to five years.

A consensus was reached that a standard VA tool should comprise questionnaires for at least three different age groups.

H. Equity and health

This analysis of the questionnaires focuses mainly on the disease-related question pattern of the VA tools, which is important to derive the direct causes of death. A large burden of disease and mortality can, however, be attributed to poverty and inequity. Furthermore, poor communities are more vulnerable to die from certain specific causes such as malaria and TB.

Reliable data on the relationship of equity and health are missing in many low- and middle-income countries. Thus inclusion of socioeconomic characteristics and detailed account of access to, and use of, health services in the VA questionnaire would help to elicit more information in this area. Some of the questionnaires and the INDEPTH standard tool have already incorporated a separate inequity module.

IV. Mortality classification systems

This section analyses the similarities and differences of the various mortality classification systems that are in use at demographic surveillance sites worldwide.

Many researchers who described the development process of verbal autopsy tools start from a classification of the most common lethal diseases in the geographical area of interest.(*6,7*) Such causes of death tabulation lists are the core of any verbal autopsy tool: they aid in structuring the questionnaire and, based on them, the questions that elicit information about the disease of interest can be designed. They help to develop diagnostic algorithms and they also serve as a decision aid for assigning causes of death. However, multiple field sites and researchers stated that they have not developed or

used a mortality classification system and with some exceptions[5] it remains unclear how some researchers or field sites developed questionnaires without an underlying COD list.

A. Framework

According to the theoretical framework on which they are based, the COD tabulation lists can be divided into two main categories: the lists that follow the organ system based on the revision of the Tenth International Classification of Diseases (ICD-10) and the others that group the disease entities by pathophysiological mechanisms.

COD lists that group the diseases by pathophysiological mechanisms and similar risk factors have the advantage that they account for misclassification: e.g. while some classification systems list meningitis and pneumonia under their respective organ systems the pathophysiological mechanism-oriented classification systems categorize them as infectious diseases. This has the advantage that if a certain disease is commonly misclassified on an individual level important information of the burden of disease on group level can still be obtained and aids decision-makers for public health planning and resource allocation.

Four of the 14 reviewed lists use the organ system-based framework to classify the various disease entities (e.g. meningitis is listed with diseases of the central and nervous system), seven other sites grouped them by pathophysiological mechanism (e.g. infectious or noncommunicable diseases are grouped together) and in three cases no framework could be distinguished. One field site assigns codes from the entire ICD-10 and one other field site has a similar approach by not using a mortality classification system for assessing COD in order not to restrict coders to the listed causes of death. After the physicians reach a diagnosis the corresponding ICD-10 codes are assigned. The representative from this field site reported that this approach resembles best the pathway of diagnosing in hospital and offers the physicians maximum freedom to reach a COD but makes a standardization difficult. It was agreed that a similar effect can be obtained by including categories for "other specified diseases" in a standardized mortality classification list.

There was also consensus on a minimum COD list (instead of the ICD-10 or ICD-10 short list) as a starting point ideally following a pathophysiological framework and mapping the diseases to the corresponding ICD-10 codes. The point that this may cause extra efforts as the coding by the pathophysiological system is an additional step in assigning ICD-10 codes was taken into consideration but as the advantages to account for misclassification outweighed the disadvantages most participants preferred the method described above.

B. Content

Three of the used mortality classification systems were designed from different research groups and organizations as templates for standardization. These three lists are more

[5] Huong et al. (*8*) Huong DL, Minh HV, Byass P. Applying verbal autopsy to determine cause of death in rural Vietnam. *Scand J Public Health Suppl* 2003; **2:** 19-25 explain in their paper how he retrospectively developed a COD tabulation list for a VA questionnaire.

complete compared to the other classifications systems and one was used by some of the INDEPTH sites as a template. Tables 1 and 2 of Appendix IV compare and give an overview of the different lists. Some sites use well-refined instruments that allow for misclassification, while other sites use unstructured forms that only list a few diseases or disease categories and do not follow any theoretical framework.

Some field sites excluded different causes of death from their list because on one hand some questions may be culturally inappropriate (e.g. to ask for abortion) and on the other hand the exclusion may reflect the low burden of disease or death from this disease. In contrast, causes of death that match the research interest of a site may be listed in more detail.[6]

Another reason for exclusion of certain diseases may be the believed inability of the VA tool to elicit information on a certain cause of death: for example, one of the COD tabulation lists has only very few disease categories for the class of neoplasms and, for example, cancer of the small and large intestine, rectum and anus are combined as neoplasm of the gastrointestinal tract.[7] These differences in content emphasize the need for validation studies so that, in future, questions will be excluded based on evidence rather than subjective criteria.

C. Layout and user-friendliness

Many factors influence inter- and intra-coder reliability which is essential for the success of verbal autopsy tools. Similarly to questionnaire classification systems (if they are used as coding schemes) which are not clearly laid out and/or user-friendly they can increase the likelihood of misclassification. The lack of visual structure of some lists and very elaborate coding schemes which require the use of a manual or instructions may increase misclassification by coders.

In the many cases where mortality classification systems are used as coding schemes to derive COD they also need to be pretested.

V. Deriving causes of death

A. Physician review

A paper from Quigley et al. (*10*) gives an overview of the different methods to derive causes of death (figure 1). The dominating methodology internationally as well as at reviewed field sites for deriving the causes of death is physician review, is where doctors assign the cause by evaluating the information from the questionnaires based on their clinical experience. However this method is often criticized for its subjectivity and low reliability. The published levels of inter-observer reliability are generally high but may merely reflect the expectation of the individual reviewers, who are aware of the epidemiological pattern and characteristics of diseases in their area. (*11*) Alternatives

[6] Anonymous: personal communication.
[7] INDEPTH preliminary working mortality classification system.

are needed as physician review is relatively cost-ineffective and not feasible if large numbers of questionnaires have to be assessed.

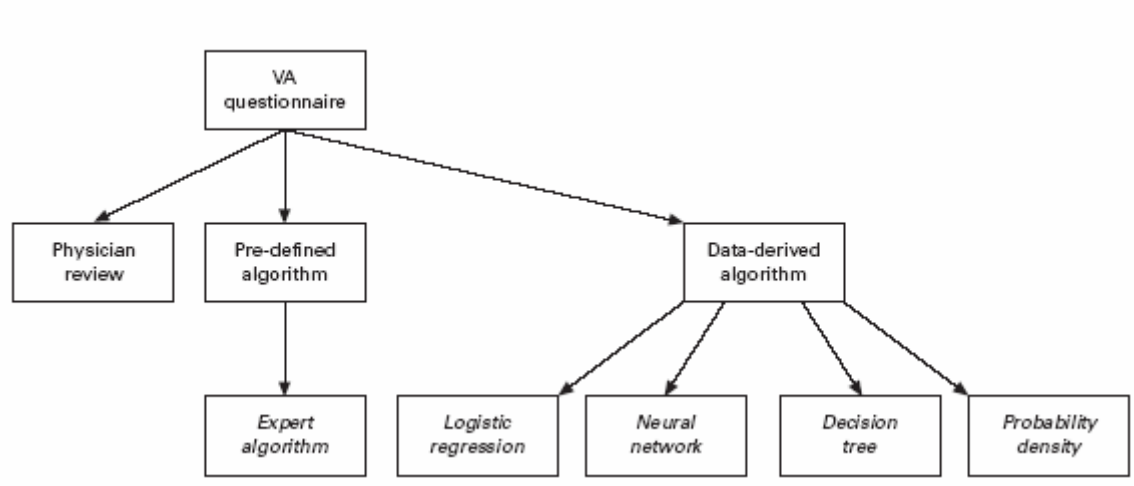

Figure 1. Methods for assigning causes of death using VA data by Quigley et al. (*10*)

B. Data-derived algorithms - Methods

Data-derived algorithms are potential alternatives to physician review but they are not yet ready for use on a larger scale.

