Assessing a new approach to verbal autopsy interpretation in a rural Ethiopian community: the InterVA model

Mesganaw Fantahun, a,b Edward Fottrell, b Yemane Berhane, a,b Stig Wall, b Ulf Högberg, c,d & Peter Byass d

Objective Verbal autopsy (VA) — the interviewing of family members or caregivers about the circumstances of a death after the event — is an established tool in areas where routine death registration is non-existent or inadequate. We assessed the performance of a probabilistic model (InterVA) for interpreting community-based VA interviews, in order to investigate patterns of cause-specific mortality in a rural Ethiopian community. We compared results with those obtained after review of the VA by local physicians, with a view to validating the model as a community-based tool.

Methods Two-hundred and eighty-nine VA interviews were successfully completed; these included most deaths occurring in a defined community over a 1-year period. The VA interviews were interpreted by physicians and by the model, and cause-specific mortality fractions were derived for the whole community and for particular age groups using both approaches.

Findings The results of the two approaches to interpretation correlated well in this example from Ethiopia. Four major cause groups accounted for over 60% of all mortality, and patterns within specific age groups were consistent with expectations for an underdeveloped high-mortality community in sub-Saharan Africa.

Conclusion Compared with interpretation by physicians, the InterVA model is much less labour intensive and offers 100% consistency. It is a valuable new tool for characterizing patterns of cause-specific mortality in communities without death registration and for comparing patterns of mortality in different populations.

Keywords Autopsy; Interviews; Questionnaires; Cause of death; Probability; Mortality/trends; Ethiopia (source: MeSH, NLM).
Mots clés Autopsie; Entretien; Questionnaires; Cause décès ; Probabilité ; Mortalité/orientations; Éthiopie (source: MeSH, INSERM).
Palabras clave Autopsia; Entrevistas; Cuestionarios; Causa de muerte; Probabilidad; Mortalidad/tendencias; Étiopía (fuente: DeCS, BIREME).

Introduction Verbal autopsy (VA) — the interviewing of family members or caregivers about the circumstances of a death after the event — is a useful tool in areas where routine death registration is non-existent or inadequate. Processing VA interview material to arrive at a cause of death (COD) has often used physician review, although this demands a considerable amount of (often scarce) physician time, and frequently more than one physician is used to interpret each VA to increase the possibility of obtaining an objective consensus. Other methods of interpretation of the VA, such as algorithms and neural networks, have been explored inconclusively in terms of validity. A new probabilistic approach has recently been proposed that has given promising preliminary results. This approach applies Bayes’ theorem to VA interpretation in an attempt to overcome concerns associated with expert assessment while also addressing the limitations of algorithmic approaches. This model has already been evaluated on a preliminary basis in Vietnam and is now known as InterVA.

Data on deaths in the community have been collected at the Butajira Rural Health Program (BRHP) Demographic Surveillance Site (DSS) since 1987. The routine death registration form records perceived causes of death, but this is of limited use. A better understanding of cause-specific mortality in such communities is a prerequisite for reducing unacceptably high rates of mortality. Reliable VA methods can be used to assess trends in cause-specific mortality, mortality differentials between population groups and the effects of interventions. For example, a study of childhood mortality in a Senegalese DSS provided useful information on the mortality structure responsible for a major peak in mortality. Efficient and valid tools for the assessment of cause-specific mortality are also necessary for monitoring progress towards achieving the Millennium Development Goals.

The aim of this study was to assess the performance of the InterVA model...
on a series of community-based VA interviews from rural Ethiopia, comparing cause-specific mortality as determined by the model and by local physicians, and identifying areas for further development of the model.

Methods

The BRHP was established in 1987 by selecting a sample of ten communities (with probabilities proportional to size) in the Butajira District, some 130 km south-west of Addis Ababa, Ethiopia. Butajira is set in the Ethiopian Rift Valley, with altitudes ranging from 1750 m to 3400 m above sea level, and local communities depending largely on subsistence agriculture and some cash-crop cultivation. Although predominantly rural, the district includes the market town of Butajira, which accounts for about 10% of the total population, as reflected in the sample.

At the time of this study, which was conducted between August 2003 and July 2004, the sample constituted about 15% of the population of the district, and comprised about 45,000 people. In this prospective study, about 350 deaths were expected, 325 were actually recorded and VA interviews were successfully conducted for 89% of these (289 cases).

