Assigning HIV/AIDS as a cause of adult death using verbal autopsy: performance of three methods and their effects on estimates of HIV/AIDS-related mortality

A thesis submitted by

Christopher Patrick Grollman

for the degree of Doctor of Philosophy

of the University of London

2014

Completed in the Department of Disease Control with support from the Economic & Social Research Council, the Health Metrics Network and the Wellcome Trust (through the Alpha network)

Statement of own work

I confirm that the work presented in this thesis is my own. Where information has been derived from other sources, I confirm that this has been indicated in the thesis.

I have read and understood the School's definition of plagiarism and cheating given in the Research Degrees Handbook. I declare that this thesis is my own work, and that I have acknowledged all results and quotations from the published or unpublished work of other people.

I have read and understood the School's definition and policy on the use of third parties who have contributed to the preparation of this thesis by providing copy-editing and/or proof-reading. I declare that no changes to the intellectual content or substance of this thesis were made as a result of this help, and that I have fully acknowledged all such contributions.

I have exercised reasonable care to ensure that the work is original and does not to the best of my knowledge break any UK law or infringe any third party's copyright or other intellectual property right.

Signed: …….... Date: .....

Name: CHRISTOPHER PATRICK GROLLMAN

Student identification number: LSH204415

Abstract

Background In sub-Saharan Africa and elsewhere, many people frequently still do not have access to health services that allow medical certification of the cause of their death. Health systems need estimates of cause-specific mortality for planning, and the only way to realistically obtain these in sub-Saharan Africa in the medium term will be through verbal autopsy (VA).

Methods This thesis investigates three methods for assigning HIV/AIDS as a cause of death – physician review, InterVA-4 and the Lopman algorithm – using VA data on 15–59 year-olds from two demographic surveillance systems in Tanzania (Kisesa) and Zimbabwe (Manicaland). The performance of the methods is assessed against the reference standard of known HIV status, allowing the calculation of performance metrics including specificity.

Results The estimated proportion of adult deaths due to HIV/AIDS varied between methods, from 30–53% in Kisesa, and 58–73% in Manicaland. It was not possible to conclude with certainty which estimate was most accurate, nor was there any relationship between the estimated proportions and the performance measured by validity metrics. The methods had variable performance, with physician review having the highest specificity, followed by InterVA-4 and the Lopman algorithm. Findings were broadly consistent with the published literature. Analysis of the Lopman algorithm provided a clear illustration of the problems of using data-derived methods, even where reference-standard data are available to train them.

Conclusion Using validity to assess the quality of real-world VA findings is flawed. Cause-specific mortality estimation should move from seeking single best estimates based on assessment of validity to seeking plausible estimates using synthesis of multiple sources of data – including VA.

Acknowledgements

I have completed this project with support from many people. For their guidance and input regarding the thesis and the wider PhD process I must thank Daniel Chandramohan, Basia Zaba and Carine Ronsmans; Emma Nabavian, Jenny Fleming and James Lewis; and Clara Calvert, Alma Adler, Matthew Darlison and Ian Timæus. Thanks to Robert Thomson and Andy Sloggett for encouraging me in the first place. For practical help in data preparation and in forming my understanding of the origins of the data I thank Denna Michael, Lucas Ng'winamilla, Raphael Isingo, Mark Urassa, Benjamin D Clark, Tom Lutalo, JB Bwanika, Milly Marston, Emma Slaymaker, Simon Gregson and Helen Smith. For help with translations I thank David Obor and David Colmer. Liz Bailey, Clara Calvert, Julia Grollman, Tim Grollman and Andrei Morgan helped with copy-editing.

Part of this research was funded by a studentship from the Economic and Social Research Council. The Health Metrics Network of the World Health Organization gave funding for my participation in two Alpha workshops.

I would not have completed this work without the generous support and love of many people, especially Liz Bailey, Jenny Cresswell, Marina Daniele, DnA, Jess Drader, Karen Frances, Michal Frances, Seb Funk, Nina Gray, Joe Grollman, Julia Grollman, Tim Grollman, Gwen Knight, Barbara Mariotti, Hanaan Marwah, Andrei Morgan, Suchismita Roy, Natti Russell, Rosemary Scott, Cat Towriss and Sam Wilkinson – thank you all.

I have been well supported by the staff in central services within LSHTM, and special mention must go to the refectory staff, not least Kerry Carlarne-Overett and Joe Linanne for the motivation provided by their excellent rock cakes.

~

Medical care is a basic ingredient of social justice and human flourishing; lack of access to medical care is a fundamental reason for the existence of verbal autopsy as a source of cause-of-death data. I dedicate this thesis to all people struggling, in public health and beyond, to ensure that nobody who needs medical care has to go without.

Table of contents

List of tables

List of figures

List of appendices

List of main abbreviations

AIDS	acquired immunodeficiency syndrome
ART	antiretroviral therapy
COD	cause of death
CSMD	cause-specific mortality distribution
CSMF	cause-specific mortality fraction
DSS	demographic surveillance system
HIV	human immunodeficiency virus
ICD-10	International Statistical Classification of Diseases, Tenth Revision
NCD	non-communicable disease
PHMRC	Population Health Metrics Research Consortium
SVR	sample vital registration
VA	verbal autopsy
WHO	World Health Organization

Statement of contribution

To create this thesis, I did:

1. Decide the research objectives;
2. Design the analysis plan;
3. Decide the methods to be used;
4. Translate Kisesa VA data from several formats into a standard data specification;
5. Link and clean household enumeration data, VA data and HIV status data from Kisesa;
6. Conduct all analyses; and
7. Write the thesis.

