

Verbal Autopsy interpretation: a comparative analysis of the InterVA model versus physicians' review in determining causes of death in the Nairobi DSS.

Samuel O Oti^{1§}, Catherine Kyobutungi¹

¹African Population and Health Research Center, P.O. Box 10787 GPO-00100,
Nairobi, Kenya

[§]Corresponding author

Email addresses:

SOO: soti@aphrc.org

CK: ckyobutungi@aphrc.org

Abstract

Background

Developing countries generally lack complete vital registration systems that can produce cause-of-death (COD) information for health planning in their populations. As an alternative, Verbal Autopsy (VA) –the process of interviewing family members or caregivers on the circumstances leading to death, is often used by Demographic Surveillance Systems (DSS) to generate COD data. Physicians review (PR) is the commonest method of interpreting VA but this method is a time- and resource consuming process and is liable to producing inconsistent results. The aim of this paper is to explore how a computer-based probabilistic model (InterVA) performs in comparison with PR in interpreting VA data in the Nairobi Urban Health DSS (NUHDSS).

Methods

Between August 2002 and December 2008, a total of 1823 VA interviews were reviewed by physicians in the NUHDSS. Data on these interviews were entered into the InterVA model for interpretation. Cause-specific mortality fractions were then derived from the COD data generated by the physicians and by the model. We then estimated the level of agreement between both methods using Kappa statistics.

Results

The level of agreement between individual causes of deaths assigned by both methods was only 35% ($\kappa=0.27$, 95% CI: 0.25 – 0.30). However, the patterns of mortality as determined by both methods showed a high burden of infectious diseases including HIV/AIDS, TB and Pneumonia in the study population. These mortality patterns are consistent with existing knowledge on the burden of disease in poor underdeveloped communities in Africa

Conclusions

The InterVA model showed ~~very~~ promising results as a community-level tool for generating COD data from VAs. We recommend further refinement to the model, its adaptation to suit local contexts, and its continued validation with more extensive data from different settings.

Background

Developing countries generally lack consistent, timely and reliable information on the levels and cause of death patterns in their populations [1]. Information about causes of death is needed by health managers and policy makers at every level of governance – from local to national – to plan for, prioritize and address the health needs of their people [2]. In developed countries, this information is made available usually through well established vital registration systems [3]. Unfortunately, for most of the developing world, where already scarce resources need to be carefully and optimally allocated, vital registration systems are weak and information on causes-of-death is simply incomplete or non-existent [1, 2].

Alternative sources of cause-of-death (COD) information are from Demographic Surveillance Systems (DSS) which in the last ten years have been receiving a lot of attention for their ability to provide invaluable field data on mortality patterns in developing countries [4, 5]. DSS monitor and track demographic and health indicators in a population within a defined geographical area [6]. Typically, DSS record vital events of births, deaths and migrations within the population under surveillance, at regular intervals ranging from quarterly to annually. Although DSS collect information from defined populations, the vital data that they generate can be linked to similar data from other DSS or sample registration systems within the same country or region to produce more representative data [7]. For instance, the International Network of field sites with continuous Demographic Evaluation of Populations and Their Health (INDEPTH) has produced a monograph series that presents comparative age-specific mortality patterns for INDEPTH field sites across Africa and Asia [8]. Such information has been used successfully for health planning

in resource constrained settings. For instance in Tanzania, COD data was utilized for district and national health planning and resulted in reductions in child mortality [9]. [Data from HDSS has also been used to produce life tables for developing countries \[8\] and provided data for estimating regional and global disease burdens and setting priorities \[10-12\].](#) Hence, DSS can and do play a very important role in producing critical information that can be used for planning and management in developing countries which lack routine vital registration systems.

DSS commonly use a methodology known as Verbal Autopsy (VA) to generate COD data. The process entails interviewing the primary caregivers of recently deceased persons to gather information on the circumstances surrounding a death [13][10]. It is based on the premise that the primary caregiver –usually a family member– can recall, volunteer and recognize symptoms experienced by the deceased that can be interpreted later to derive a probable cause of death [14][11]. [The purpose of VA is to describe the cause of death structure at the population level rather than to diagnose the cause of death at individual level.](#)

A VA data collection instrument is used to conduct the VA interview. This instrument is usually a questionnaire which has an open-ended/narrative section for recording a verbatim account of the circumstances surrounding death, and a closed section with filter questions of symptoms and signs of disease and or injury [15][12]. There are numerous and diverse VA questionnaires being used by various DSS and sample registration systems but recently there has been attempts to harmonize these tools internationally [13, 14, 16][10, 11, 13]. Specifically, recent efforts by the World Health Organization (WHO), INDEPTH and other partners to standardize VA data

collection process have resulted in the elaboration of key characteristics of VA data collection instruments [13, 15][10, 12].

Apart from the various tools used for collecting VA data, there are also different methods for interpreting this data to derive probable causes of death. These include physician review, algorithms, and use of neural networks [14][44]. Physician review (PR) is the most commonly used method and typically involves the independent review of VA data by one or more local physicians. These physicians go through each completed VA questionnaire and using the International Classification of Diseases Version 10 (ICD-10) list or an abridged version, assign the single most probable cause of death to each case [17-22][14-19]. ~~Although there have been various attempts at validating validity of PR- with limited success is well established[19, 21] but,~~ there are several concerns that arise from using this methodology to interpret VA data. Firstly, physicians may differ systematically in their methods of interpreting VA data based on their ~~different~~ training, experiences and or perceptions of local epidemiology. Hence there may be inter and intra-coder variability between physicians which may lead to inconsistencies in COD data and also hinder reliable temporal and spatial comparisons of mortality [23, 24][20, 21]. Secondly, the PR process often demands a considerable amount of often scarce physician time, and often incurs considerable costs for remunerating these physicians.

Consequently, various alternative methods to physician's review of VA data have been introduced. These include the use of expert/data driven algorithms, neural networks and a computer based probabilistic model known as InterVA (*Interpreting Verbal Autopsy*). Algorithms and neural networks are said to have the advantage of

being quicker, more transparent and more consistent in comparison to PR [20, 25, 26][17, 22, 23]. However, both of these methods have been explored inconclusively in terms of their validity and hence their use is still not very widespread [20, 26][17, 23]. The use of the InterVA model to interpret VA data is a relatively new methodology that has been recently explored successfully in a number of settings. This computer program is based on Bayes' probability theorem and is said to have the distinct advantage of achieving maximum consistency in interpreting VA data [27-29][24-26]. It also requires minimal time and labour resources especially in comparison to the PR method. Moreover it is freely available in the public domain making it ideal for resource constrained settings [30][27].