The verbal autopsy interview and questionnaire

The VA questionnaire used was prepared by WHO and INDEPTH (International Network of field sites with continuous Demographic Evaluation of Populations and Their Health in developing countries) and was adapted to the local situation. The questionnaire included open narrative and closed questions. The narrative was used to record free explanations of the circumstances of death while the closed questions dealt with specific symptoms and conditions. Five interviewers administered the questionnaire to those who had witnessed the deaths and/or took care of the deceased. The interviewers had all completed secondary (high school) education and had previous experience in community-based data collection. Training of the interviewers emphasized issues such as preferred respondents, period of interviews, approaching grieving respondents and compiling narrative material (ensuring that duration, frequency, severity and sequence of the symptoms were mentioned). Pretesting resulted in small modifications to facilitate understanding of the study population. Each interviewer was assigned two communities on the basis of his/her previous place of work and experience. Details of deaths were notified by DSS staff, who also regularly reviewed death registration forms and BRHP registers to identify deaths for which VA interviews could be carried out. A supervisor coordinated their activities, oversaw the data collection process, checked questionnaires for completeness and consistency, and conducted random quality checks by re-interviewing about 5% of the respondents.

Data were collected between 45 and 60 days after death in most cases. Forty-five days is considered to be the usual period of mourning in the study area. Two months was chosen as an upper guideline to minimize recall bias concerning details of symptoms and circumstances of death. In about 15% of the cases, data collection took place between 4 and 5 months after death, for operational reasons.

Two physicians reviewed each VA form independently to assign one or more causes of death, and subsequently met to reach consensus for cases where there were differences of opinion. If no physician consensus could be reached after discussion, the COD was regarded as indeterminate. The physicians were trained in procedures to assign causes of death and given details of the study area, population and the surveillance system. They were briefed on the common local terms used to express signs, symptoms, causes and conditions of death. They were also provided with a list of ICD-9 (International classification of diseases, ninth revision, http://www.cdc.gov/nchs/icd9.htm) causes of death and required as much information as possible to identify main and underlying causes of death. The physicians were not specifically briefed about the probabilistic model, in order to preserve their independence. The causes of death assigned by the physicians were coded and entered into a database.

Data analysis

The probabilistic model, as previously described, requires the extraction of a defined set of indicators (signs, symptoms, history, circumstances) as the input to the model, and which can be derived from both the open narrative and closed questions in the VA interview. Running the model on these indicators then generates a database with up to three likely causes of death for each case, together with respective likelihoods. Table 1 shows an example of how likely causes of death for an individual case are determined stepwise by the model as successive indicators are processed. In this case, the final causes of death were human immunodeficiency virus/acquired immune deficiency syndrome (HIV/AIDS) with pulmonary tuberculosis.

Since the objective of this study was to determine the cause-specific mortality fraction (CSMF) for each important COD in the community, individual cases that were assigned more than one cause (by physicians or by the model) were taken to represent two or three fractional causes. For the physicians, for cases with two causes of death each was assigned a weight of 0.5, while for the model the respective likelihoods were used for apportionment where there were two or three causes. Thus the two approaches processed the same basic data from the VA questionnaire independently, resulting in individually assigned causes of death by both methods, which were then aggregated to respective CSMFs at the community level.

Ethical approval for the study was obtained from the Faculty of Medicine, Addis Ababa University and the Ethiopian National Ethics Committee.

Results

A total of 289 VA interviews were successfully completed. Infants (aged less than 1 year) accounted for 48 cases (16.6%), children aged 1–14 years for 101 cases (34.9%) and adults 140 cases (48.4%). Of the deaths in childhood, 87% occurred before the fifth birthday. Deaths in rural areas amounted to 235 cases (81.3%). For the 289 VA interviews, the physicians reached consensus without further discussion in 201 cases (69.3%). In four cases, no consensus was reached after discussion, resulting in an indeterminate outcome. In 57 cases (19.7%), two causes of death were assigned. In the model’s output, 84 cases (29.1%) were assigned three causes and 73 (25.3%) had two causes. Examination of the range of causes assigned by the physicians and by the model, respectively, revealed some important differences in terminology. The physicians differentiated between “pneumonia” and “sepsis”, while the model used a single category
for these conditions. The physicians also tended to use more etiological descriptions in some cases, for example “shigellosis” in preference to “bloody diarrhoea.” Fig. 1 represents a comparison of principal CSMFs. The physicians’ stated causes are listed in the left-hand column and the model’s descriptors are listed in the right-hand column. The two central columns represent a rationalized version of the physicians’ and model’s CSMFs respectively, for the sake of easier comparison across common categories.