I worked with others to:

1. Decide on the data sources; and
2. Decide eligibility criteria for deaths occurring in the demographic surveillance systems.

I did not:

1. Collect any of the primary data; or
2. Design the data specifications.

1. Introduction

1. Summary

Few studies have investigated population-based cause-specific mortality distributions disaggregated by HIV status. There is little work comparing the performance of methods of interpreting verbal autopsy (VA) data for diagnosing HIV/AIDS, assessed against known HIV status. This thesis uses population-based data from two demographic surveillance systems to analyse VA findings against known HIV status, compares the performance of three methods of interpreting VA data, and discusses the implications of the choice of methods and of metrics used to assess performance.

2. Importance of knowing the causes of deaths occurring in populations

Knowing what causes of death occur in populations is important for determining the priorities and necessary resources of public health systems, and for assessing the impact of interventions and investments[1-5]. Information on cause of death for general adult populations in sub-Saharan Africa remains severely lacking[2, 6, 7]. Empirical data are important to provide inputs to, and assess the plausibility of, modelling exercises[8]. Vital registration systems, that record numbers of deaths and their causes, are weak-to-non-existent, and where they exist their incomplete and selective coverage hampers the population-level application of the data they provide[9]. Sample vital registration systems, that aim only at coverage of a (usually representative) geographical sample of the national population, have proved useful in India and China[10, 11], and are being introduced in Tanzania and Zambia[12, 13]. Studies that aim to ascertain a representative picture of the cause structure of mortality in a setting are often based in demographic surveillance systems (DSSes) – research programmes that enumerate and investigate deaths and other vital events in all locally resident people[3, 4].

3. HIV/AIDS as a cause of death among adults in sub-Saharan Africa

Human immunodeficiency virus (HIV) causes reduced immune defence against a range of infections, resulting in a set of morbidities collectively known as HIV-related disease[14], the most advanced of which is acquired immunodeficiency syndrome or AIDS. This thesis will use the term "HIV/AIDS" to encompass diseases in which HIV is the underlying cause. References to "AIDS" alone are to the syndrome and not to all HIV/AIDS-related disease. References to "HIV" alone are to the virus.

HIV/AIDS is the leading cause of adult mortality in sub-Saharan Africa[6], responsible for an estimated 1,200,000 deaths in 2012[15]. Community-based studies in eastern and southern Africa, prior to the introduction of antiretroviral therapy (ART), have reported attributable mortality fractions above 40% and up to 80% in some age groups[16-25]. Two morgue-based studies in Republic of Congo showed population attributable fractions (PAFs)* for HIV/AIDS of

* A population attributable fraction is the proportion of mortality in a population that would be avoided if an exposure were removed. Using the prevalence of the exposure in the population and the mortality rate ratio between the exposed and unexposed groups, it is calculated as:

$$PAF = \frac{prevalence\ of\ exposure \times (rate\ ratio - 1)}{1 + prevalence\ of\ exposure \times (rate\ ratio - 1)}$$

24% and 40%[26]. The age-adjusted mortality rate ratio for HIV-positive people in a rural Tanzanian cohort compared to HIV-negative people in 1994–1996 was 14.5 for people aged 15+, and almost 18 for people aged 15–44[18]. A cohort study using HIV status in Malawi found individuals who were HIV-positive at baseline to have a hazard ratio for death at 10-year follow-up fifteen times that of HIV-negative individuals[27]. A cohort study of the general adult population in eastern Zimbabwe found a hazard ratio of 12.1 in the period 1998–2005[25].

Access to ART was very low until 2003[28], since when availability has increased, which has had an important effect on mortality. For example, a study in four largely rural populations in eastern and southern Africa compared mortality in the period 1–2 years following ART rollout with mortality in the pre-ART period: all-cause mortality rate ratios ranged from 0.58 in Karonga, Malawi, to 0.79 in KwaZulu-Natal, South Africa[29]. Post-ART mortality rate ratios among HIV-positive people were 0.35 in Masaka, Uganda, and 0.69 in Kisesa, Tanzania, compared to pre-ART, with no change among HIV-negative people[29]. In Karonga, in 3–4 years following the introduction of ART the mortality rate due to AIDS (as ascertained by verbal autopsy) fell by 68%[30]. In KwaZulu-Natal, where ART provision began in 2004[31], HIV/AIDS-related mortality fell from 15.4 per 1000 person years in 2000 to 10.9 per 1000 person years in 2009[32].

In the period 1–2 years following ART rollout, mortality among HIV-positive people was still an order of magnitude higher than among HIV-negative people in Masaka and Kisesa[29], which may partly reflect incomplete access to or uptake of ART among those eligible. Comprehensive facility-based enquiries into deaths among pregnant/postpartum women in South Africa suggest that mortality is substantially higher among HIV-positive women than HIV-negative women, even among those HIV-positive women who are not eligible for ART due to not yet having the level of immune suppression at which ART treatment is initiated[33].

Monitoring the level of HIV/AIDS-related mortality is an indispensable part of monitoring cause-specific mortality in sub-Saharan Africa[4, 34], and requires identifying deaths in which HIV is the underlying cause[20, 35].

I. Development of definitions of HIV/AIDS over time

In 1985 in Uganda, Serwadda and colleagues reported on cases of "Slim's disease"[36], the same disease as the "AIDS" that had been reported in western populations. Also in 1985, the World Health Organization (WHO) adopted the Bangui definition for case surveillance of AIDS where HIV testing was not available[37], consisting of a list of exclusion criteria, and "important signs", "very frequent signs" and "other signs". Each sign had a score attached and the total score in a patient presenting for clinical diagnosis determined whether AIDS was diagnosed. Those signs formed the basis for the 1994 WHO case surveillance definition, which was a simple algorithmic formulation using the presence of major and minor signs, with an "expanded" definition for use with patients who had a positive test for HIV[38]. These definitions were envisaged as the basis for sentinel surveillance in healthcare facilities[38], though several signs required laboratory support for diagnosis, which may have limited their application[39].