The aim of this paper is to explore how the InterVA model performs in comparison with PR in interpreting VA data collected by the Nairobi Urban Health and Demographic Surveillance site (NUHDSS). Since its inception in 2002, the NUHDSS has relied on PR to interpret its VA data and this paper therefore seeks determine the suitability of the InterVA model as a viable alternative to physician interpretation of VA data.

Methods

Study Area and Population

Since 2002, the African Population & Health Research Center (APHRC) has been operating the NUHDSS. The NUHDSS covers the two urban informal settlements (slums) of Korogocho and Viwandani both located about five to ten kilometers from Nairobi, the capital city of Kenya. Viwandani and Korogocho each occupy an area of 0.45 and 0.52 km² respectively and are inhabited by about 60 000 people from over 15 ethnic groups. The population in Viwandani is mainly comprised of labor migrants working in the neighboring industrial area, while that of Korogocho is mainly comprised of long-term settlers mostly engaged in the informal sector. These slum settlements, like most others in Nairobi, are characterized by relatively high crime rates, drug and alcohol abuse, risky sexual behaviors, high unemployment rates, poor access to health facilities, low school participation and extreme poverty compared to other urban residents as well as their rural counterparts [31][28]. Data on individual and household core demographic events (birth, death, in-migration and out-migration) in the two slums are collected on a four-monthly interval –also known as data collection rounds. In addition to the routine data collections rounds, the DSS also integrates the ~~Verbal Autopsy~~ [VA](#) process for COD ascertainment.

Verbal Autopsy

Deaths are usually identified during the DSS data collection rounds by trained field-interviewers who complete a one-page death registration form (DRF) and then inform their field supervisors about the deaths. These supervisors are experienced field-interviewers who have a minimum qualification of a bachelor's degree. They conduct the VA interviews using a VA questionnaire developed in conjunction with other

INDEPTH sites. This questionnaire has two formats; one for deaths among children less than five years of age and the other for deaths of persons five years and older. The latter has an additional section on maternal deaths. The questionnaire covers the background characteristics of the deceased and the respondent as well as structured filter questions on specific signs and symptoms experienced by the deceased up to the point of death. There is also an open section in the form that allows for recording of a narrative account of the events leading to the death. On average, it takes about 30-45 minutes to administer the VA questionnaire if no interruptions occur during the interview.

Before a VA interview is conducted, the field supervisor visits the household in his/her zone where a death has occurred as soon as he/she learns of the event and consoles the bereaved family. He then assesses the situation and decides whether the timing is appropriate to conduct the interview or not. If it is not appropriate he/she makes an appointment with the family to return at an agreed later date –usually 3 to 4 weeks later. However VA interviews may be conducted as much as six months after death due to operational reasons. At the first visit, the field supervisor identifies a ‘credible respondent’ –usually a spouse or relative– who will participate in the interview. If the deceased is a child, the preferred credible respondent is usually a parent. Several revisits may be made to the household until a credible respondent is identified. After five such visits or if it is established that the remaining household members are no longer resident in the area, a credible neighbour is interviewed if he/she is willing. Otherwise the verbal autopsy is coded as missing and no cause of death is assigned to such cases.

Interpretation of VA questionnaires

All completed VA questionnaires are collated and sent to three local physicians for interpretation. At least one of the three physicians is a full-time researcher employed with APHRC while the other two are consultants or medical officers in public or private practice who review VA data on a contractual basis. Each physician independently reviews all the VA questionnaires and assigns a single COD based on ICD-10. The complete ICD-10 list has ~~more than 12,420~~ ~~155,000~~ unique codes for diseases, signs, symptoms, abnormal findings, complaints, social circumstances and external causes of injury [32][29]. Hence, for practical purposes, the physicians use an abridged version of the ICD-10 list which was modified in such a way that uncommon causes of death in the study area are collapsed into broader categories. This modified list has 60 codes for possible COD (See additional file 24). If two of the assigned COD for each VA questionnaire are identical, this is taken as the final COD for the deceased. However, if all three of the assigned COD are different, the physicians hold a consensus meeting and review all such cases. In cases where consensus is not reached even at these meetings, the COD is classified as “indeterminate”.

The InterVA model

The InterVA model is a probabilistic model based on Bayes' theorem that seeks to define the probability of a cause (C) given the presence of a particular indicator (I), represented as P(C|I). This probability can be stated as:

$$P(C|I) = \frac{P(I|C) \times P(C)}{P(I|C) \times P(C) + P(I|\bar{C}) \times P(\bar{C})}$$

where P(!C) is the probability of not (C). Therefore, for a set of VA symptom level data or indicators (I1...In) and each possible cause of death resulting from these indicators (C1...Cm), there is an associated indicator Ij and cause Ck, whose probability of occurrence at population level can be determined. For each case therefore, the probability of Ck is *ab initio* the value found among all deaths in total, which gives the cause specific mortality fraction (CSMF). For each case however, and each applicable indicator, the above theorem, can modify the probability of Ck. Thus, the VA model adjusts the probability of each likely cause according to a matrix of P((I1...In)|(C1...Cm)) and then produces a summary listing of as many as three possible causes and their corresponding likelihood value [28][25]. Full details of the InterVA model and how it was developed based on the above theorem have been described in previous studies [27-29][24-26]. In summary, the model is run using computer software –Visual FoxPro– which provides a user interface into which a set of 100 indicators must be entered for each VA case in order for the model to generate a COD. These indicators are basically specific information comprising reported symptoms, signs and medical history which need to be extracted from completed VA questionnaires. Some examples of the required indicators include ‘Any difficulty in breathing?’, ‘Any weight loss?’ and ‘Any coughing with blood?’ among others. Thus,

when the model is run on these indicators, it automatically generates a listing of any of 30 probable COD for each verbal autopsy case (See additional file [24](#) for COD listing). A maximum of three probable causes of deaths and their corresponding likelihoods (in percentages) are presented in the list.

Additionally, the InterVA model has ~~provisions~~ [an in-built facility](#) to adjust for the prevalence of malaria and HIV/AIDS in any setting such that before running the model, the prevalence of HIV/AIDS and malaria in the study population can be set as 'high' or 'low'. This [was introduced during a process of refining the original model to address underlying conceptual issues of VA data collection and interpretation. The Delphi technique using a panel of experts was utilised to develop consensus on key conceptual issues of cause of death classification and VA usage including adjustment for large variations in prevalence of the malaria and HIV/AIDS at the population level between regions. This adjustment significantly improved the performance of the model and increased the model's potential to be applied in different settings \[27, 28\]. Details of the methods and how the in-built facility was developed are beyond the scope of this paper.](#) For our study we set the prevalence of malaria to be 'low' and that of HIV/AIDS to be 'high'. Previous research in our study population has demonstrated that the prevalence of malaria is less than 0.5%, while HIV/AIDS prevalence is as high as 12.4% [\[33, 34\]](#)~~[30, 31]~~.