The application of the various analytical techniques to derive COD based on linear and other discriminatory techniques (logistic regression), probability density estimation and decision tree and rule-based methods (including artificial neural networks) depends on the intended use. (*12*)

In brief, four factors influence the choice of the classification method (=analytical technique):

- The purpose of the COD analysis
  Is the outcome needed to
  1) compare patterns of mortality to inform health policy (multiple causes of death are compared);
  2) to compare patterns over time or between geographical areas or intervention groups (single or multiple COD are compared); or
  3) formulate interventions to lower mortality.

Here the key distinctive feature is the number of causes of deaths of interest. If mortality for a single cause has to be derived the method of choice is logistic regression which also has the advantage that the software required is easily available and the method is

familiar to epidemiologists. If information on more than one cause of death is needed the other factors that have to be taken into account are:

- The number of COD per case
- The characteristics of the VA dataset

Some of the characteristics that have an influence are the kind of data used for the analysis (categorical, ordinal, continuous) or if the dataset is comprised of many undetermined or rare causes of death.

- The need for a particular classification rule of an analytical technique to be comprehensible

Medically trained users tend to interpret VA classification rules of analytical techniques within the framework of their existing knowledge and to distrust unintelligible ones. Therefore analytical techniques with comprehensible, interpretable classification rules will be easier to implement successfully.
But classification rules derived from different classification methods do not equally fulfil this criterion: for example probability density estimation rules are most incomprehensible because the importance of the different signs and symptoms and the way in which the information is combined are not made explicit. The decision finding seems to happen in a kind of black box for the user.
In contrast are decision tree classification rules better comprehensible because they follow a comparatively simple dichotomous rule.

The question which of the above discussed data-derived methods shows the best performance still remains unanswered. In the published literature there are no analyses of datasets that are large enough for meaningful comparisons yet.(*12*)

C. Data-derived algorithms - Issues

The development and validation of data-derived methods for verbal autopsies is currently at a level that leaves many questions open. Although there are a few validation studies for these methods it remains unclear how valid and accurate these comparisons to the gold standard really are. For most studies the same dataset was used to generate and validate the algorithms and one study that used logistic regression only yielded a similar CSMF as the physician review after adjustment of algorithm cut-off points. (*10, 13,14)* If data-derived algorithms are to be applied on a large scale, further research and validation against a different dataset is needed.

Validation studies have demonstrated that regression-based data-derived algorithms only estimate CSMFthat are comparable to physician review on the population level as they do not have sensitivities and specificities high enough for use at the individual level.(*10, 13*) Nevertheless, experts in the field of verbal autopsy[8] argue that the value of data-derived methods may be underestimated because for data-derived validation studies the closed-ended sections alone were evaluated while most often information of the closed- and open-ended part was used for the review by physicians. In addition,

[8] London School of Hygiene and Tropical Medicine, personal communication.

physicians more often have information on drugs and can take reported discharge diagnoses into account. This may lead to higher values of sensitivities and specificities for the manual review by clinicians.

The argument in favour of data-derived methods is their relative cost-effectiveness in contrast to clinician review and the advantage that they will yield a higher reliability because the criteria for any given COD is fixed a priori. They will, therefore, be consistent over time and between sites. Nevertheless, valuable information from open-ended parts of the question forms will be lost unless this method is developed further to take the verbatim into account for deriving diagnoses (see annex I).

D. Expert algorithms

Expert algorithms which are predefined algorithms developed by a panel of physicians are repeatable and thereby overcome the inconsistencies of clinician review and are time effective. Nevertheless, there are critical voices doubting the validity of pre-defined algorithms as they often include symptoms that do not help to discriminate between two diseases (e.g. fever is both a cardinal symptom for malaria and pneumonia). (*15,16,17*)

Using the Bayesian approach to define the probability of a given cause in the presence of a particular symptom/sign based on personal experiences of experts and compiling a set of probabilities on a semiquantative scale using the Delphi method seems promising to a diagnosis from the VA. (*18*)

Byass et al. (*18*) tested this method on 189 VA interviews from Vietnam. They compared the outcome of the diagnosis reached by these probabilistic models against the diagnosis reached by physicians from VA questionnaires. Over 70% of the COD corresponded and the results were even better (over 83%) if the categories "unknown" or "old age" from the human coders were excluded: in 60% the most probable causes matched directly between the two methods and in another 23.8% the physicians' opinion was among the three most likely causes derived by the Bayesian models. Although this Bayesian model-based approach seems to be promising, it needs to be validated in large datasets using diagnoses from medical records as the "gold standard". The validating study datasets (both children and adults) that are currently available could be used for such validating studies.

E. The practice

Some of the methods described as an alternative to the physician review may provide cost-effective alternatives in the future. So far, field experience with these tools is limited and from the 18 field sites that have supplied information on their methodologies to derive COD 16 use physicians' coding, one site employs a social scientist[9], another one a medical assistant with three years of training to evaluate the questionnaires, the COD are then reviewed by two physicians.[10] Only one site[11] uses

[9] Rakain DSS employs a social scientist for deriving COD.
[10] This double checking at the MATLAB/Indonesia DSS site will not be continued in the future due to limited financial resources.
[11] ACIDS DISS INDEPTH field site South Africa.

expert algorithms but it is not compulsory for physicians to use them for coding. One field site in Vietnam has piloted the probabilistic method, described above, to derive causes of death.

VI. Validation studies

A. What is a validation study?

The term "validity" means the degree to which an assessment instrument measures what it is supposed to measure. However, some studies with the heading "validation study" in peer-reviewed journals do not assess sensitivity or specificity of a VA tool compared to a gold standard. For example, one validation study measures "inter-observer reliability"(*19*).

B. Validation studies - issues

We reviewed eight "classic" adult validation studies and 13 for children published in peer-reviewed journals. Some more sensitivity and specificity values for certain symptoms and diseases are cited in the literature based on personal communication.(*5*) Furthermore, one WHO report lists validations for a few COD in Lebanon.(*20*)

As reported in the literature one has to face limitations in performing validation studies. (*7,9,15,17,21-27*) Due to the scarcity of medical information outside hospital settings in many developing countries only one option currently exists to validate VA study results: the comparison to hospital data. This is not the optimal source of information because of :

- **Information bias**
  Relatives of a patient who died in hospital may be better informed about the cause of death and may therefore be able to better answer the VA questions. On the other hand, separation from the deceased during hospitalization may impair knowledge of signs and symptoms.(*28*)

  It is especially difficult to obtain accurate information on diseases or causes of death that are stigmatized like AIDS: participants at the Talloires meeting reported that relatives are often reluctant to report AIDS-related deaths which leads to an underestimation of the case-specific mortality fraction (CSMF) and reduction of the specificity for other febrile illnesses like malaria. It is especially difficult to obtain correct estimates for maternal mortality because abortion, violence and suicide during pregnancy are frequently underreported.