Definitions adopted since by WHO for clinical assessment of the staging of HIV-related disease require a positive HIV test and the presence of one or more clinical conditions[14, 40]. In the International Statistical Classification of Diseases, Tenth Revision (ICD-10)[41], the widely used global standard for classifying diseases, "HIV/AIDS is the underlying cause of death when an HIV-positive individual dies from a co-morbid condition resulting from the HIV infection (codes B20–B24)" (Birnbaum et al 2011: 278[42]). Despite the definitions adopted by WHO, "AIDS" cases are inconsistently defined in surveillance activities in resource-constrained settings[43]. There is no single standard definition of "HIV mortality", "AIDS mortality" or "HIV/AIDS", and terms are used inconsistently and interchangeably in the literature, even within publications (e.g. Groenewald et al 2005[20]).

Even in South Africa, which has the best vital registration system in sub-Saharan Africa, researchers and activists suspect there is substantial under-reporting of HIV/AIDS in official cause-of-death statistics[35, 44], in part due to people's HIV status not being known and their classification as having died due to HIV/AIDS therefore not being possible in health-system-based reporting mechanisms.

II. Definition of HIV/AIDS as a cause of pregnancy-related death

Maternal mortality is the term used to describe those pregnancy-related deaths that are due either to a direct obstetric complication, or to an "indirect maternal" cause. The latter is a

cause that does not derive from the fact of pregnancy but that is aggravated by pregnancy[41]. Pregnancy-related mortality is the term used to describe the deaths of women who are pregnant or die within six weeks of the end of a pregnancy, regardless of the cause of their death.

HIV/AIDS is associated with high pregnancy-related mortality in sub-Saharan Africa: Le Coeur and colleagues found a mortality rate ratio for HIV-positive to HIV-negative women of 3.9 (95% confidence interval 1.7–8.8) in Republic of Congo in 2001[45], and Zaba and colleagues found a mortality rate ratio of 8.2 (5.7–11.8) across six general population cohorts in eastern and southern Africa[46]; large proportions of deaths reported to the South African Confidential Enquiries into Maternal Deaths have been due to infections among HIV-positive women, with a smaller proportion due to AIDS following the criteria in WHO clinical stage 4[33, 47].

There is little evidence that HIV/AIDS is aggravated by pregnancy[48-50]. Despite this, HIV/AIDS is widely treated as an "indirect maternal" cause of mortality among pregnant or postpartum women: a systematic review conducted as part of this thesis[51] found that 12 population-based studies in sub-Saharan Africa reported HIV/AIDS among the causes of pregnancy-related mortality; of these, 10 defined HIV/AIDS as an "indirect maternal" cause, one did not categorise causes, and one categorised HIV/AIDS as a "non-obstetrical cause". Only four of these 12 studies reported criteria defining an HIV/AIDS death, three of which used "clinical signs/symptoms" with no further elaboration. In similar work investigating facility-based studies (unpublished) I found that of 92 facility-based studies, 37 did not mention HIV; 36 categorised HIV as an "indirect maternal" cause of death; 17 made no distinction between causes of death and only two reported HIV-related deaths in a category other than "indirect maternal". One used four broad non-exclusive categories: HIV/AIDS-related, obstetrical, medical and anaemia[52]; the other reported "Obstetric complications" and "Nonobstetric conditions", including HIV/AIDS[53]. There is a strong indication that HIV/AIDS-related deaths may be driving high maternal mortality ratios and obscuring assessment of the success of Safe Motherhood programmes[44, 50, 54]. The concept of "indirect maternal" causes of death has been questioned[49, 55], and authors have variously suggested reporting both ICD-10 codes and pregnancy status for deaths of women of reproductive age[49], or reporting both immediate causes and HIV status for pregnancy-related deaths[51].

4. Ascertaining data on the causes of deaths

There are three main bases for empirical cause-of-death estimates: postmortem autopsy conducted by pathologists (referred to henceforth as simply "autopsy"); clinical diagnosis by physicians; and verbal autopsy (described in detail below). In addition there are modelled estimates, based on empirical data where available[6, 56]. Modelled estimates based on non-disease-related socio-economic, geographic and anthropometric indicators have been made for causes of childhood deaths[57] but not for adults.

I. Autopsies

Of these methods, autopsies are most rigorous and require the greatest level of resources and infrastructure. They are consequently rare in sub-Saharan Africa, largely limited to investigating selected deaths or research series[53, 58]; autopsy investigations in the general population have to date been confined to hospital settings[59]. There have been calls both for greater investment in routine autopsies[5] and for community-based autopsy studies that would provide the best possible reference standard for validating verbal autopsy methods[60]. Efforts to create a "gold standard" facility-based validation dataset using rigorous clinical assessment were driven in part by the belief that a community-based autopsy project would not happen[61]; nonetheless, a pilot is currently taking place under the aegis of the Indepth network[62]. Due to the expense involved, autopsy is not a candidate for large-scale cause-of-death recording.

II. Clinical diagnoses

Facility-based cause-of-death data are also an imperfect source of population-level cause-specific mortality: first, causes of death assigned in health facilities often have poor sensitivity and specificity when compared with autopsy findings[58, 59, 63], and may frequently fail to detect HIV/AIDS as the underlying cause of deaths[20]. Perhaps more importantly than this, many people in sub-Saharan Africa do not die in health facilities, meaning the recording of deaths is incomplete and likely to be biased[9, 10, 64].