Data analysis

Between August 2002 to December 2008, 1823 VA questionnaires were reviewed by physicians who assigned COD to each case. The required indicators from each of these questionnaires were extracted and entered into the InterVA model to automatically generate COD.

There ~~were~~are a total of 60 possible causes of death assigned by physicians while the InterVA model only assigns 27 causes (see additional files 21). Therefore, to allow for meaningful comparison between physicians and the model, we re-categorized all causes in both methods into 14 main groups of causes for two main reasons. First, to have comparable cause of death categories between both methods being analyzed; where possible we retained the categories that were common to both methods. For instance, 'Malaria' and 'Meningitis' which are common to both physician review and InterVA were retained as standalone causes. In cases, where there were no direct correlates, we had to collapse and or re-categorize the causes of death into cause groups to match each other in a broad sense. For instance, the InterVA model has only one broad category of 'maternity-related deaths' representing all types of pregnancy related deaths. However, the physicians coded causes such as 'eclampsia' and 'ante-partum and post-partum haemorrhage.' Such causes were therefore recoded into one broad category of 'maternity-related deaths' so that we could compare with the corresponding InterVA category. For instance, 'kidney disease', 'diabetes' and 'malignancies' were all grouped into "chronic/Non Communicable Diseases (NCD)". And causes like 'rabies', 'tetanus,' and 'typhoid' were grouped into "other acute/infectious diseases". However, frequently occurring conditions such as pulmonary tuberculosis, HIV/AIDS, and pneumonia were left as stand-alone causes. Second, it was more important to us that the model and the physicians can first and foremost arrive at broad agreement in identifying cause of death groups with the most public health importance at population level than individual level causes. Hence, causes such as 'kidney disease' and 'cancers' were recoded as 'chronic diseases' while causes like rabies, tetanus and typhoid were grouped into other acute/infectious diseases..

We then determined the cause-specific mortality fractions (CSMF) of using the InterVA model and physician review. We conducted our analysis for the general population and by two main age groups: children aged less than 5 years and for adults aged 18 years and older. While it is possible to conduct the analysis across various age categories, we decided to focus on the under-five and adult deaths due to high levels of mortality in these age groups from preventable conditions such as diarrheal disease and HIV/AIDS respectively [35][32]. Such preventable conditions are of huge public health significance especially in developing countries and so we were particularly interested in investigating how the model performed within the selected age groups. Thereafter, we estimated the level of agreement between InterVA and physician assigned COD using Kappa statistics. All analyses were carried out using STATA version-10 statistical software. In all our analyses, we only considered the most probable COD assigned by the model rather than all three possible causes. This is because the COD assigned by the physicians included only a single cause of death.

Results

A total of 1823 VA interviews were successfully completed and reviewed by physicians for the period between August 2002 and December 2008. Children aged less than 5 years accounted for 572 cases (31.4%) and adults aged 18 years or older were 1166 (64%). Of the under-5 deaths, 384 (67%) occurred before the first year of life. Overall, males accounted for 1020 (56%) deaths and Korogocho slum had the majority (66%) of deaths in the two slum areas.

For the 1823 deaths, physicians successfully assigned a single cause in 1443 (79%) cases on the first attempt. However after holding a consensus meeting, they successfully assigned single causes of death to a further 252 (14%) cases. Therefore, in total, physicians assigned a single cause of death in 1,695 cases (93%). No consensus was reached in 128 cases (7%) which were coded as “indeterminate” by the physicians. The InterVA model assigned a single or primary cause of death in 1290 cases (70.8%), two causes of death in 152 cases (8.3%) and three causes in six cases (0.3%). In 375 cases (20.6%), the InterVA model assigned the cause of death as “indeterminate”.

A direct comparison of the causes of death assigned by the physicians with and by the primary causes of death assigned by the InterVA model (Table 1) shows that overall, there was direct agreement in 630 cases or 35% ($\kappa=0.27$, 95% CI: 0.25 – 0.30). If all “indeterminate” cases were dropped, the level of agreement increased to 47% ($\kappa=0.40$, 95% CI: (0.36 – 0.42)). The level of agreement was 32% for deaths in children aged less than five years and 36% for adults 18 years and older.

[Table 1: About here]

We did some further analysis to compare the level of agreement between the physicians and the InterVA model among the 152 cases which each had two likely causes of death as assigned by the model. We did not analyze cases with three likely causes because they were too few (only 6 cases). Our results showed that among these 152 cases, the level of agreement was 31% ($\kappa = 0.23$, 95% CI 0.19 – 0.25) between the physicians and the primary cause assigned by the model. However the level of agreement drops to 26% ($\kappa = 0.17$, 95% CI 0.11 - 0.25) if we compare the physicians' causes with the second most likely causes of death as assigned by the model.

Figure 1 shows the major COD categories and the numbers of deaths in each category as assigned by physicians and the InterVA model. Overall, majority of deaths were attributable to infectious diseases including pulmonary tuberculosis, pneumonia, and HIV/AIDS. Specifically, 1051 cases (58%) were attributed to infectious causes by physicians and 1109 cases (61%) by the InterVA model. However, 249 cases (14%) were attributed to non-communicable diseases by physicians while only half of this number, 126 cases (7%), was attributed as such by the InterVA model. Physicians assigned 257 cases (14%) to injuries while the model assigned this cause to 127 cases (8.8%). It is noteworthy that the InterVA model attributed significantly more causes of death to pulmonary tuberculosis (401 cases) compared to physicians (131 cases). However, physicians attributed more causes of death to HIV/AIDS (419 cases) compared to InterVA (310 cases).

For children aged less than five years, there were 572 deaths that were assigned causes by both physicians and the model. Figure 2 shows the CSMF as assigned by

physicians and the model. Of all deaths in this age group, majority (382 cases or 67%) were due to infectious diseases as assigned by physicians. An even larger number of cases (446 or 78%) were attributed to infectious diseases by the model. A notable difference between the two methods is the frequencies with which deaths in this age group were attributed to HIV/AIDS. The InterVA model attributed 110 out of 572 (19.2%) deaths to HIV/AIDS whereas the physicians only attributed 10 (1.7%) deaths to HIV/AIDS. On the other hand, physicians attributed 105 (18.4%) deaths to diarrheal diseases while the model attributed only 48 (8.4%) deaths. Also, the model assigned twice as many deaths to pulmonary tuberculosis as the physicians. Other infectious diseases such as pneumonia, malaria and measles only showed slight differences in proportions assigned by the two methods. Injuries and non-communicable diseases accounted for less than 8% of deaths as assigned by both methods.