- **Sample selection**
  Hospital populations may differ from the mostly rural study populations in many aspects. In low income countries patients that can afford the direct and indirect costs of hospital stays often tend to be richer and/or better educated which may influence the ability to answer VA questions. The socioeconomic status has also a major impact on health and higher socioeconomic strata in many countries have different risk factors and may, therefore, be less affected by some and more

affected by other diseases. These epidemiological differences between hospital and general population may distort the result of validation studies.

Other factors are that CSMF may be lower in hospital as patients that would have died in the community are saved or cases that are severely ill or injured are not brought to hospital when the relatives assume that they will die soon - which reduces the CSMF. On the other hand patients may only decide to stay in hospital when they are already heavily affected by a disease. This scenario may be reflected in an upwards biased CSMF.

- **Cut-off points**
  There are no internationally agreed cut-off points above which sensitivities and specificities deem to be acceptable.(*7*) Some authors have used cut-off >90% for sensitivity and specificity as acceptable. Another option is to use a combination of sensitivity and agreement between the VA estimates of CSMF and "true" CSMF. For example, a sensitivity of <50% and the difference between the VA estimate of CSMF and the true CSMF is within 20% points of the true CSMF.

- **Quality of used reference standards**
  Many authors list the unreliability of the used gold standard as one of the major limitations of their study.(*17,28*) Information from various sources is used as gold standard for validation studies: most researchers rely on the diagnosis of the underlying or main cause of death from the clinical records, others review the case history and use, where available, laboratory results to come to a diagnosis. (*5,7,9,17,22,24,28-30*)

  However, the reliability of these reference diagnoses remains often unclear - especially in settings where there are often very limited diagnostic possibilities. In addition, sometimes the hospital records do not contain explicit diagnoses of stigmatized diseases. One validation study describes the report of proxy terms like immunocompromised, Cryptococcus meningitis and retroviral disease in medical records in order to avoid the term AIDS in the official records.(*32*) During the discussions of gold standards for VA validation studies many of the participants argued that this term should be replaced by the term "reference standard" because of the differing quality of information from clinical records.

Only one study uses the result from necropsies as a reference standard, which is believed to give the most accurate information on the underlying COD.(*25*) Nevertheless this method is only rarely applicable in many settings in developing countries.(*15*) Another option described in publications is to nest the validation study in a prospective clinical study that defines certain necessary criteria (clinical examination and, often, certain laboratory results) or algorithms for the diagnosis of a disease in the clinic. In case of a death the COD stated on the clinical record is reproducible by these defined standards and used as a reference diagnosis.(*9,24*)

In conclusion, the generalization of findings of hospital-based validation studies to the general population is limited. Many researchers highlighted the need for community-based validation studies at the Talloires meeting.

The comparability of validation study findings across settings is also restricted due to the use of different methods.(*24*) One question that is still unanswered is how the

duration of recall affects sensitivity and specificity of the VA tool. The minimal and maximal time range between death and interview varies between validation studies but only a few tested the influence of the recall period on validity and reliability. Mirza et al.(*34*) claim in their paper that there is no significant difference in reliability between longer and shorter recall intervals. However, the sample sizes for her hypothesis testing are small. Her findings are in accordance with another study that has found little association of recall period on sensitivity and specificity and assumes that the memory of events leading to a death is retained for long periods, possibly up to three years.(*6*)

Similarly, there is only limited information of the influence of the interviewer's education on verbal autopsy results. While some claim that they should be medically trained others prefer interviewers with basic school education as medical knowledge may bias the result towards a certain differential diagnosis the interrogator has in mind.(*8*) The effect of many other variables on validity and reliability of the VA tool remains unclear.

One of the largest evaluation studies that has examined the effect of age, sex, relationship and language of the respondents has not found a significant effect of these variables on the sensitivity and specificity of VA.(*6*) However, the accuracy of the verbal autopsy tool has been better if the respondents looked after the deceased during the final illness. The correspondence between the VA and gold standard CSMF has also been better if the respondents spoke the same language as that used for the questionnaire. Questionnaires should be translated into the locally-used language whenever possible, but in some areas this may not be feasible due to the variety of languages and local dialects that are spoken.

Another issue raised by participants of the meeting was whether questions on information on the discharge sheet should be included in the questionnaire or whether this is an additional source of error. Some of the field sites have already included such questions in their tools. However, contextual information is, in many cases, better and more often available from urban areas where people have better access to hospital care and may bias the results.

There was general agreement at the meeting that internationally standardized procedures for verbal autopsies and validation studies are needed and the representatives of the different field sites presented and discussed their experiences with factors like time from death to interview and type of interviewer. Many useful suggestions were made on these factors and there was an agreement that validation of these factors will help in creating evidence-based VA tools and processes.

C. Validation studies - children

In contrast to the validation studies of adults most child VA validation studies do not only validate the COD but also various algorithms to derive the cause of death. Although all the methodological inconsistencies explained above between the studies make a comparison difficult, a general trend can be observed in validation studies for death in children.

Algorithms for some of the main childhood diseases with very distinct features like measles or neonatal tetanus (NNT) perform well and yield high sensitivities and

specificities while the values for diseases with overlapping symptoms like diarrhoea, malaria and acute respiratory infection (ARI) are in the lower range.(*15,17,24,26*) Tables 1-3 displays values for diseases where VA performs well with high sensitivities and specificities or of diseases that are of major public health importance even if the values are in the lower range.

There are, especially for neonatal deaths - with the exception of NNT - very few algorithms that yield high specificities or sensitivities for VA. Two of the studies do also include severely ill children and one of the authors claims that the results of such studies are better generalizable as they reflect more accurately the actual mix of severity and causes of diseases in community settings.(*24*)

D. Validation studies - adults

There are only a few (eight) reports of causes of adults deaths and even less studies that validate algorithms for adults.(*7,23*) The most comprehensive validation studies were published in 1998 (*7,23*) and since then only the results of one more (*28*) big validation study were reported. These studies demonstrate as well that diseases with unique features have higher sensitivities and specificities when validated.

Chandramohan et al. have found the highest validities for the group of direct maternal causes (82%, 98%), injuries, tetanus (77%, 100%) and rabies (83%, 100%).(*7*) Kahn et al.(*28*) not only report high values for these disease entities but in contrast also high ones for some infectious diseases, especially diarrhoea (sensitivity=100, specificity=100) and pulmonary TB (sensitivity=92%, specificity=99%). This high sensitivity to detect certain infectious disease deaths may be due to the absence of malaria and typhoid in the geographical area of the second study as diseases with rather unspecific signs and symptoms may increase misclassification.

But to obtain a more complete picture the results from more than two studies are needed. In addition, Hosegood (*32*) reports much higher values for the use of VA for AIDS compared to other reviewers - but the sample size of the study is a third of the size of source one from the table and epidemiological patterns of other infectious diseases have to be taken into account as well. There are some more peer-reviewed published papers but they validated a small number, in some cases only one cause of death.(*31-33*) The question has to be asked whether the validation of one or a few causes of death equals the validation of the whole tool as some authors claim.(*34*)

VII. The expanding role of verbal autopsy

In areas where most deaths take place at home verbal autopsy is used as a substitute and to derive mortality patterns of the researched area. In the absence of accurate data from other sources many also envision the improvement of the VA tool for comparing mortality patterns over time and/or between geographical areas.(*12*)

So far, measurement of trends between areas is not possible due to the different VA tools in use. But even if this uncertainty factor is overcome by standardized methods

there are many other reasons why the application of the VA process for monitoring trends is debatable.