This leaves verbal autopsy as the only candidate for deriving population-level cause-specific mortality data in sub-Saharan Africa, at least for the foreseeable future[5, 9, 12, 61, 65].

III. Verbal autopsies

Verbal autopsy (VA) is a public health tool, used for estimating cause-specific mortality. Reviews have periodically presented the limitations that are both inherent in VA, and that are related to individual methods of data-gathering and analysis[7, 65, 66]. The issues were explored in detail with regard to VA for maternal deaths in the report of a WHO expert workshop[67], and the process of developing a VA instrument was described by Zimicki[68].

In brief, VA consists of deriving causes of death based on symptoms suffered by deceased people prior to their deaths; symptoms are reported by someone familiar with the experience of the deceased before death, such as a caregiver or relative. VA is based on the premise that symptoms (and circumstances) prior to death are determined by the cause of death, and that causes of deaths induce more-or-less distinct sets of symptoms in ill people, which can be "recognized, remembered and reported by lay respondents" (Chandramohan et al 1994: 213[7, 69]). The VA process involves several steps: a structured questionnaire to collect reports of symptoms (although unstructured lay reports have also been used[70]); training of VA interviewers; identification of deaths; a visit by the interviewer; selection of a respondent (ideally someone who was present while the deceased was dying); conduct of the interview; and interpretation of the resulting data to derive causes of death.

Each of these steps can introduce some variation in the outcomes of the VA analysis – for example, the frequency with which deaths are ascertained in a population may affect the cause-specific mortality estimates if the cause of death is related to the probability of household dissolution and therefore the ability to record the death or to find an appropriate respondent. Researchers in Malawi found that household dissolution was far more likely for women whose husbands had died and been HIV-positive than for similar women whose deceased husbands were HIV-negative (26% vs 5%, $p<0.001$)[71].

Stigma is associated with HIV-positive status or having died of HIV/AIDS[72], and there are reports in the literature of family pressure on physicians to record cause of death as other than HIV/AIDS when writing death certificates, due both to stigma *per se* and to practical consideration such as access to health insurance[10, 20]. The use of euphemisms has resulted in substantial under-estimation of HIV/AIDS as a cause of death in South Africa[73]. It is plausible that this is mirrored in an unwillingness on the part of respondents to report pre-mortem diagnoses of HIV/AIDS to VA interviewers investigating the deaths of loved ones, particularly if

interviewers lack sensitivity to how traumatic the experience of VA interview, or the perceived revelation of HIV-positive status, may be for the respondent[74]. Closed questions on the presence of symptoms, in addition to or instead of narrative accounts and reports of pre-mortem diagnoses, are important to assigning a cause of death by VA[65].

Authors have suggested that, by contrast, increased knowledge of HIV/AIDS may lead VA respondents to tend toward reporting symptoms associated with HIV/AIDS, leading to overestimation of the proportion of mortality due to HIV/AIDS[75]. However, no empirical data to support this are presented.

VA is widely used for population-level cause-of-death estimation in demographic surveillance systems, including in the DSS networks Indepth and Alpha (Indepth: International Network for the Demographic Enumeration of Populations and Their Health, www.indepth-network.org; Alpha: Analysing Longitudinal Population-based HIV/AIDS data on Africa, www.lshtm.ac.uk/eph/dph/research/alpha). The method is also used to estimate cause-of-death in non-research sample vital registration (SVR) systems in China, India, Tanzania and more recently Zambia[10, 11, 13, 76]. Sample vital registration is a means of obtaining nationally representative cause-of-death data, routinely collected, where vital registration is weak[13]. Data collection methods in DSS and SVR systems are broadly similar, and knowledge gained from DSSes can help the development of further routine sample vital registration[3]. Such systems have a key role in monitoring public health, including progress toward the Millennium Development Goals[77]. An example of how SVR data are used to inform health planning at the district level can be seen from the Adult Morbidity and Mortality Project (AMMP) in Tanzania. That project produced district-level burden-of-disease profiles[†] that were an important input to health planning in the districts where the AMMP operated (Gregory Kabadi, personal communication). Population-based estimates of cause-specific mortality using VA have also been obtained in urban areas through burial surveillance[70, 78].

VA is used to estimate both causes of deaths of individuals and distributions of causes of death in populations. VA is most frequently a "population-level" tool in that the outcome of interest is a distribution of causes of death across a population, but actual population-level estimation – bypassing the individual level – is rare: in VA it has only been done by King and Lu[69], while

[†] Accessible via http://research.ncl.ac.uk/ammp/project_outputs/districtmortality.html, last accessed 25th March 2014.

others have estimated population-level deaths in clusters of causes without using VA[79]. All other methods of interpreting VA data (see next section) aggregate individual-level causes to obtain population-level cause fractions. Individual deaths are either assigned a single cause and summed across the population, or assigned multiple, fractional causes with the fractions of each individual death assigned to each cause summed to give the population-level fraction for that cause[68, 80].

At the individual level VA is not, to date, used to suggest to respondents the causes of death of their individual deceased loved ones – indeed, the use of InterVA-4 for investigating individual deaths is explicitly discouraged in its user guide[81]. VA has been used to assess the causes of individual deaths in clinical trials, but even in such applications the results are aggregated across groups of deceased people. Acknowledging concerns about the potential for individual-level misclassification to affect trial outcomes, authors using VA to assign causes to deaths of adult women in a randomised controlled trial have observed that misclassification is likely to affect both intervention and control arms equally[82].