There were notable differences in the proportions of deaths assigned to preterm/perinatal conditions and malnutrition. There were three times more deaths (72 cases) attributed to preterm/perinatal causes by physicians in comparison to the model (24 cases). Also, there were 37 cases attributed to malnutrition by physicians while the model did not assign any cause of death to malnutrition. Among deaths to children less than five years old, the model arrived at an “indeterminate” cause in 90 cases (15.7%) while physicians did the same in 37 cases (6.3%).

For the 1166 deaths in persons aged 18 years and older, we found that infectious causes accounted for just over half of all deaths as assigned by physicians (54%) and the model (53%). The main difference again lies in the proportions of deaths attributed to tuberculosis and HIV/AIDS (see figure 3). Physicians assigned 401 cases

(34%) to HIV/AIDS and 117 cases (9.9%) to PTB. However, the model assigned 194 cases (16%) to HIV/AIDS and 365 cases (31%) to PTB. As regards non-communicable diseases, we found that physicians attributed twice as many deaths than the model – that is, 18% and 9% respectively. A difference of similar magnitude was also found in deaths attributed to injuries –that is, 19% attributed by physicians and 11.7% by the model. Finally, we found that physicians arrived at an “indeterminate” cause in 86 cases (7%) whereas the model did so in 275 cases (23%).

Discussion

The level of agreement between physician review and the InterVA model was less than 40% of the 1823 deaths interpreted by both methods. Other similar studies have shown higher levels of agreement, ranging from 50% to 83%, albeit with much smaller sample sizes [27-29][24-26]. However, it is ~~very~~ encouraging that the overall CSMF for the major causes of death in our study population were picture of CSMF for the major causes of death in our study population was somewhat quite similar as determined by both methods. This is despite the fact that the InterVA model was applied independently and at a single time point to VA data which had been previously reviewed by different physicians at various time points over a period of almost eight years. Even more encouraging is the finding that the patterns of mortality data generated by both methods were consistent with those found in previous studies on burden of mortality in slum populations and other disadvantaged populations in less developed countries [35, 36][32, 33].

However, there are several differences and contradictions in the assignment of causes of death by either method that need to be discussed further. Firstly, we found that the InterVA model more frequently identified pulmonary tuberculosis as a cause of death than was done by physicians. On the other hand physicians more frequently identified HIV/AIDS as a cause of death than the model. This however is not entirely surprising as there is a great deal of overlap between both disease conditions in terms of clinical symptoms and signs [37][34]. Furthermore, the re-emergence of pulmonary tuberculosis in several countries of the world is spurred by the HIV/AIDS pandemic [38][35]. This underlies the high level of interconnectedness between both disease entities. Moreover, from a public health perspective, control and prevention of either

disease cannot be considered without due regard to the other [39][36]. Hence, what is critical is that the composite burden of both diseases in any population comes out clearly and the InterVA model achieved this as successfully as the physicians did.

As regards non-communicable diseases, we found that physicians identified this category as a cause of death almost twice as much as the model. Further inspection of the data revealed that one-quarter of the physician-assigned non-communicable disease deaths were actually identified as PTB or HIV/AIDS by the model (See additional file 12). An even slightly larger proportion was identified by the model to be “indeterminate”. A similar magnitude of differences was also observed in the identification of deaths due to “injuries”. Again, physicians assigned more than twice the number of injuries than the model. However, unlike the non-communicable disease cases, the model settled for a diagnosis of “indeterminate” in virtually all the cases which differed with the physicians’ diagnoses of injury. The reasons for this difference in coding non-communicable diseases are not immediately clear and it is difficult to say whether or not the physicians’ diagnosis was more appropriate than the model.

However, for the injuries it is clear that the model did not have enough information to arrive at a diagnosis and simply settled for indeterminate. This is largely due to the design of the VA questionnaires which is structured in such a way that the richest source of clear detail information on the circumstances surrounding injuries is the narrative section of the questionnaire. The structured/close-ended part of the questionnaire is more suited for symptoms and signs of disease rather than circumstances leading to injury. Hence, most of the completed VA questionnaires for

cases of injury have very sparse detail in the structured part. Physicians, unlike the model, may therefore have ~~the~~ benefited ~~offrom~~ reading the open narrative section of the VA ~~form~~ questionnaire and as this ~~could~~ would certainly have contained more information on injuries than what was captured in the closed-ended part of the questionnaire which was used for the model.

In children less than five years of age, we find that the model identified HIV/AIDS as a cause of death at a much higher frequency than physicians who mostly diagnosed such cases as either diarrheal diseases or malnutrition. A similar finding has also been demonstrated in previous studies. The explanation given in these studies is that considering that malnutrition and diarrheal disease are major causes of under-five mortality in poor and undeveloped settings [40][37], it is likely that the model overestimated HIV/AIDS prevalence in children while physicians may have underestimated it. We suspect that a similar possibility may have occurred in our study. We however cannot fully explain why the model did not assign any cause of death to malnutrition. We observed that majority of the cases of malnutrition diagnosed by physicians were identified as HIV/AIDS ~~and PTB~~ by the model.

Specifically, of the 37 cases of malnutrition diagnosed by physicians, the model assigned 18 cases to HIV/AIDS, five to TB, four to Pneumonia, four to Malaria, and six to other conditions. This is not so surprising considering that symptoms like chronic diarrhoea and weight loss are common features ~~of~~ to some of these conditions as well as to both HIV/AIDS and malnutrition. Furthermore, most of these conditions are very likely to either co-exist with or complicate malnutrition. ~~Our~~ Also, our experience in VA interpretation suggests that there is weak reporting of certain key features of malnutrition such as abnormal hair changes, abdominal swelling and

oedema which may have impaired the model's ability to arrive at this diagnosis. It may therefore be easier for physicians to spot malnutrition as an underlying cause of death especially through the additional benefit of the narrative section of the VA data not considered by the model.

Other differences in the interpretation of VA data by physicians and by the model include assignment of “preterm/perinatal” deaths in children by either method. We observed that about one-third of the cases identified by physicians as “preterm/perinatal” deaths were assigned as “pneumonia/sepsis” by the model. This finding was similar to what was observed in another study in a rural community in Ethiopia [29][26]. Just like in that study, we also believe that the reasons for our observed differences in interpretation arose most probably from differences in definitions of this category of deaths. Specifically, it should be emphasized that, for ease of analysis, the term “preterm/perinatal” was used broadly to include all causes of death in the early neonatal period including conditions like birth asphyxia, birth injuries, congenital deformities, neonatal sepsis and neonatal jaundice among others. Furthermore, it should be noted that the InterVA model's age grouping has ‘< 4 weeks old’ as its lowest age group. This means that unlike physicians, the model is unable to distinguish deaths in the perinatal period (first seven days of life) from other neonatal deaths.