Disease/ Condition	Sensitivity (%)	Specificity (%)	Predictive value + (%)	CSMF (VA-tool) (%)	CSMF (gold standard) (%)	Difference (%)	Algorithms Used	Country
Measles[26]	90	96	84	17.5	11.1	58.3	None	Kenya
ALRI[34]	71	92	83	30.8	36	14.4	None	Kenya
Diarrhoea[28]	86	100	100	NA	NA		None	South Africa
Malnutrition[28]	86	100	100	-	-		None	South Africa
Kwashiokor[28]	100	100	100	-	-		None	South Africa
Malaria[21]	71	100	100	-	-			Malawi
Cerebral Malaria[15]	72	85		-	-		Fever, convulsions or loss of consciousness	Namibia
Anaemia[21]	83	93	71	-	-		None	Malawi
Accidents[28]	100	97	80	15	12	25	None	South Africa

NA: not available

Table 1: VA validation studies in children

Disease/ Condition	Sensitivity (%)	Specificity (%)	Predictive value + (%)	CSMF (VA-tool) (%)	CSMF (gold standard) (%)	Difference (%)	Algorithms Used	Country
Neonatal tetanus[30]	100	93	NA	NA	NA		Age <28 d and convulsions or spasms	Bangladesh
Diarrhoea[5]	91	93	-	-	-		Local term for diarrhoea	Nicaragua
Preterm births[5]	79	85	-	-	-		Pregnancy ended early	Nicaragua
AIDS[21]	60	87	43					
LBW/severe malnutrition[5]	82	80					Pregnancy ended early or baby was very small at birth	Bangladesh

NA: not available

Table 2: VA validation studies in neonates

Disease/ Condition	Sensitivity (%)	Specificity (%)	Predictive value + (%)	CSMF (VA-tool) (%)	CSMF (gold standard) (%)	Difference (%)	Algorithms Used	Country
Malaria[7]	39	97	68	8.8	15.7	44	None	Ethiopia
Meningitis[7]	64	98	78	7.3	8.9	18	None	Tanzania
TB[7]**/ Pulmonary TB**[28]	57 / 92	99 / 99	81 / 92	6.7 / NA	9.5 / NA	30 / NA	None	Tanzania / South Africa
AIDS[33]	85.9	92	92	47	50.3	7	None	Uganda
Injuries[7]	100	100	100	6.8	6.8	0	None	Ethiopia
Direct maternal causes[7]	90	99	69	3.2	4.1	29	None	Tanzania
Ante/Post Partum Haemorrhage[7]	100	97	56	7	4	75	None	Combined data from Ethiopia, Ghana, Tanzania

NA: not available

Table 3: VA validation studies - adults

- Misclassification

  Misclassification affects estimates of changes of CSMF over time and between population groups and depends on the mix of prevalence diseases, which varies within and between regions and lowers the accuracy of the used instrument. The performance of the VA tool depends on the coexistence of diseases with similar features: for instance, can the presence or absence of a disease without a very specific clinical presentation like malaria influence the accordance between CSMFs for infectious diseases due to an increase in misclassification.(*28,35*).

  If sensitivity and specificity are known, it is theoretically possible to adjust for misclassification.

- Validity

  Sensitivity and specificity depend among other things on disease characteristics which vary between regions. This is again best exemplified by malaria.

The symptomatology of Malaria differs in high transmission areas where chronic *Plasmodium falciparum* infection is common and it is more often accompanied by anaemia than cerebral malaria which has more specific symptoms and is therefore easier to detect by verbal autopsy.(*36*)

Average estimates of malaria mortality across Africa between 1982 and 1998 have been published based on data from demographic surveillance but have not been adjusted for differences in mortality between areas with different transmission or for other possible changes of disease incidence and prevalence (e.g. AIDS) over time.(*36*) In an attempt to use verbal autopsy for trend estimates Korenromp et al. adjusted specificities using results from published validation studies but there is some inconclusiveness in their argument. For the adjustment of the specificity they assume a linear inverse relationship between specificity and CSMF. However, this theory is based on seven validation studies whose methodology was subject to many limitations as described above. Furthermore, they extrapolate the sensitivity of VA by averaging the reported sensitivities ranging from 45-75% to adjust the VA estimates of CSMF. This is in contradiction to the notion that the sensitivity of physician-review changes depending on the prevalence of a condition. This theory applies especially if the prevalence of a disease is very low because the physician may not diagnose rare diseases unless they have very distinct features.

VIII. Conclusion - Current status of Verbal Autopsy

The review of verbal autopsy methods has shown that the internationally-used verbal autopsy tools are of varying quality. Although there were previous attempts by different organizations to introduce standardized methods the VA instruments differ between and within the organizations that use them. Examples discussed in this review demonstrate

that in analogy to good medical practice VA would benefit from a standard for "good verbal autopsy practice". The participants of the Talloires meeting concluded that one internationally-used core standard VA tool is needed. For this, not only a standardized tool has to be developed but also further research on the factors that influence reliability and validity as agreed in the VA workshop are needed.

Summary

There was agreement on the following points at the WHO meeting on verbal autopsy tools:

I. Questionnaire

- Both the open- and closed-ended section of the VA questionnaire will be useful and necessary components of the core VA questionnaire
- Sections/core questionnaires for three different age groups (neonates, children and adults) are needed and these questionnaires should ideally not exceed more than 2 pages in order to ensure the compliance of interviewers, respondents and the field sites involved.
- Questions on risk factors and service use will help to elicit information for public health planning and resource allocation.

II. Mortality classification system

- The mortality classification system list should ideally list COD by pathophysiological categories and link them to the corresponding ICD-10 codes.
- Categories for "other specified diseases" should be included in the list in order not to restrict coders to COD only mentioned in the mortality classification system.

III. Validation studies

- In order to ensure a valid standard VA tool that yields accurate data, successive improvement of the core tool with information from studies is needed. Detailed information and studies on the following points are needed:
  - "lay" vs. medically-trained professionals as interviewers
  - open- vs. closed-ended section vs. both
  - optimal time span from death to interview
  - data derived algorithms
  - expert systems
- More data from validation studies - especially on adults - are needed. Internationally standardized methods for validation studies with
  - similar reference standards
  - similar time frames until the interviews
  - similar cut-off points to evaluate the usefulness of the VA tool

  will help to get measures of accuracy that are better comparable.