Diversity in the methods used has long been recognised as a potential limitation to comparability of VA findings[7, 57, 67, 83], and there have been many calls for standardisation of VA procedures[12, 57, 65, 84]. WHO released a set of standardised tools in 2007[85], but these were not universally adopted. In response to concerns over the impractical length of its 2007 standard questionnaire, WHO has undertaken an extensive review of VA methods in current use, and developed a shorter questionnaire, the authors of which have called on VA practitioners to adopt as a matter of urgency[84]. Authors have investigated the effects of using different questionnaires, by comparing results using the full range of symptoms available in a dataset with the results using only those symptoms that would have been available using another questionnaire. They have found the effect to be minor[86] or moderate[87].

The method used to interpret VA data and assign causes of death ("the interpretation method") is a source of substantial variation in the findings of VA studies[88-95], and it is the variation in methods that is perhaps further from resolution than the issue of which questionnaire to use.

Many VA studies have assessed HIV/AIDS as a cause of adult or all-age mortality in sub-Saharan Africa[34, 64, 88, 89, 96-102]. VA has been used to measure trends in mortality within a single DSS[32], though the WHO has urged caution over using VA for monitoring trends[83]. Definitions

of which deaths are due to HIV/AIDS, and therefore estimates of HIV/AIDS-related mortality, depend on the interpretation method and thus are affected by variability between those methods.

5. Methods of interpreting VA data

Interpretation methods developed to date fall into two broad categories: expert-opinion-based methods and data-derived methods. Expert-opinion-based methods can be further split into review of VA data by a physician or other reviewer ("physician review") and automated methods using pre-defined algorithms[103] (Table 1). The most important practical difference between expert-opinion-based methods and data-derived methods is that data-derived methods must be trained using a "reference-standard" cause of death – a "true" cause known for each death – in addition to the VA data, while expert-opinion-based methods can be used with VA data alone. Few studies combine interpretation methods[18], and they do not always explain how methods are combined [104, 105].

Expert-opinion-based methods	**Data-derived methods**
Applied using VA data only	*Require VA data & reference standard*
Physician review	Lopman algorithm
	King-Lu direct estimation
Pre-defined algorithms	Simplified symptom pattern*
Simple expert algorithms, e.g. those of WHO (1994[38]) or Lulu and Berhane (2005[99])	Random forests*
	Tariff*
InterVA	Neural networks
	Probability density
	Logistic regression

Table 1: Methods of interpreting VA data. *IHME methods

I. Expert-opinion-based methods

i. Physician review

Physician review can be seen as the classic method of interpreting VA data, used in the vast majority of early studies[7] and still widely used today. The method consists of physicians assigning causes of death for VA transcripts based on their reading of the reported symptoms and free-text narrative account, sometimes supplemented with additional information such as known HIV status[30]. It is relatively expensive and time-consuming, and being based on human cognition, lacks reliability and reproducibility[106-108]. There have been several cheaper, quicker and more reliable automated methods proposed[65], and in 2011 many VA researchers agreed on the need to convincingly move away from treating physician review as the default interpretative method[109], although more recent statements have been more equivocal[110, 111]. While some automated interpretation methods are more widely used than others, there is no consensus on which is best, and innovation continues – indeed, there is an argument that no single method can satisfy all of the uses of cause-of-death data[65, 83, 112, 113]. With this uncertainty about which method should "replace" physician review, there is some resistance to moving away from physician review[19].

ii. Simple expert algorithms

Among expert-opinion-based methods other than physician review, the 1994 WHO algorithm[38] has been applied several times for cause-of-death estimation – in unaltered[75], simplified[89] and modified forms[45]. Algorithms developed by Lulu and Berhane[99] and by Quigley and

colleagues[103] have been applied by other researchers investigating mortality among adults, though the way in which they have been applied has sometimes been unclear[105].

iii. InterVA

The most frequently used non-physician review method is the Bayesian probabilistic algorithm InterVA, developed by Peter Byass and colleagues[114, 115]. Its current iteration is InterVA-4[102], available in the public domain via *www.interva.net*. InterVA has been applied frequently by authors involved in its development[64, 88, 116, 117] and others[32, 118, 119]. InterVA-4 has also been made compatible with the shortened WHO standard VA questionnaire[84].

II. Data-derived methods

Most non-physician review interpretation methods have had limited, if any, use by VA practitioners investigating adult mortality. Of the data-derived methods listed in Table 1, most have only been applied on a small set of research datasets. Methods using neural networks, probability density and logistic regression have been presented in validation studies alongside physician review and expert algorithms[103, 120]. The Lopman algorithm has been presented in development and validation studies only[22, 87]. The direct estimation method of King and Lu[69] has been applied for child deaths but not for adult deaths[91].

There has been a resurgence of interest in VA in recent years[109], in which a major contribution has been the work of researchers at the Institute for Health Metrics and Evaluation (IHME) led by Chris Murray, who have developed several new methods[90, 92, 95] as well as creating a "gold standard" dataset for validation (discussed below). Unfortunately, these methods are proprietary and not publicly available for use outside the team that developed them, who presented them in the August 2011 series on verbal autopsy in Population Health Metrics (www.pophealthmetrics.com/content/9/August/2011) and in a recent paper comparing computerised methods[121]. Data-derived methods have a barrier to use due to requiring a reference standard of "true" cause-of-death for their application; in practice, facility-based medical records have been used to train many data-derived methods[61, 103], which is the same reference standard that is frequently used to validate VA output.