The low level of agreement between physicians' review and InterVA model in this study highlights several limitations of either method. As far as InterVA is concerned, it is important to reiterate that the model has a pre-defined set of indicators which it

depends upon to arrive at causes of deaths. Hence these indicators need to be extracted from the VA data and fed into the model. Some of these indicators were not available in the VA data as there were no specific questions in the VA questionnaires that collected any information on these indicators. For instance the VA questionnaires do not have any questions on the history of the deceased's alcohol or tobacco use, raised or lowered fontanelles, symptoms or signs of herpes, excessive thirst, excessive hunger, history of liver disease, history of sickle cell anemia and history of cancer. Also there are several questions in the VA questionnaires that were not built into the InterVA- indicator set. For instance the VA questionnaires have useful questions pertaining to drug/treatment history and health service utilization. It is therefore unclear how the absence of the above information may have affected the performance of the InterVA model. There is therefore scope, on one hand, to revise the VA questionnaires to include the missing InterVA indicators and on the other hand, to expand the list of indicators already in the model to include more information from the VA questionnaires.

With regards to physicians, we must consider that, as previously mentioned, they have the added advantage of being able to utilize the narrative section of the VA questionnaires. The narrative section often contains a lot of redundant details which the physicians have to sift through to get the most relevant facts pertaining to the death under review. The InterVA model can only utilize the narrative section if key words from the text are manually extracted and fitted into the model. For a large number of VA questionnaires, this would be a laborious and time consuming process that would preclude the argument that the model is a time-saving alternative to PR. Also, the physicians are able to use their clinical skills and experiences to make a final

judgment between disease entities that may have strikingly similar symptoms and signs. However, physicians may also be influenced by their own experience-driven biases which have traditionally raised issues of inter-observer reliability and the hindrance that this poses to temporal and regional comparisons of COD. Furthermore, it is clear that PR of VA data is not a “gold-standard” and this has raised concerns of validating other interpretation methods against it. It is precisely for this reason that we have avoided using measures such as sensitivity, specificity and positive predictive value in our comparative analysis of InterVA model against PR data. Additionally, while it would have been ideal to validate the model with respect to hospital COD as “gold-standard” as has been done in other studies [\[17, 19\]](#)~~[14, 16]~~ we could not do so in our study. This was because the majority of deaths in the slums occur outside hospital settings (about 80% in the NUHDSS). Hence, the deaths that occur in a hospital may differ selectively from those outside it and this will inherently bias the validation process.

As we had mentioned previously, PR is both time consuming and labor intensive and prone to inter-observer variation. Neither of these issues are a problem with the InterVA model which provides 100% consistency, minimal level of effort and very short turnaround time in its application to VA interpretation. Additionally, physicians often provide a single cause of death per case, with no place for secondary or underlying causes. In clinical practice, the certification of deaths requires the designation of primary or immediate and secondary or underlying causes of death. The InterVA model is able to provide as many as three possible causes of death, hence giving it an advantage over the traditional PR process. ~~However, with-~~

proper training, physicians who code VA data can easily assess underlying,
immediate and contributing causes of death.

Conclusions

This study has been able to ~~successfully~~ apply the InterVA model to a large number of VA deaths unlike previous studies which have used much smaller number of deaths. Overall, the performance of the model in VA interpretation was only satisfactory for a few conditions which however are of huge great public health significance in the context of the study area setting ~~is satisfactory. This is encouraging given the fact that the model and the VA questionnaires used in this study were developed independently of each other. Therefore, We are of the opinion that~~ Although the model may be limited in its ability to identify COD at the individual level, it ~~is clearly a ve~~ has some potential as ~~ry an~~ innovative community-level tool for identifying COD patterns. This is particularly highlighted in our findings of somewhat comparable CSMFs for the commonest COD as interpreted by the model and PR. From a public-health perspective, this is ~~very crucial~~ useful because the model paints a fairly accurate picture of the disease burden in the study population. Potentially therefore, t ~~This information will be~~ could be invaluable to guiding health policy, programmes and interventions in a resource constrained and data-deprived settings such as the slums of Nairobi.

However, there is certainly a lot of scope for improving the model and specifically, making allowances for local context-specific features to be built therein. A more robust model may be one that for instance, is tailor made to fit a standardized VA questionnaire such as the Sample Vital Registration using Verbal Autopsy (SAVVY) questionnaires developed by the WHO. There is also the small but important issue of how to adequately incorporate the multiple probable causes of death generated by the model, which our study did not address. Hence, the next steps will be to refine the

model, continue its validation with more extensive data from different settings and give further thought to the interpretation and analysis of multiple causes of death for individual cases. It will also be useful for the developers of the model to consider programming the model in such a way that it ~~presents~~computes credible intervals that ~~reflect~~ uncertainty estimates in the model's predictions. In its current version, the model only presents point predictions which may be misleading to its users.

Competing interests

The authors declare that they have no competing interests.

Authors' contributions

SOO did the literature review, data analysis and wrote the first draft of the manuscript.

CK contributed to the data analysis plan, participated in coding verbal autopsies, writing of the paper and interpretation of the findings.

Acknowledgements

We wish to acknowledge the contribution of APHRC's dedicated field and data management teams. We acknowledge the contributions of Dr. Alex Ezeh and Dr. Eliya Zulu to the conceptualization of the NUHDSS. The authors are supported by a grant from the Wellcome Trust UK (grant number GR078530AIA). Work in the NUHDSS is supported by grants from the Rockefeller Foundation, and the William and Flora Hewlett Foundation.