References

1. Gray RH, Smith G. Barss P. The use of verbal autopsy methods to determine selected causes of death in children. (Occasional paper No. 10). Baltimore: Institute for International Programs, Johns Hopkins University, 1990.
2. Verbal Tools for Adults Deaths, Workshop Report: London School of Hygiene and Tropical Medicine, 1993.
3. Verbal Autopsy Tools for Adult Deaths. London School of Hygiene and Tropical Medicine, 1996.
4. Verbal Autopsies for Maternal Deaths. Report of a WHO Workshop. London School of Hygiene and Tropical Medicine, 1994.
5. Anker M, Black RE, Coldham C, et al. A standard verbal autopsy method for investigating causes of death in infants and children: World Health Organization, 1996.
6. Chandramohan D. Verbal autopsies for assessing causes of adult and maternal death: development and validity of a model tool. PhD Thesis: London School of Hygiene and Tropical Medicine, 2002.
7. Chandramohan D, Maude GH, Rodrigues LC, Hayes RJ. Verbal atuopsies for adult deaths: their development and validating in a multicentre study. *Trop Med Int Health* 1998; **3**(6): 436-46.
8. Huong DL, Minh HV, Byass P. Applying verbal autopsy to determine cause of death in rural Vietnam. Scand J Public Health Suppl 2003; 2: 19-25.
9. Marsh DR, Sadruddin S, Fikree FF, Krishnan C, Darmstadt GL. Validating of verbal autopsy to determine the cause of 137 neonatal deaths in Karachi, Pakistan. *Paediatr Perinat Epidemio* 2003; **17**(2): 132-42.
10. Quigley MA, Chandramohan D, Setel P, Binka F, Rodrigues LC. Validity of data-derived algorithms for ascertaining causes of adult death in two African sites using verbal autopsy. *Trop Med Int Health* 2000;**5**(1)**:**33-9.
11. Todd J, Balira R, Grosskurth H, et al. HIV-associated adult mortality in a rural Tanzanian population. *Aids* 1997;**11**(6)**:**801-7.
12. Reeves BC, Quigley M. A review of data-derived methods for assigning causes of death from verbal autopsy data. *Int J Epidemiol* 1997;**26**(5)**:**1080-9.
13. Quigley MA, Chandramohan D, Rodrigues LC. Diagnostic accuracy of physician review, expert algorithms and data-derived algorithms in adult verbal autopsies. *Int J Epidemiol* 1999;**28**(6)**:**1081-7.
14. Boulle A, Chandramohan D, Weller P. A case study of using artificial neural networks for classifying cause of death from verbal autopsy. *Int J Epidemiol* 2001;**30**(3)**:**515-20.
15. Mobley CC, Boerma JT, Titus S, Lohrke B, Shangula K, Black RE. Validation study of a verbal autopsy method for causes of childhood mortality in Namibia. *J Trop Pediatr* 1996;**42**(6)**:**365-9.
16. Quigley MA, Armstrong Schellenberg JR, Snow RW. Algorithms for verbal autopsies: a validation study in Kenyan children. *Bull World Health Organ* 1996;**74**(2)**:**147-54.
17. Kalter HD, Gray RH, Black RE, Gultiano SA. Validation of postmortem interviews to ascertain selected causes of death in children. *Int J Epidemiol* 1990;**19**(2)**:**380-6

18. Byass P, Huong DL, Minh HV. A probabilistic approach to interpreting verbal autopsies: methodology and preliminary validation in Vietnam. *Scand J Public Health Suppl* 2003;**62:**32-7.
19. Benara SK, Singh P. Validity of causes of infant death by verbal autopsy. *Indian J Pediatr* 1999;**66**(5)**:**647-50.
20. Sibai AM. National Burden of Disease Study:Mortality in Lebanon:
A Standard Verbal Autopsy Method for investigating Causes of Death in Adults: Department of Epidemiology and Population Health Faculty of Health Sciences, American University of Beirut, 2002
21. Nykanen M, Tamaona W, Cullinan T, Van Oosterzee V, Ashorn P. Verbal autopsy as a technique to establish causes of infant and child mortality. *East Afr Med J* 1995;**72**(11)**:**731-4.
22. Chandramohan D, Maude GH, Rodrigues LC, Hayes RJ. Verbal autopsies for adult deaths: issues in their development and validation. *Int J Epidemiol* 1994;**23**(2)**:**213-22.
23. Chandramohan D, Rodrigues LC, Maude GH, Hayes RJ. The validity of verbal autopsies for assessing the causes of institutional maternal death. *Stud Fam Plann* 1998;**29**(4)**:**414-22.
24. Kalter HD, Hossain M, Burnham G, et al. Validation of caregiver interviews to diagnose common causes of severe neonatal illness. *Paediatr Perinat Epidemiol* 1999;**13**(1)**:**99-113.
25. Rodriguez L, Reyes H, Tome P, Ridaura C, Flores S, Guiscafre H. Validation of the verbal autopsy method to ascertain acute respiratory infection as cause of death. *IndianJ Pediatr* 1998;**65**(4)**:**579-84.
26. Snow RW, Armstrong JR, Forster D, et al. Childhood deaths in Africa: uses and limitations of verbal autopsies. *Lancet* 1992;**340**(8815)**:**351-5.
27. Todd JE, De Francisco A, O'Dempsey TJ, Greenwood BM. The limitations of verbal autopsy in a malaria-endemic region. *Ann Trop Paediatr* 1994;**14**(1)**:**31-6.
28. Kahn K, Tollman SM, Garenne M, Gear JS. Validation and application of verbal autopsies in a rural area of South Africa. *Trop Med Int Health* 2000;**5**(11)**:**824-31.
29. Coldham C, Ross D, Quigley M, Segura Z, Chandramohan D. Prospective validation of a standardized questionnaire for estimating childhood mortality and morbidity due to pneumonia and diarrhoea. *Trop Med Int Health* 2000;**5**(2)**:**134-44.
30. Kalter H. The validation of interviews for estimating morbidity. *Health Policy Plan* 1992;**7**(1)**:**30-9.
31. Kahn K, Tollman SM, Garenne M, Gear JS. Who dies from what? Determining cause of death in South Africa's rural north-east. *Trop Med Int Health* 1999;**4**(6)**:**433-41.
32. Hosegood V, Vanneste AM, Timaeus IM. Levels and causes of adult mortality in rural South Africa: the impact of AIDS. *Aids* 2004;**18**(4)**:**663-71.
33. Kamali A, Wagner HU, Nakiyingi J, Sabiiti I, Kengeya-Kayondo JF, Mulder DW. Verbal autopsy as a tool for diagnosing HIV-related adult deaths in rural Uganda. *Int J Epidemiol* 1996;**25**(3)**:**679-84.
34. Mirza NM, Macharia WM, Wafula EM, Agwanda RO, Onyango FE. Verbal autopsy: a tool for determining cause of death in a community. *East Afr Med J* 1990;**67**(10)**:**693-8.

35. Chandramohan D, Setel P, Quigley M. Effect of misclassification of causes of death in verbal autopsy: can it be adjusted? *Int J Epidemiol* 2001;**30**(3)**:**509-14.
36. Korenromp EL, Williams BG, Gouws E, Dye C, Snow RW. Measurement of trends in childhood malaria mortality in Africa: an assessment of progress toward targets based on verbal autopsy. *Lancet Infect Dis* 2003;**3**(6)**:**349-58.
37. Bennet AE, Ritchie K. Questionnaires in Medicine. A guide to their design and use. Oxford University Press, 1975.
38. Maccoby, Belson in Bennet AE, Ritchie K. Questionnaires in Medicine. A guide to their design and use. Oxford University press, 1975.