III. The differences between interpretation methods: validity, accuracy and feasibility

The differences between interpretation methods lie in the validity and accuracy of their results and the feasibility of their use. Validity and accuracy may be different when methods estimate individual-level causes and aggregate to obtain population-level cause-specific mortality fractions (CSMFs): even if methods assign incorrect causes to individual deaths (imperfect validity), these may cancel out in population-level estimates, not affecting accuracy[65]. Cancelling out may not be purely fortuitous, as causes of death with similar presentations are most likely to be mistaken for one another, and errors in population-level CSMFs have been lower than errors in individual-level classification[122]. Where it has been considered, authors have tended to attempt to correct for misclassification[123, 124] or emphasise the importance of minimum levels of validity[66, 98, 125, 126] rather than assume misclassification to be bi-directional[75]. In practice, measuring the accuracy of cause-specific mortality estimates requires an assessment of validity, which I will consider in more detail in the next section.

Assessing validity for individual deaths is conceptually easy where each death is assigned a single cause that can be compared to a reference standard, but is more difficult when individual deaths are assigned multiple, fractional causes, as in applying InterVA. Authors validating CSMDs assigned by InterVA have taken several approaches: some validated at the level of individual death, treating as correct deaths where any of the most likely three fractional causes matched the reference standard[116]; some compared the summed fractional probability in the population with the single-cause-per-death CSMD in the reference standard (for calculating CSMF accuracy)[93]; sometimes the approach taken is unclear[101, 102].

The requirements for using different interpretation methods vary, making their use in a given setting more or less feasible. Physician review requires available physicians to review the VA transcripts and assign causes of death. Places with the greatest need for VA tend to be those where human resources for health are most scarce.

The IHME authors have described the difficulty of using Simplified Symptom Pattern (SSP) in the field, due both to its needs for a large training dataset and great computational power[90]; they have recommended Tariff, partly due to its parsimony and ease of implementation[95, 121]. The fundamental limitation to using data-derived methods to estimate causes of death in populations is that they require training on reference-standard data, which is often unavailable.

6. Validating verbal autopsy findings

Many studies have sought to validate the findings of VA interpretation methods, including their estimates of HIV/AIDS-related mortality[94, 96, 100, 103]. It is important to recognise that there is no ultimate way of determining true cause of death: the cause of death is a judgement arrived at with more or less rigour, and validation exercises seek to assess the findings of one VA method against the findings of a method of greater rigour. Validating VA findings requires three components: metrics to measure validity; data in which to make the validation; and a reference standard against which that output can be compared.

I. Validation: metrics for measurement

Validation metrics are attributes of an interpretation method applied to a dataset. The metrics used most frequently to judge the validity of methods of interpreting VA data have been sensitivity and specificity, as well as positive predictive value (PPV) and the kappa statistic (used to measure inter-rater agreement, adjusting for chance agreement)[84]. Several authors have suggested or endorsed levels of acceptable sensitivity and specificity[66, 98, 110, 126], though these are all arbitrary. When estimating multiple causes of death, high specificity is more important than high sensitivity in minimising misclassification between causes[124, 127]. When validating mortality due to HIV/AIDS using HIV status, sensitivity is a less meaningful metric than specificity as HIV-positive people can die of non-HIV/AIDS-related causes of death[96].

Studies have frequently reported more than one metric of performance[110], as different metrics are useful for answering different questions[65]. To illustrate, in 1990 Kalter and colleagues[113] wrote of their intention

> to determine a number of diagnostic algorithms, each of which might be appropriate for a particular purpose. For example, high sensitivity might be desired in a clinical situation where one planned to treat all suspect cases of an illness, whereas high specificity would be important to minimize misclassification, if overtreatment was a concern. (p381)

It has long been recognised that sensitivity, specificity and other standard validation metrics are vulnerable to the true cause-specific mortality distribution (CSMD) in the dataset[124, 127-129]: different CSMDs mean different patterns of misclassification, and therefore the performance of an interpretation method in terms of sensitivity and specificity will vary depending on the

dataset. This severely limits the meaning of comparisons of these metrics for VA interpretation methods applied to data from different places or points in time[129]. Validated causes of death tend to come from health facilities, which have a different cause-composition of deaths compared to the general population. This vulnerability to the true CSMD in the dataset means that estimates of validity derived from health facilities cannot be used to adjust for misclassification in community-based VA findings[123].

Alternative validation methods have been proposed by the Population Health Metrics Research Consortium (PHMRC): chance-corrected concordance (CCC) for assigning causes to individual deaths, and CSMF accuracy for population-level cause-specific mortality fractions (CSMFs)[129]. These are not like-for-like replacements for existing metrics – rather, they are proposed to fulfil articulated requirements of VA validation metrics: first, CCC is proposed to measure "the fraction of true deaths from a cause that are correctly assigned to that cause" (Murray et al 2011: 4–5[129]). CCC is better than sensitivity for this purpose as it accounts for the likelihood of deaths being assigned to the same cause by chance; for datasets with large numbers of possible causes of death, the CCC of a method approximates to the sensitivity of that method. In contrast, when there are few possible causes, CCC is much lower than sensitivity, reflecting the greater likelihood of chance agreement between the assigned and true cause[129]. CCC can be calculated for each cause of death in a dataset, and the authors propose that the CCC of a method across all causes be calculated as the average (mean) of the cause-specific CCC values[129]. Murray and colleagues also propose partial CCC (PCCC), for use when methods output several causes of death of ranked likelihood, rather than a single cause of death.

CSMF accuracy is proposed to measure the overall performance of a method for estimating CSMFs, across all causes. This assesses the degree to which, across all causes, the false positive and the false negative diagnoses of each cause balance out and return an accurate CSMF. It is scaled from zero to one, with zero being the maximum possible CSMF error in that dataset and one being perfect estimation of CSMFs. The absolute CSMF accuracy is important in the application of these methods to public health, as necessary public health resources are proportional to the absolute size of a public health problem, and "the negative consequence [of inaccurately estimating CSMFs] scales to the absolute error in cause estimation, not the relative error" (Murray et al 2011: 7[129]).