There was a general similarity between the two approaches in terms of cause-specific mortality, but with some important differences. For all deaths in this community, four major groups of causes accounted for more than 60% of all mortality, as determined both by the physicians and by the model. These were pneumonia/sepsis, pulmonary tuberculosis, malaria, and diarrhoeal disease/malnutrition (Fig. 1). Among infants in the first year of life, prematurity/low birth weight, perinatal causes and pneumonia/sepsis accounted for four out of five deaths, as assessed both by the physicians’ review and the model. For children, pneumonia/sepsis, diarrhoeal disease/malnutrition and malaria were the major causes of death. Even among adults, infectious causes (pneumonia/sepsis, tuberculosis, malaria and diarrhoeal disease) accounted for more than 50% of deaths according to both methods.

There were also good rank correlations between causes of death assigned by the physicians and by the model, both for the whole community (Fig. 1) and for specific age groups, particularly when including the “others” category. This category included different diseases assigned with a low frequency (e.g. other infectious diseases, other respiratory problems, kidney or urinary problems, malignant, other digestive problems) and a group of cases for which the COD was indeterminate (four cases by physicians’ diagnosis and five by the probabilistic method).

Some important differences included the lower frequency with which meningitis and malaria were assigned as a COD by the model than by the physicians. Perinatal problems appeared to be responsible for a higher proportion of deaths among infants according to the physicians. On the other hand, HIV/AIDS was less frequently diagnosed by the physicians than by the model.

Fig. 2 shows principal CSMFs according to the model for the urban area (Butajira town) and the surrounding rural areas. In the town, the model detected higher rates of HIV/AIDS, tuberculosis and cardiovascular mortality than in the rural area, all of which are credible findings.

Discussion

Comparison of the physicians’ review and the probabilistic model in determining causes of death showed that both approaches yielded very similar findings for the major CSMFs in this community, despite that the fact that the two approaches were applied independently to the same data, and that the model was built without direct reference to the data for Butajira. Patterns of mortality revealed were consistent with those anticipated for an underdeveloped population in sub-Saharan Africa. This approach would be equally applicable to cause-specific mortality rates, given appropriate denominator data. However, there are some differences in interpretation that deserve further attention. There is a need for intelligent interpretation of categorizations that may be largely equivalent without being exactly the same, for example, the preference of the physicians for “shigellosis”, imputing a likely specific etiology to reports of bloody diarrhoea, but in the absence of any microbiological data.

Diagnosis of the disease group comprising malnutrition, kwashiorkor, marasmus and diarrhoea was an important and more complex issue for these data. This version of the InterVA model was not designed to differentiate between kwashiorkor and marasmus, and presented a single possible category of “malnutrition”. However, most people dying as a result of malnutrition would also have a strong history of diarrhoea reported in the VA interview; marasms would be reported as losing weight, and people with kwashiorkor may be reported as having weight loss or no weight loss, oedema, and skin and hair changes. This whole complex area is something that needs further attention in the probabilistic model, although the results for the composite group of malnutrition, marasmus, kwashiorkor and diarrhoea obtained by the physicians and by the model were still comparable.

Meningitis in young infants may not be accompanied by the traditional signs of the disease (neck stiffness, fever, severe headache and other meningeal signs), or the signs may not be noted by

---

Table 1. Example of verbal autopsy interpretation showing selected causes of death from the InterVA model

<table>
<thead>
<tr>
<th>Indicator</th>
<th>Probability of selected causes of death</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Acute cardiac</td>
</tr>
<tr>
<td>Unconditional probability</td>
<td></td>
</tr>
<tr>
<td>1st: Male aged 15–49 years</td>
<td>0.005</td>
</tr>
<tr>
<td>2nd: Final illness lasted more than 3 weeks</td>
<td>0.000</td>
</tr>
<tr>
<td>3rd: Oral candidiasis</td>
<td>0.000</td>
</tr>
<tr>
<td>4th: Coughing up blood</td>
<td>0.000</td>
</tr>
<tr>
<td>5th: Cough for more than 3 weeks</td>
<td>0.000</td>
</tr>
</tbody>
</table>
the caregiver, which makes it difficult to diagnose the disease using specific indicators. In principle, these signs might be expected in older children, but the physicians (rightly or wrongly) perhaps used more general symptoms of acute infection to lead towards a diagnosis of meningitis.