References

1. [Setel P.W., Macfarlane S.B., Szreter S., Mikkelsen L., Jha P., Stout S., et al. \(2007\). A scandal of invisibility: making everyone count by counting everyone. *Lancet*.370:1569-77.](#)
2. [Byass P. \(2007\). Who needs cause-of-death data? *PLoS Medicine*.4.](#)
3. [Declich S., Carter A.O. \(1994\). Public health surveillance: historical origins, methods and evaluation. *Bulletin of the World Health Organization*.72\(2\): 285–304. .](#)
4. [Baiden F., Hodgson A., Binka F.N. \(2006\). Demographic surveillance sites and emerging challenges in international health. *Bulletin of the World Health Organization*.84\(3\).](#)
5. [Carrel M., Rennie S. \(2008\). Demographic and health surveillance: longitudinal ethical considerations. *Bulletin of the World Health Organization*.86\(8\):577-656.](#)
6. [Byass P., Berhane Y., Emmelin A., Kebede D., Andersson T., Högberg U., et al. \(2002\). The role of demographic surveillance systems \(DSS\) in assessing the health of communities: An example from rural Ethiopia *Public Health*.116\(3\):145-50](#)
7. [INDEPTH. International Network of field sites with continuous Demographic Evaluation of Populations and Their Health in developing countries; Founding Documents. \[cited 2009 August 27\]; Available from: \[http://www.indepth-network.org/index.php?option=com_content&task=view&id=43&Itemid=129\]\(http://www.indepth-network.org/index.php?option=com_content&task=view&id=43&Itemid=129\).](#)
8. [INDEPTH. INDEPTH Monograph series: Population, Health and Survival at INDEPTH Sites. Ottawa: IDRC; 2002](#)
9. [de Savigny D., Kasale H., Mbuya C., Reid G. In focus: Fixing Health Systems. Ottawa: International Development Research Centre; 2005.](#)
10. [Korenromp E.L., Williams B.G., Gouws E., Dye C., Snow R.W. \(2003\). Measurement of trends in childhood malaria mortality in Africa: an assessment of progress toward targets based on verbal autopsy. *Lancet Infectious Diseases* 3:349-58.](#)
11. [Morris S.S., Black R.E., Tomaskovic L. \(2003\). Predicting the distribution of under-five deaths by cause in countries without adequate vital registration systems. . *International Journal of Epidemiology* 32:1041-51.](#)
12. [Murray C.J.L., Lopez A.D., editors. The Global Burden of Disease. A comprehensive assessment of mortality and disability from diseases, injuries and risk factors in 1990 and projected to 2010. The Harvard School of Public Health on behalf of the World Health Organization and the World Bank. Boston: Harvard University Press; 1996.](#)
13. [Baiden F., Bawah A., Biai S., Binka F., Boerma T., Byass P., et al. \(2007\). Setting international standards for verbal autopsy. *Bulletin of the World Health Organization* 85\(8\).](#)
14. [Soleman N., Chandramohan D., Shibuya K. \(2006\). Verbal autopsy: current practices and challenges. *Bulletin of the World Health Organization*.84:239-45.](#)
15. [WHO. WHO technical consultation on verbal autopsy tools. Geneva; 2005.](#)
16. [Garenne M., & Fontaine, O. \(2006\). Potential and limits of verbal autopsies. *Bulletin of the World Health Organization*.84\(3\).](#)
17. [Chandramohan D., Maude D., Rodrigues L.C., Hayes R.J. \(1998\). Verbal autopsies for adult deaths: their development and validation in a multicentre study. *Tropical Medicine and International Health*.3:436-46.](#)
18. [Gajalakshmi V., Peto R. \(2006\). Verbal autopsy procedure for adult death. *International Journal of Epidemiology*.35:748-50.](#)

19. [Kahn K., Tollman S.M., Garenne M., Gear J.S. \(2000\). Validation and application of verbal autopsies in a rural area of South Africa. *Tropical Medicine and International Health*.5\(11\):824-31.](#)
20. [Quigley M.A., Chandramohan D., Rodrigues L.C. \(1999\). Diagnostic accuracy of physician review, expert algorithms and data derived algorithms in adult verbal autopsies. *International Journal of Epidemiology*.28:1081-7.](#)
21. [Setel P.W., Whiting, D.R., Hemed, Y., Chandramohan, D., Wolfson, L. J., Alberti. K.G.M.M. & Lopez, A.D. \(2006\). Validity of verbal autopsy procedures for determining cause of death in Tanzania. *Tropical Medicine and International Health*.11 \(5\):681–96](#)
22. [Yang G., Rao C., Ma J., Wang L., Wan X., Dubrovsky G., et al. \(2006\). Validation of verbal autopsy procedures for adult deaths in China. *International Journal of Epidemiology*.35:741-8.](#)
23. [Ronsmans C., Vanneste A.M., Chakraborty J., Van Ginneken J. \(1998\). A comparison of three verbal autopsy methods to ascertain level and causes of death in Matlab, Bangladesh. *International Journal of Epidemiology*.27:660-6.](#)
24. [Todd J.E., De Francisco A., O'Dempsey T.J., Greenwood B.M. \(1994\). The limitations of verbal autopsy in a malaria endemic region. *Annals of Tropical Paediatrics*.14:31-6.](#)
25. [Bouille A., Chandramohan D., Weller P. \(2001\). A case study of using artificial neural networks for classifying cause of death from verbal autopsy. *International Journal of Epidemiology*.30\(515\):520.](#)
26. [Quigley M.A., Chandramohan, D., Setel, P., Binka, F., & Rodrigues, L.C. \(2000\). Validity of data-derived algorithms for ascertaining causes of adult death in two African sites using verbal autopsy. *Tropical Medicine and International Health*.5 \(1\):33–9.](#)
27. [Byass P., Huong D.L., Minh H.V. \(2003\). A probabilistic approach to interpreting verbal autopsies: methodology and preliminary validation in Vietnam. *Scandinavian Journal of Public Health*.31\(Suppl.62\):32-7.](#)
28. [Byass P., Fottrell E., Huong D.L., Berhane Y., Corrah T., Kahn K., et al. \(2006\). Refining a probabilistic model for interpreting verbal autopsy data. *Scandinavian Journal of Public Health*.34:26-31.](#)
29. [Fantahun M., Fottrell E., Berhane Y., Wall S., Högberg U., Byass P. \(2006\). Assessing a new approach to verbal autopsy interpretation in a rural Ethiopian community: the InterVA model. *Bulletin of World Health Organization*.84:204-10.](#)
30. [InterVA. \[cited 2009 August 27\]; Available from: \[www.interva.net\]\(http://www.interva.net\).](#)
31. [APHRC. Population and Health Dynamics in Nairobi Informal Settlements. Nairobi Cross-Sectional Slums Survey 2000. Nairobi; 2002.](#)
32. [WHO. International Classification of Diseases \(ICD\). \[cited 2010 May 03\]; Available from: <http://www.who.int/classifications/help/icdfaq/en/index.html>.](#)
33. [APHRC. The Economic, Health, and Social Context of HIV Infection in Informal Urban Settlements of Nairobi In press 2009.](#)
34. [Ye Y., Madise N., Ndugwa R., Ochola S., Snow R.W. \(2009\). Fever treatment in the absence of malaria transmission in an urban informal settlement in Nairobi, Kenya. *Malaria Journal* 8\(160\).](#)
35. [Kyobutungi C., Ziraba A.K., Ezeh A., Yé Y. \(2008\). The burden of disease profile of residents of Nairobi's slums: Results from a Demographic Surveillance System. *Population Health Metrics* 6:1.](#)

36. van Eijk A.M., Adazu K., Ofware P., Vulule J., Hamel M., Slutsker L. (2008). Causes of deaths using verbal autopsy among adolescents and adults in rural western Kenya. *Tropical Medicine and International Health*.13 (10):1314-24.
37. Keshinro B., Diul M.Y. (2006). HIV-TB: epidemiology, clinical features and diagnosis of smear-negative TB. *Tropical Doctor*.36(2):68-71.
38. Myers J., Sepkowitz K. (2008). HIV/AIDS and TB *International Encyclopedia of Public Health*.421-30.
39. Abdool Karim S.S., Churchyard G.J., Abdool Karim Q., Lawn S.D. (2008). HIV infection and tuberculosis in South Africa: an urgent need to escalate the public health response. *Lancet*.In press.
40. Bryce J., Boschi-Pinto C., Shibuya K., Black R.E. (2005). WHO estimates of the causes of death in children. *Lancet*.365:1147-52.