Operational issues	Physician review	Predefined expert algorithms	Logistic regression model-based data derived algorithms	Artificial neural network	Case-based reasoning system	Expert opinion-based Bayesian probability models
▪ Accuracy of cause-specific mortality estimates ▪ Repeatability of cause-specific mortality estimates ▪ Human resource requirements ▪ Logistics requirements ▪ Sustainability ▪ Time lag between VA interview and VA diagnosis ▪ Transparency of the method	▪ Validation study to date suggests that this method gives robust estimates of CSM for several common causes ▪ A study of child VA in the Gambia showed poor repeatability of diagnosis reached by a panel of physicians ▪ Labour intensive. Needs +/- 3 physician-years for coding 3000 VAs annually ▪ Computer hardware plus software for data management and statistical analysis ▪ Difficult to sustain the commitment of physicians over a long period ▪ Typically the time lag is 3-5 months ▪ Rules used by physicians are not transparent	▪ Sensitivity and specificity of this method is lesser than physician review; give robust estimates of CSM, but need further research to be applicable widely ▪ Very consistent ▪ 75% of physician time can be saved. Needs 2 data input clerks entering 3000 VAs annually ▪ Computer hardware plus software for data management and statistical analysis ▪ Data management would need close monitoring and routine maintenance of the system ▪ Typically 1-2 months ▪ Transparent and clinically credible	▪ Sensitivity and specificity of this method is lesser than physician review; give robust estimates of CSM, but need further research to be applicable widely ▪ Can be consistent if the cut off scores of the algorithms are fixed ▪ 75% of physician-time can be saved. Needs 2 data input clerks entering 3000 VAs annually ▪ Computer hardware plus software for data management and statistical analysis ▪ Data management would need close monitoring and routine maintenance of the system ▪ Typically 1-2 months ▪ Transparent, but some criteria may lack clinical credibility	▪ Accuracy is comparable to physician review; needs further work to adapt the system for AMMP ▪ Can be very consistent if a standard system is trained ▪ 80% of physician-time can be saved. Needs 2 data input clerks entering 3000 VAs annually ▪ Need to develop customized neural network systems ▪ Data management would need close monitoring and routine maintenance of the system ▪ Typically 1-2 months ▪ Not transparent; clinical credibility may not be known	▪ A new system trained to diagnose causes of death based on series of typical cases; yet to be developed. ▪ Can be very consistent if a standard system is developed ▪ 80% of physician-time can be saved. Needs 2 data input clerks entering 3000 VAs annually ▪ Need to develop customized case-based reasoning system ▪ Data management would need close monitoring and routine maintenance of the system ▪ Typically 1-2 months ▪ Not transparent; clinical credibility may not be known	▪ Shown to produce CSMF estimates consistent with physician review. Not yet validated on "gold standard" estimates ▪ Consistent ▪ 80% of physician-time can be saved. Needs 2 data input clerks entering 3000 VAs annually ▪ Need to develop customized Bayesian model-based system ▪ Data management would need close monitoring and routine maintenance of the system ▪ Typically 1-2 months ▪ Transparent, but some criteria may lack clinical credibility

ANNEX I. Strengths and limitations of selected options for coding VA

(from Chandramohan D. Second stage redesign of the AMMP verbal autopsy tool)

ANNEX II

SIGN/SYMPTOM	NUMBER OF FIELD SITES ASKING FOR SIGN/SYMPTOM
FEVER	17
• Severity	12
• Pattern	13
• Chills/Rigor	9
• Backpain/Myalgia	1
• Convulsions (fever)	1
• Duration	16
RASH	13
• Location	11
• Skin Peeling	7
• Description	12
• Red Eyes	7
• Ulcer	3
• Scaring	2
• Itchy	8
• Pins and needles	3
• Bleedings from openings	3
• Duration	15
WEIGHT LOSS	17
• Severity	9
• Appearance	2
• Duration	8
PALLOR, JAUNDICE	14
• Discoloration of eyes	12
• Duration	7
OEDEMA, SWELLING	17
• Around ankle	9
• Ulcer	7
• Puffy face	8
• Neck	7
• Armpit	9
• Groin	9
• Joints	4
• Abdomen	5
• Duration	13
COUGH	17
• Dry	2
• Productive	16
• Blood	15
• Night Sweat	7
• Time of Day	7
• Shortness of breath	14
• Noisy breathing	2
• TB-status	1
• Duration	15

Table 1: Number of field sites that ask for a specific sign or symptom and subquestions: adult questionnaires

ANNEX II

SIGN/SYMPTOM	NUMBER OF FIELD SITES ASKING FOR SIGN/SYMPTOM
CHEST PAIN	16
• Sudden/Gradually	7
• Location	11
• Resting	4
• Activity	5
• Continuous/On-off	5
• Palpitation	7
• Cough with sputum/dyspnoea	2
• Duration	15
DIARRHOEA	17
• Continuous	5
• Consistency	12
• Frequency	13
• Blood	15
• Sunken Eyes	8
• Vomiting	7
• Duration	16
VOMITING	17
• Frequency	6
• Description	17
• Continuous/On-off	3
• Duration	17
ABDOMINAL PAIN	15
• Location	12
• Severity	8
• Constipation	11
• Type	11
• Duration	12
ABDOMINAL DISTENSION	14
• Developed rapidly or slowly	11
• Duration	14
SWALLOWING	12
• Duration	9
MASS	11
• Location	8
• Duration	10
CONSCIOUSNESS	15
• Level (confused, unconscious, other)	8
• Sudden, rapid (1day), slow start	11
• Duration	10
FITS	16
• Frequency	11
• Awake/Unconscious between fits	9
• Difficulty opening mouth during fits	8
• Stiffness of whole body during fits	10
• Duration of stiffness	5
• Duration of unconsciousness	15

Table 1: Number of field sites that ask for a specific sign or symptom and subquestions: adult questionnaires

(continued)

ANNEX II

SIGN/SYMPTOM	NUMBER OF FIELD SITES ASKING FOR SIGN/SYMPTOM
PARALYSIS	16
• How long did it take to develop	5
• Lower limbs	7
• Duration	12
HEADACHE	15
• Continuous/On-Off	1
• Severity	4
• Duration	7
STIFF NECK	9
• Neck pain	5
• Duration	11
URINE COLOUR	12
• Colour change	16
• Colour	11
• Duration	9
URINE AMOUNT CHANGES	15
• Amount	10
• Difficulty or pain	15
• Type of difficulty	9
• Duration	14
ACCIDENT	12
• Type of accident	8
• Death at site of accident	5
• Received medical care	5
• Duration of survival after the accident	7
OPERATION	8
• Site	8
• When	8

Table 1: Number of field sites that ask for a specific sign or symptom and subquestions: adult questionnaires

(continued)

ANNEX II

SIGN/SYMPTOM	NUMBER OF FIELD SITES ASKING FOR SIGN/SYMPTOM
PREGNANCY	16
• Time of pregnancy	11
• Excessive bleeding before/after delivery	15
• Duration of labour	13
• Mode of delivery	14
• Difficulty delivering placenta	12
• Fever	7
• Convulsions	8
• Vaginal bleeding	6
• Baby alive	12
• How is baby	3
• Location of delivery	12
• Management	6
ABORTION	13
• Abortion-death, time relation	2
• Bleeding after abortion	3
• Fever after abortion	2
• Induced abortion	3
OTHERS	
• Ulcer/breast	10
• Irregular bleeding	9
• Menorrhagy	4
• Discharge	5

Table 2: Number of field sites that ask for a specific sign or symptom and subquestions: adult questionnaires - evaluation of obstetrics and gynaecological sections