Malaria in the Ethiopian highlands follows a somewhat atypical unstable hypopoendemic pattern,\(^\text{10}\) which may not be optimally reflected in what is intended to be a generally applicable model. This kind of endemicity tends to result in more deaths than might otherwise be expected in relatively immune-naive adults, when compared with mortality attributable to the more common childhood malaria seen in areas of higher endemicity. Choosing between the model’s options of “low” and “high” malaria prevalence did not substantially narrow the gap between physicians’ diagnosis and the probabilistic method, although at the “high” setting the model performed better.

HIV/AIDS was more frequently identified by the model than by the physicians’ review. In children, the cases assigned by the model as HIV/AIDS were identified by the physicians’ review mainly as kwashiorkor, while in adults they were generally diagnosed as tuberculosis, or tuberculosis and diarrhoea. International estimates\(^\text{11}\) and studies in poor rural African settings\(^\text{12}\) have reported that malnutrition is the major underlying COD in children. Thus it seems that the model may have overestimated the incidence of HIV/AIDS in children, while the physicians’ review may have underestimated it. This may also reflect some reluctance among health professionals to reach a diagnosis of HIV/AIDS on the basis of circumstantial evidence.

Among infant deaths there appeared to be a difference in the diagnosis of perinatal mortality made by the two approaches. Further inspection of the data revealed that the model reached a diagnosis of pneumonia in about five cases that were labelled as perinatal deaths by the physicians’ review. The physicians’ decisions may have been reached by directly applying the definition of the perinatal period (the first 7 days of life) as the causal basis for these early deaths. The model’s age grouping, does not define such a narrow interval at the beginning of life. However, in six cases for which the model considered prematurity/low birth weight to be the second likely COD, the physicians restricted themselves to perinatal causes. If prematurity/low birth weight is considered to be a perinatal problem, in accordance with the ICD-9 classification, the gap between the two approaches reduces. Thus the apparent differences in diagnoses for the very young may largely arise as a result of differences in definition. In addition, the differences in CSMFs may be attributable to the relatively small number (48) of deaths in this age group.

Another area for consideration was the model’s indicator “rash”, which,
when present, tended to lead to “measles” as a COD. This led to five conclusions of “measles” from the model, compared with 0.5 from the physicians. However, it seemed from the VA material that measles rash was well recognized by respondents as a separate entity, justifying its inclusion as a separate indicator in the next revision of the model. The model also incorporated a possibility of “sickle-cell disease”, not normally found in Ethiopia, and attributed one case erroneously to this cause, which should be generalized to include other haemoglobinopathies.

Both approaches painted a picture of a high-mortality community in which most deaths were caused by major infectious diseases. While this might be expected in rural Ethiopia, it nevertheless highlights the importance of targeting many preventable infections with improved health-care measures if this burden of communicable disease is to be reduced.13

Mortality patterns in the modest but rapidly developing market town of Butajira show some trends towards higher proportions of deaths related to HIV/AIDS and noncommunicable diseases than in its rural surroundings, as well as a higher CSMF for tuberculosis, possibly associated with more crowded living conditions. These observations may have important implications for future cause-specific trends in mortality in rural communities, as behavioural and lifestyle changes trickle out from urban areas. Age-specific differences in CSMFs followed largely predictable patterns, with infections dominating all age groups.

As validity is a major concern in any VA application,14–16 possible ways of assessing both approaches in relation to a gold standard were considered. Validation using hospital COD as a gold standard was not possible owing to the extreme rarity of hospital deaths in this setting; most of the deaths occurred in the home without any professional contact. Thus, reluctantly, we had to conclude that any formal validation was impossible. However, the good concordance in CSMFs between the two methods and the reasonableness of findings compared with other local studies13,17–19 suggest that validity for both methods is good. In addition, the clear differences between the results produced here for Ethiopia and those previously produced using data from Viet Nam3 demonstrate the model’s capacity to reflect mortality patterns in very different communities, as do the evident urban–rural differences shown in Fig. 2.