Figures

Figure 1 - Representation of the major cause of death categories derived from physicians review and InterVA model interpretation of 1823 deaths in the NUHDSS 2002 -2008

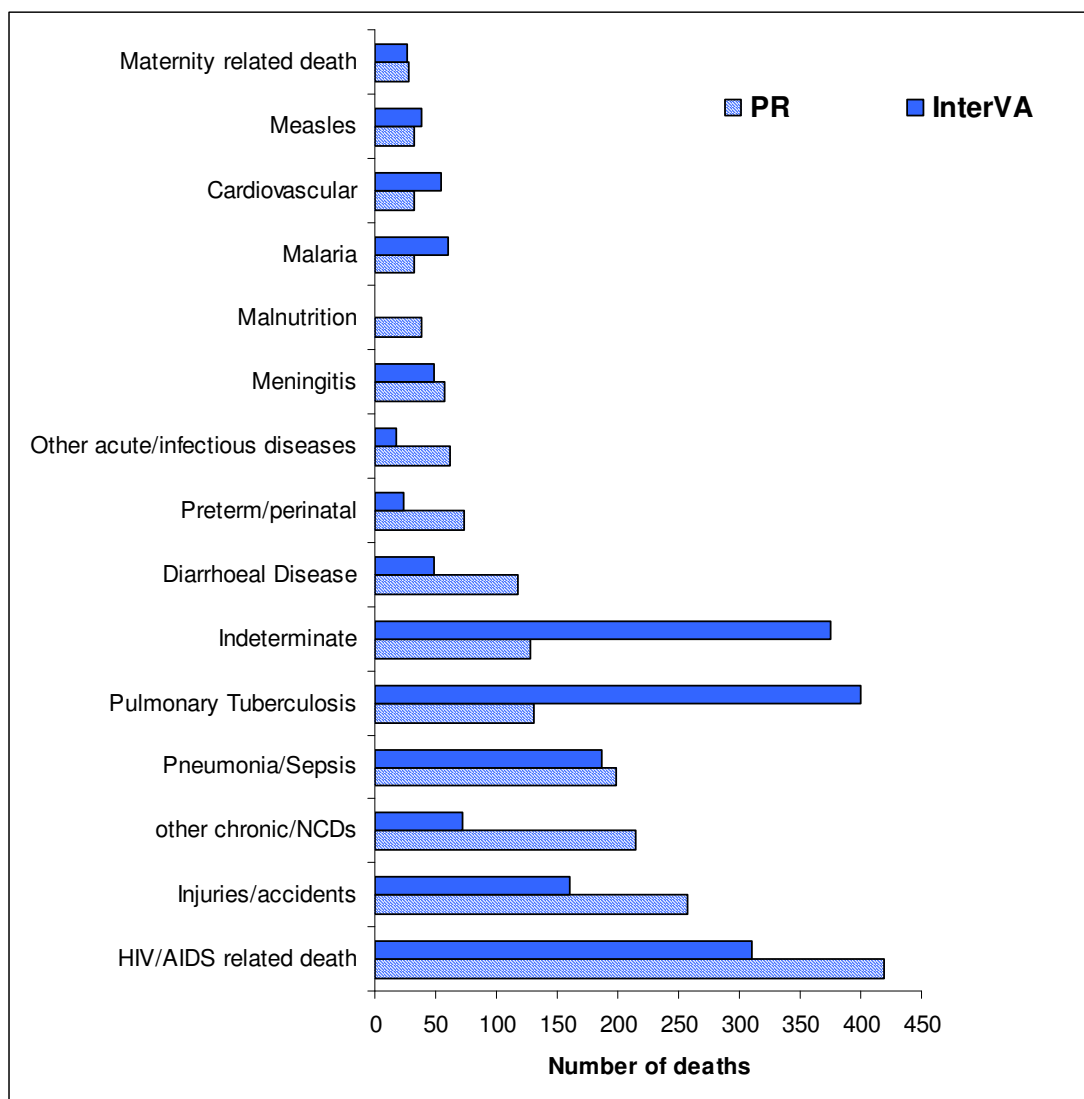
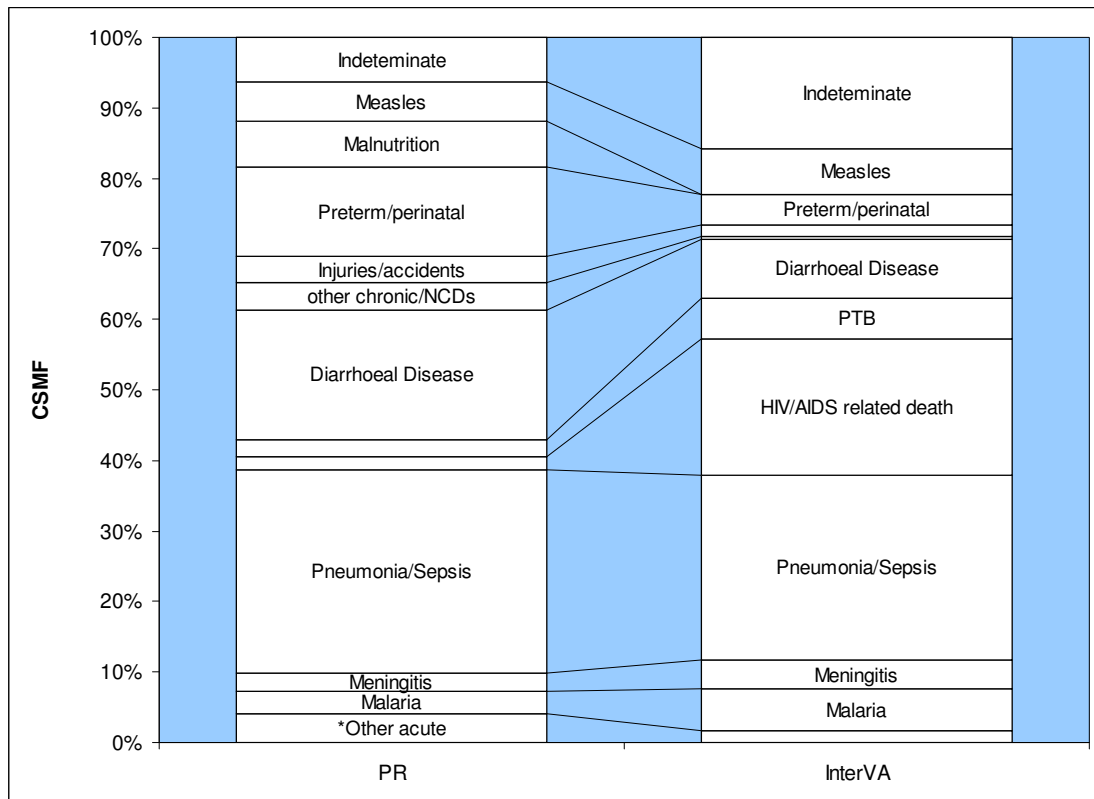
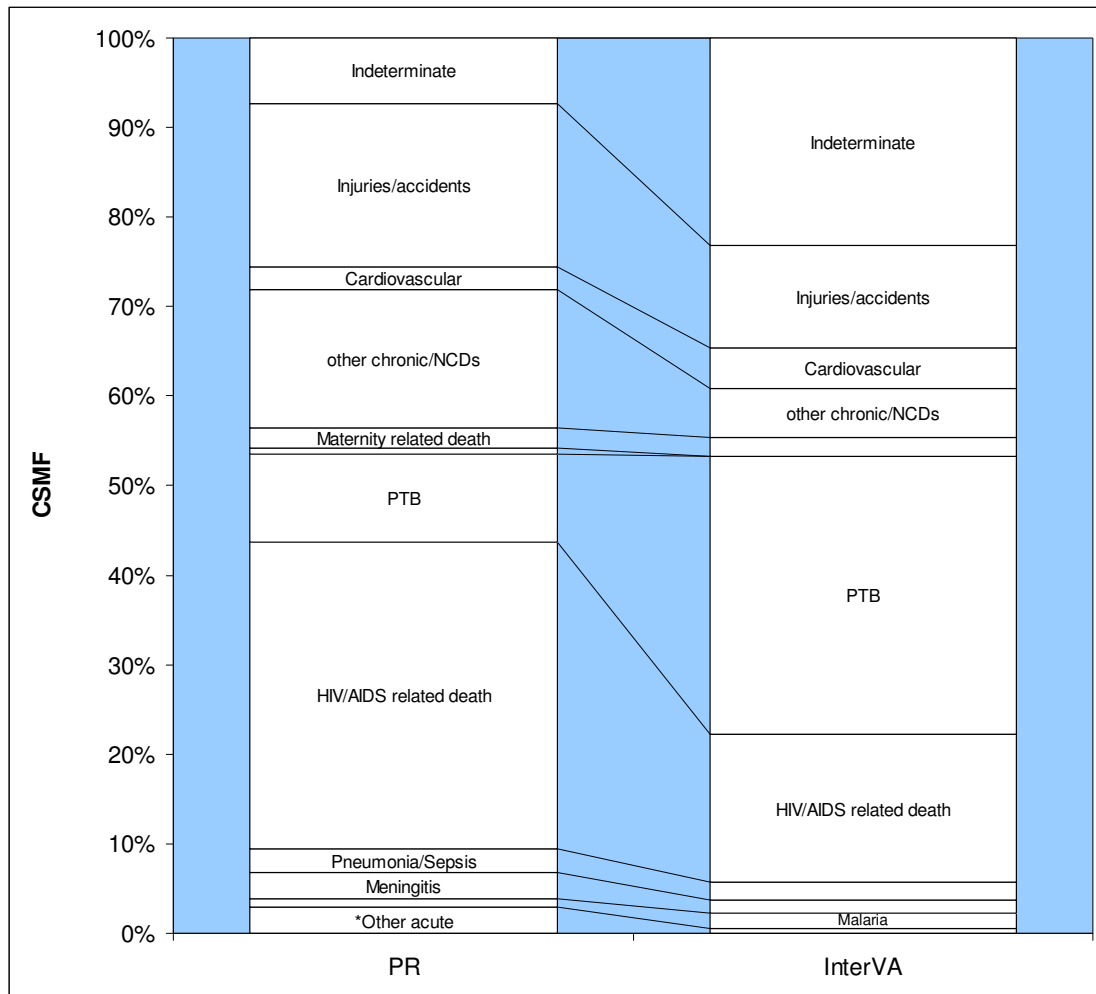