CCC and CSMF accuracy share with sensitivity, specificity, kappa and other statistics the limitation of vulnerability to the true CSMF composition of the dataset. Murray and colleagues[129] wrote, regarding CCC, that

> To avoid making the wrong inference on a method's performance, we recommend that a set of 100 or more test datasets be created with varying CSMF compositions (p6)

and regarding CSMF accuracy, that

> reporting CSMF error or CSMF accuracy for one test set would risk drawing an incorrect inference on relative performance (p8)

Murray and colleagues[121] calculated CCC, CSMF accuracy, sensitivity, specificity and kappa for several methods applied to the 500 test datasets derived from the PHMRC "gold standard" dataset, and reported the median values assigned by each interpretation method.

Desai and colleagues[111] calculated PPV, PCCC and CSMF accuracy for a number of computer-coded VA interpretation methods applied to several large datasets of VA records from various low- and middle-income settings. They did repeated resampling from the original dataset in order to obtain 30 train/test splits in which to apply the methods[111], but it is not clear whether the reported results are the median results from these repeated resamplings (as recommended by Murray and colleagues[129]), or whether they used another method.

These innovative validation metrics do fulfil two needs in judging the performance of VA interpretation methods, and are important methodological developments that clarify the aims of validating VA findings[130]. However, they do not solve the problem of validation metrics being vulnerable to underlying CSMDs and the resulting non-comparability of diverse estimates from single datasets.

II. Validation: data in which to validate

Validation studies have tended to assess the performance of one or more VA interpretation methods using the CSMD of a single, moderately sized set of facility-based validation diagnoses[98, 101, 103, 121]. A recent systematic review found only five studies from sub-Saharan Africa reporting adult or all-age mortality from all causes in datasets with ≥1000 deaths, plus one investigating a single cause in a dataset containing ≥100 deaths[110]; the exclusion criteria

for that review meant that at least three studies investigating HIV/AIDS-related mortality with ≥100 deaths were excluded for other reasons[22, 87, 126].

Murray and colleagues[129], presenting their new set of validation metrics, state that "it could be extremely misleading to draw conclusions about the performance of one method compared to another on the basis of only one test dataset" (p4); and

> In any real comparison of alternative VA methods with longer cause lists [...] drawing conclusions about which method performs better cannot be made based on one test dataset but needs to be carefully assessed for a diverse range of cause compositions in a series of test datasets. (p4)

This means that, ideally, performance would be assessed by applying methods to datasets that have diverse CSMFs and are obtained using comparable methods. Until recently there was no VA dataset large enough to allow robust comparisons between the precision of different methods[120, 131]. In recent years, large datasets have been created. The PHMRC gathered a dataset of 12,500 facility-based deaths from a total of six sites in India, Tanzania, Mexico and the Philippines using a standardised procedure, for the purpose of validating VA interpretation methods. Reference-standard causes of death were assigned using stringent diagnostic criteria including medical imaging, laboratory tests and pathology findings[61]. From this large dataset, the PHMRC created 500 pairs of training and testing datasets with randomly varying CSMF compositions, to represent a wide range of mortality scenarios.

While the PHMRC validation dataset contains identified errors and inconsistencies[86], it is likely to be of higher quality than medical records derived from facilities used in any other VA validation exercise. However, it has one important limitation for ascertaining mortality due to HIV/AIDS. The authors describe "AIDS" and "AIDS with TB" as two of the causes of death in their mortality classification[129]. The definition they use to obtain their clinical gold standard HIV/AIDS deaths is equivalent to the definition of stage 4 HIV disease (AIDS) in the WHO clinical staging[14], which is the strictest definition available. Almost all of the "AIDS" deaths in their validation dataset (493/501) meet this criterion. This is problematic because cause-of-death studies ought to ascertain the underlying causes of deaths[20], but not all deaths of which HIV is the underlying cause and which might usefully be ascribed to "HIV/AIDS" meet the clinical criteria for stage 4 disease[33, 132, 133]. These criteria may mean that classification of deaths in the reference standard is more restrictive than it should be, and that deaths in which

HIV infection is the underlying cause have been ascribed to other causes. Such a limitation may also explain the apparent over-estimation of HIV/AIDS-related mortality by InterVA 3.2 in this dataset[93]. This limitation of the PMHRC validation dataset has also been noted by Peter Byass[86].

Desai and colleagues[111] claim to have assessed the performance of VA interpretation methods using the largest dataset so far created for this purpose, containing over 24,000 deaths from a range of settings in low- and middle-income countries[111]. However, those authors did not create a pooled dataset across all these settings, presumably because the data were collected using heterogeneous methods, so the resulting assessments of performance were conducted separately on each dataset. The largest of the datasets they used, a subset of the Million Deaths Study in India, contained over 12,000 deaths. Their overall dataset is not presented as a validation dataset, as it does not contain data on the true causes of deaths. Nonetheless, as noted, Desai and colleagues[111] did analyse it using the validation metrics proposed by Murray and colleagues[129].

For investigating mortality due to HIV/AIDS, a dataset comprising over 17,500 deaths, with HIV status from research testing known for 29% (5000) of the deceased, has been collected from six demographic surveillance systems in the Alpha network[102], potentially providing an important validation resource.