ANNEX III

SIGN/SYMPTOM	NUMBER OF FIELD SITES ASKING FOR SIGN/SYMPTOM
ACCIDENT/INJURY	9
NORMAL GROWTH	9
WEIGHT LOSS	15
MALFORMATION	1
REPEATED EPISODES OF ILLNESS	7
DIARRHOEA	17
FREQUENCY OF DIARRHOEA	11
BLOOD IN STOOL	15
SUNKEN EYES	12
COUGH	17
BREATHING DIFFICULT	15
BREATHING FAST	15
CONVULSIONS/FITS	17
CHEST INDRAWING	17
WHEEZING	13
NO GRASP	8
NOT RESPONDING TO VOICE	8
NOT FOLLOWING MOVEMENTS WITH EYES	7
BULGING FONTANEL	13
SKIN RASH	15
LOCATION OF SKIN RASH	12
DESCRIPTION OF SKIN RASH	13
CRACKING/PEELING SKIN AFTER RASH	10
MEASLES	10
VERY THIN	5
WASTING/MARASMUS	7
SWELLING OF BODY PARTS	16
SWELLING OF WHOLE BODY	8
PALE	12
PALE PALMS	11
WHITE NAILS	8
JAUNDICE	7
FEVER	16
VOMITING	11
TRUSH	9
HAIR COLOUR CHANGE (RED)	**7**

Table 1: Signs and symptoms - children

ANNEX III

SIGN/SYMPTOM	NUMBER OF FIELD SITES ASKING FOR SIGN/SYMPTOM
ACCIDENT/INJURY	14
MALFORMATION	15
COMPLICATION OF LATE PREGNANCY	7
DURATION OF PREGNANCY	4
TETANUS VACCINATION OF MOTHER	3
SIZE AT BIRTH OR BIRTH WEIGHT	14
BRUISES/ SIGNS OF INJURY	7
DIFFICULTY BREATHING	10
DIFFICULTY BREATHING SOON AFTER BIRTH	7
DIFFICULTY BREATHING	7
DIFFICULTY SUCKING	4
DIFFICULTY SUCKING	10
DIFFICULTY SUCKING AFTER A PERIOD OF	8
CRIED AFTER BIRTH	17
SPASMS AN CONVULSIONS	14
BULGING FONTANEL	14
SKIN COLOUR (PURPLE/PALE)	2
YELLOW EYES/SKIN	16
RED UMBILICAL CORD STUMP	15
RED/ HOT OR PEELING SKIN	10
RASH WITH PUS FILLED BLISTERS	12
BLEEDING FROM ANY SITE	8
DIARRHOEA	14
STOOL FREQUENCY	8
BLOOD IN STOOL	7
FAST BREATHING	13
CHEST INDRAWING	14
FEVER	13
COUGH	15
VOMITING	12

Table 2: Signs and Symptoms - Neonates

ANNEX IV

FIELD SITE	STRUCTURE	DIVIDED IN MAIN CAUSES & SUBGROUPS	SECTION ON ADULTS & CHILDREN	USER FRIENDLINESS	CLEAR LAY-OUT	NOT INCLUDED IN LIST	MISCLASSIFICATION WITHIN PATHOPHYSIOLOGICAL GROUP POSSIBLE
ACDID DSS	N/A	N/A	N/A	N/A	N/A	N/A	N/A
AMMP	ICD-10 oriented	No	No	Yes	No	Rabies, Epilepsy	No
LSHTM list	By pathophysiological mechanism	Yes	Yes	Yes	Yes	Rabies, Epilepsy, NNT	Yes
INDEPTH (temporary)	By pathophysiological mechanism	Yes	Yes	Yes	Yes	NNT	Yes
IFKARA DSS	Use INDEPTH one	Yes	Yes	Yes	Yes		Yes
FILABAVI DSS	Unclear	No	No	No		N/A, very short list	No
KARONGA DSS	By pathophysiological mechanism	Yes	Yes	No	No	Accidental poisoning	Yes
KISUMU DSS	ICD-10 based	Yes	Yes	Yes	Yes	N/A, very short list	No
NAIROBI DSS	By pathophysiological mechanism	Yes	Yes	Yes	Yes	Hypertension, Abortion	Yes
OUBRITENGA DSS	ICD-10 based	Yes	Yes	Yes	No		No
RUFIJI DSS	N/A	N/A	N/A	No	No		No
PURWORJEO DSS	By pathophysiological mechanism	Yes	No	Yes	Yes	N/A, very short list	Yes
NAVRONGO DSS	Use INDEPTH one but have separate one for children	Yes	Yes	Yes	yes		Yes
SRS INDIA	ICD-10 based	N/A	N/A	N/A	N/A	N/A	No

Table 1: Comparison of mortality classification systems I

ANNEX IV

Mortality classification system	Number of field site using as a template	list 1 compared to list 2: disease categories missing on list 1	list 2 compared to list 1: disease categories missing on list 2	list 3 compared to list 1: disease categories missing on list 3
List 1 (LSHTM list before modification)		Rabies, Epilepsy, NNT, Congestive Heart Failure		
List 2 (INDEPTH working COD list)	**4** (Karonga, Navrongo, Purworejo, APHRC)		Leishmaniasis, Appendicitis, NNT, Intestinal Obstruction, Ca. of Lip, Oral Cavity, Pharynx GI. Ca less detailed, Obstructive Disorders of Urinary tract, Mental and Behavioural Disorders, all other Endocrine Disorders, Unintentional Injury list not detailed	
List 3 (AMMP proposed list)				Ca. of Ovaries, Appendicitis, Intestinal Obstruction, Obstructive Disorders of Urinary Tract. (does also not include Rabies, Epilepsy)

Table 2: Comparison of mortality classification systems II

ANNEX IV

Site	Template used	Missing categories if compared to INDEPTH list	Additional categories if compared to INDEPTH list	Comment
Karonga	INDEPTH	Epilepsy, Accidental poisoning,		
Navrongo	INDEPTH	Puerperal sepsis, hypertension, Ca of lung, Epilepsy	Very detailed COD list for children,	
Purworjeo	INDEPTH	Missing from Communicable diseases: Meningitis, Aids, Rabies, Non Communicable Diseases are mainly headings, Abortion, Obstructed labour		
Kisumu	None	N/A	N/A	Is a coding sheet for hospital diagnoses and therefore contains diseases that are hard to diagnose by VA: e.g.BPH
Filabavi	None	N/A	N/A	short, unstructured list
APHRC, Nairobi	INDEPTH	Abortion, Puerperal Sepsis, Hypertension	Acute and chronic complications of Diabetes, Typhoid	
OUBRITENGA	ICD 10	N/A	N/A	
RUFIJI	Various (ICD and INDEPTH list)	N/A	N/A	No underlying framework: mix of ICD 10 and INDEPTH list
ACDIS DSS	N/A is a compilation of algorithms			

Table 3: Comparison of mortality classification systems III

ANNEX V. LIST OF PARTICIPANTS

WORLD HEALTH ORGANIZATION

VERBAL AUTOPSY MEETING 2-3 NOVEMBER 2004

TALLOIRES, FRANCE, 02 - 03 November 2004

LIST OF PARTICIPANTS

Temporary Advisers

Dr Frank Ekow BAIDEN
Clinical Research Officer
Navrongo Health Research Center
PO Box 114
Navrongo UER
GHANA