Philosophies underlying the application of VA in assessing community-based mortality have not always been clear. The term “autopsy” implies some equivalence with a pathologist’s postmortem, in which the major objective is an unambiguous conclusion for a specific individual. However, in terms of the epidemiological and public health implications of cause-specific mortality, the
emphasize is not so much on the individual's COD but on eliciting the causes that have a major impact on communities as a whole. The approach taken here reflects this whole-community emphasis, by allowing attributions of cause to be apportioned for individuals with multiple likely causes. While the physicians had the advantage of being able to consider detailed, nuanced information by going through the questionnaire and using their clinical skills and experiences in determining cause(s) of death, they may have also been influenced by their own biases, particularly for less obvious causes of death for which decisions had to be made between equally likely diagnoses. This might have contributed to some of the discordance observed between the two approaches. It seems fair to conclude that the InterVA model, subject to some limitations identified above, represents a valuable new tool in the quest to characterize patterns of cause-specific mortality in communities without death registration. While its use may be more limited for identifying COD at the individual level and its performance in terms of determining CSMFs may not be obviously better than that of the local physicians in this setting, it offers two major advantages as community-level tool for identifying COD patterns. It is much less labour-intensive, since at an estimate of 5 minutes per physician per VA interpretation, even a relatively small study such as this demands more than one week of physician time purely for VA interpretation. At the same time, precisely the same model can be applied to VA material from a range of settings, or over extended periods of time, without introducing any interobserver variation in interpretation. These two advantages outweigh the possible losses associated with subtle interpretation carried out by physicians, at least in a public health context. Possibilities for future comparisons of patterns of mortality at different surveillance sites, for example, among various INDEPTH members, are thus created, paving the way for building a more reliable global overview of patterns of mortality.

Current developments
This exercise in Butajira provided some important indicators for modifications to be made at a detailed level, which would be needed in the next version of the model, and which have now been implemented. This revised version of the InterVA model is now available for general release (www.interva.net). The development of more specialized versions of the model (e.g. specifically for maternal deaths) is in progress. A version of the model that can be used on a handheld computer (PDA) will also be available for download from this website; this system could be used during the VA interview and thus eliminate the need to transfer the VA data into the model after the interview. We hope that making these resources generally available will stimulate progress towards consistent comparisons of patterns of mortality.

Acknowledgements
We acknowledge Dr Derege Ketema and Dr Amha Fantaye, who assigned causes of death.

Funding: This study received financial support from the Swedish Agency for Research Co-operation with Developing Countries (SAREC).

Competing interests: none declared.
esos resultados con los obtenidos por médicos locales a los que se dio a examinar las AV, con miras a validar el modelo como instrumento comunitario.

**Métodos** Se completaron con éxito doscientas ochenta y nueve entrevistas de AV, que incluían la mayor parte de las defunciones ocurridas en una determinada comunidad durante un periodo de un año. Las entrevistas de AV fueron interpretadas por médicos y mediante el modelo, y se procedió a calcular las fracciones de mortalidad por causas específicas para el conjunto de la comunidad y para grupos de edad particulares utilizando los dos sistemas.

**Resultados** Se observó una buena correlación entre los resultados de los dos sistemas de interpretación en este ejemplo de Etiopía.

Cuatro grandes grupos de causas explicaban más del 60% de toda la mortalidad, y los perfiles internos observados en grupos de edad específicos fueron compatibles con lo que podía esperarse para una comunidad subdesarrollada de alta mortalidad del África subsahariana.

**Conclusión** En comparación con la interpretación realizada por médicos, el modelo InterVA requiere mucho menos trabajo y tiene una coherencia del 100%. Es un nuevo y valioso instrumento para caracterizar la distribución de la mortalidad por causas específicas en comunidades sin registros de defunción y para comparar las pautas de mortalidad de distintas poblaciones.

Melhus

Cet article présente les résultats de l’étude d’autopsie verbale réalisée en Éthiopie. Le modèle InterVA a été utilisé pour interpréter les AV. Les résultats obtenus ont été comparés avec ceux obtenus par les médecins locaux. Un bon accord a été observé entre les deux systèmes d’interprétation.

**Méthodes** Des entretiens d’autopsie verbale (AV) ont été réalisés dans une communauté d’Éthiopie sur une période d’un an. Les AV ont été interprétées par des médecins et avec le modèle InterVA, les mortalités par cause spécifique ont été calculées pour la communauté en entier et pour des groupes d’âge particuliers.

**Résultats** Une bonne corrélation a été observée entre les résultats des deux systèmes d’interprétation.

**Conclusions** Le modèle InterVA, qui nécessite beaucoup moins de travail que les interprétations par des médecins, est un nouveau et précieux outil pour caractériser la distribution de la mortalité par cause spécifique dans les communautés sans registres de décès et pour comparer les patterns de mortalité entre différentes populations.

**Références**