Figure 2 - Cause-specific mortality fractions for 572 deaths of children aged less than five years in the NUHDSS from 2002 - 2008, derived from verbal autopsies interpreted by physicians and by the InterVA model



*Other acute/infectious diseases. PTB, Pulmonary Tuberculosis.

Figure 3 - Cause specific mortality fractions for 1166 adult deaths in the NUHDSS from 2002 - 2008, derived from verbal autopsies interpreted by physicians and by the InterVA model



*Other acute/infectious diseases. PTB, Pulmonary Tuberculosis.

Tables

Table 1 Summary of case-by-case agreement between PR and InterVA

	Cases (deaths) in agreement	Kappa (95%CI)
Children aged less than five years	181 (31.64%)	0.224 (0.183 - 0.253)
Adults (18 years and older)	204 (35.59%)	0.254 (0.222 - 0.282)
Overall	630 (34.56%)	0.271 (0.247 - 0.293)

CI, Confidence Interval.

Additional files

~~Additional file 1 – Spreadsheet showing cause of death categories (including abridged ICD-10 code list) by physicians and InterVA model~~

~~Additional file descriptions text: MS Excel format.~~

~~Filename: *COD_disease_codes_PRvsInterVA.xls*~~

Additional file 12 – Table showing a comparison of cause of death assessment by PR and most likely cause by InterVA model, for 1823 verbal autopsies from NUHDSS.

Additional file descriptions text: Landscape table in MS Word 2003 format.

Filename: *Table_2_cross_tab.doc*

Additional file 2 – Spreadsheet showing cause of death categories (including abridged ICD-10 code list) by physicians and InterVA model

Additional file descriptions text: MS Excel format.

Filename: *COD_disease_codes_PRvsInterVA.xls*

Additional file 3 – Graph showing a comparison of cause of death assessment by PR and most likely cause by InterVA model, for 65 verbal autopsies in persons aged between 5 and 15 years from NUHDSS.

Additional file descriptions text: MS Excel format

Filename: *Graph_5_15years.xls*

Additional files provided with this submission:

Additional file 1: Table_2_cross_tab.doc, 78K

<http://www.pophealthmetrics.com/imedia/1899481448389653/supp1.doc>

Additional file 2: COD_disease_codes_PRvsInterVA.xls, 58K

<http://www.pophealthmetrics.com/imedia/9884371243896545/supp2.xls>

Additional file 3: Graph_5_15years.xls, 43K

<http://www.pophealthmetrics.com/imedia/1416907741389654/supp3.xls>