Telephone No.: 0023324 4591181
Email address: frankbaiden@hotmail.com
baidenf@yahoo.co.uk

Dr Peter BYASS
Guest Professor
Umea Int. School of Public Health
Epidemiology and Public Hlth Sciences
Dept. of Public Health & Clinical Med.
Umea University
S-90185 Umea
SWEDEN

Telephone No.: 0046907853345
Email address: peter.byass@epiph.umu.se

Dr Peter BYASS
Honorary Professor, IMMPACT
Dugald Baird Centre for Research on
Women's Health
Department of Obstetrics and Gynaecology
University of Aberdeen
Aberdeen Maternity Hospital
Cornhill Road
Aberdeen AB25 2ZL, UK

Telephone No.: 00441949 836016
Email address: p.byass@abdn.ac.uk

Dr Daniel CHANDRAMOHAN
Disease Control and Vector Biology Unit
London School of Hygiene and Tropical Medicine
Keppel Street
London WC1E 7HT
UK

Telephone No.: 0044 207 927 2322
Fax No.: 0044 207 580 9075
Email address: Daniel.Chandramohan@lshtm.ac.uk

Dr Greet DIELTIENS
Institute for Tropical Medicine
Nationalestraat 155
2000 Antwerpen
BELGIUM

Telephone No.: 0032 32476 304
Fax No.: 0032 32476 258
Email address: gdieltiens@itg.be

Dr Rajesh DIKSHIT
Visiting Scientist
International Agency for Research on Cancer
150 cours Albert Thomas
Lyon 69008
FRANCE

Telephone No. : 0033472738030
Fax: 0033 472 738 320
Email address: dikshit@iarc.fr

Dr Vendhan GAJALAKSHMI
Director
Epidemiological Research Center
New No. 37, Outer Circular Road
Kilpauk Garden Colony, Cnennai 600 010
Tamil Nadu
INDIA

Telephone No.: 0091 98 401 60050
Email address: gajaerc@rediffmail.com
Fax No.: 0091 44 225 78621

Dr Yusuf HEMED
Director
AMMP
PO Box 65243
Dar es Salaam
UNITED REP. TANZANIA

Telephone No. : + 255 22 215 3388
Fax No. : +255 22 215 3385
Email address: hemed@ammp.or.tz
maharage2000@yahoo.com

Dr Prabhat JHA
Canada Research Chair in Hlth & Dev.
Public Health Sciences
University of Toronto
70 Richmond Street East
3rd Floor, Toronto, Ontario M5C 1N8
CANADA

Telephone No.: 1 416 864 6042
Fax No.: 1 416 864 5256
Email address: prabhat.jha@utoronto.ca

Dr Kathleen KAHN
Agincourt Health & Population Unit
School of Public Health
University of Witwatersrand
7 York Road
Parktown 2193
Johannesburg
SOUTH AFRICA

Telephone No.: 27 11 717 2617
Fax No. : 27 11 717 2084
Email address : KAHNK@SPH.WITS.AC.ZA

Dr Henry KALTER
Department of International Health
Johns Hopkins Bloomberg School of
Public Health
615 N. Wolfe Street, Rm E8132
Baltimore, MD 21205
USA

Telephone No. : +9729955 0184
Email address : hkalter@jhsph.edu

Dr Rajesh KUMAR
Professor of Community Medicine
School of Public Health
Post Graduate Inst. of Medical Education
and Research
Chandigarh
160 012
INDIA

Telephone No. : 0091 172 3139948
Fax No. : 0091 172 274 4401
Email address : rajeshkum@sancharnet.in

Dr Osamu KUNII
Professor of Global Health
Research Center for Tropical Infectious
Diseases
Nagasaki University Institute of
Tropical Medicine
Sakamoto 1-12-4
Nagasaki 852-8523
JAPAN

Telephone No. : +81 95 849 7869
Fax No. : +81 95 849 7853
Email address : kunii@jp.org

Dr Rosalind PARKES
Project Leader
Cryptococcal Study
MRC Programme on AIDS
c/o UVRI
P.O. Box 49, Entebbe
UGANDA

Telephone No. : 256 782 3253
Email address : rosalind.parkes@mrcuganda.org

Dr Chalapati RAO
Lecturer
University of Queensland
Herston Road, Herston 4006
Brisbane
AUSTRALIA

Telephone No. : 61 7 3346 4623
Fax No. : 61 7 3365 5442
Email address : C.Rao@sph.uq.edu.au

Dr Philip SETEL
Deputy Director
Measure Evaluation
Carolina Population Center
CB 8120
CB 8120 University Square East
Chapel Hill, NC 27516-3997
USA

Telephone No. : +1 919 966 7541
Fax No. : +1 919 966 2391
Email address : psetel@unc.edu

Dr Nadia SOLEMAN
London School of Hygiene and Tropical Medicine
Keppel Street
London WC1E 7HT
UK

Email address : NADIA.SOLEMAN@lshtm.ac.uk

Dr James WHITWORTH
Head of International Activities
Science Funding
The Wellcome Trust
215 Euston Road
London, NW1 2BE
UK

Telephone No. : +44 207 6118854
Fax: +44 207 611 7288
Email address: j.whitworth@wellcome.ac.uk

Dr Gonghuan YANG
Director
Chinese Academy of Preventive Medicine
Division of DSP
27 NanWei Road
Beijing 100050
CHINA

Telephone No. : 8610 6317 1866
Fax No. : 8610 6317 0894
Email address : yangghuan@sina.com

Dr Stan YODER
Demographic and Health Surveys
11785 Beltsville Drive
Calverton, MD 20705
USA

Telephone No. : +1 301-572-0840
Email address : Paul.S.Yoder@orcmacro.com

World Health Organization

Headquarters

Dr Rajiv BAHL
Medical Officer
Telephone No. : +4122 7913766
Email address: bahlr@who.int

Dr Paul R. KOWAL
Scientist
Telephone No. : +4122 7914379
Email address: kowalp@who.int

Ms Agnes PRUDHOMME
Statistician
Telephone No. : +4122 7913612
Email address : prudhommea@who.int

Dr Ian William SCOTT
Technical Officer
Telephone No. : +4122 791325
Email address: scotti@who.int

Mr Brian WILLIAMS
Epidemiologist
Telephone No. :+4122 7914680
Email address: williamsbg@who.int

Dr Lara WOLFSON
Technical Officer
Telephone No. : +4122 7911857
Email address: wolfsonl@who.int

Ms Hongyi XU
Statistician
Telephone No. : +4122 7913658
Email address: xuh@who.int

Dr Jan Ties BOERMA
Director
Telephone No. : +4122 7911481
Email address: boermat@who.int

Dr Somnath CHATTERJI
Scientist
Telephone No. : +4122 7913202
Email address: chatterjis@who.int

Ms Mie INOUE
Statistician
Telephone No. : +4122 7912309
Email address: inouem@who.int

Mr André C. L'HOURS
Technical Officer
Telephone No. : +4122 7912843
Email address: lhoursa@who.int

Mrs Doris MA FAT
Statistician
Telephone No. :+4122 7912841
Email address: mafatd@who.int

Mrs Susan PICCOLO
Telephone No. :+4122 7912855

Secretary

Email address: piccolos@who.int

Dr Kenji SHIBUYA
Scientist

Telephone No. : +4122 7912370
Email address: shibuyak@who.int