The preference for large datasets does not mean that one cannot meaningfully measure the validity of findings in smaller datasets where only one CSMD is feasible, but rather that one cannot use smaller datasets to conclusively or rigorously compare the performance of interpretation methods. Instead, validation in smaller datasets serves to inform the interpretation of the CSMD findings produced by applying individual VA interpretation methods.

i. Issues with using facility-based data to validate VA output

Hospital-based cause-of-death data are clearly important to validate VA findings: Murray and colleagues state “We simply have no way to figure out the true cause of death for deaths that have occurred in the community with no contact with health services” (2011: 13[61]); and Chandramohan and colleagues (1998[134]) wrote that “the causes of community-based deaths are only obtainable through post-mortems for each community-based death, therefore the validity of verbal autopsy is only studied in hospital deaths”.

The main problem with facility-based data is that the CSMD in facilities differs from that in the community, making it impossible to adjust misclassification in community-based estimates using validation metrics derived in facilities[123]. Aleksandrowicz and colleagues[122] demonstrate that this is a real concern: in data from the Million Deaths Study in India, the CSMF composition of deaths in hospitals differed dramatically from that of deaths at home, including odds ratios of 0.6 for tuberculosis (95% CI 0.5–0.6), 0.7 for chronic respiratory disease (0.6–0.7) and 3.4 for maternal conditions (2.9–4.0). There is no reason to think those findings would not be similar in sub-Saharan Africa. By contrast, one group of authors have tried to adjust cause-specific mortality fractions for malaria using estimates of mean sensitivity drawn from the literature, and an assumed relationship between specificity and the observed proportion of malaria deaths in the unadjusted VA analysis. The authors regard their adjusted estimates as plausible, but present no sensitivity analysis investigating the potential effect of basing the adjustment on different assumptions, so conclusions cannot be drawn. Their adjusted estimates have very wide confidence intervals, particularly at higher observed unadjusted proportions, so it is in any case unclear how useful the adjusted estimates are[135].

III. Validation: Reference standards

Researchers have sometimes used the findings of physician review to validate proposed automated methods, either stating that these are probably accurate[103] or acknowledging the inherent limitation in using physician review as a reference standard[136]. Others have used physician-review findings to validate the findings of lay reports of cause of death[70] – or to assess the performance of automated methods in the absence of a true cause-of-death needed for formal validation[111]. Physician review has imperfect reliability and repeatability[65, 137]; it may also suffer from biases introduced by the regional and clinical specialisms of the reviewing physicians[100], but on the other hand may benefit from knowledge of the local context[138]. It is not ideal as a reference standard – particularly as newer methods are being developed explicitly to enable a move away from physician review for interpretation of VA data[109].

Methods of interpreting VA data are frequently validated against medical records for people who have died, or at least spent time, in health facilities[34, 61, 98, 100, 101, 113, 125]. Authors sometimes attempt to adjust for[100] or ensure[61, 101] the quality of these records. Medical records are not an ideal reference standard either, as physicians frequently make incorrect diagnoses[59, 63, 83, 106]. In addition to the inherent risk of incorrect diagnosis, medical records

may be biased away from recording HIV/AIDS as a cause of death, due to stigma in such a diagnosis[20, 34, 42]. Nevertheless, as Daniel Chandramohan put it in 2011, "hospital diagnosis of COD [cause-of-death] based on defined clinical and laboratory criteria are the only useful gold standard available at present for validating VAs"[130].

No assessment of the validity of VA against autopsy findings has been reported to date; it is worth noting that even pathologists have imperfect reliability and repeatability[106].

For a limited number of causes of death, some studies have used more specific indicators of true disease, giving greater confidence in the accuracy of the reference-standard diagnoses. For example, a study in India that did not validate its general VA findings due to the absence of a reference standard was able to validate diagnoses of cancer through record linkage with a cancer registry[11]. A similar means of having greater confidence in the validity of VA findings, particularly in their specificity, is the use of known HIV status to validate findings of HIV/AIDS-related mortality. I will now consider this in more detail.

i. HIV status as a reference standard

Several studies have used known HIV status as the sole external standard to validate VA findings of HIV/AIDS (Table 2)[19, 87, 89, 96, 102]. Three studies[19, 89, 96] used HIV status alone as the reference standard, taking deaths of HIV-negative people to constitute the denominator for specificity, and deaths of HIV-positive and HIV-negative people diagnosed with "HIV/AIDS" to calculate the positive predictive value. Todd and colleagues also calculated sensitivity, while Kamali and colleagues stated "Since a proportion of HIV-positive subjects are bound to die of non HIV-related causes we considered the sensitivity of the tool less meaningful" (p682). Mayanja and colleagues[19] did not seek to ascertain deaths due to HIV/AIDS in their analysis of VA data – rather, their aim was to assess how well physician review could ascertain HIV-positive status among deceased people: "Since we did not have data on the actual cause of death, we defined as 'HIV-associated' any death occurring in a person who was HIV-seropositive" (p3). Byass and colleagues[102] calculated only specificity, combining HIV-status data with VA data: they excluded from the denominator for specificity those deaths of recently HIV-negative people where a premortem diagnosis of HIV/AIDS was reported in the VA. Finally, Lopman and colleagues[87] calculated sensitivity and specificity. They treated HIV-positive people with VA evidence of obstetric causes of death or injuries as not having died

from HIV/AIDS, excluding them from the denominator for calculating sensitivity and including them in the denominator for calculating specificity.

Clearly, deaths of HIV-negative people should always be considered true negative cases. Regarding the definition of positive cases, even studies that only calculate specificity make an implicit judgement about the cause of deaths of HIV-positive people: to exclude them from the denominator of the specificity calculation is to say that none of them ought to be considered true-negative HIV-related deaths. A reference standard consisting only of HIV status, as used by all the above studies apart from Lopman *et al.*, with no information on the cause of death for HIV-positive people, means VA is discriminating solely between deaths of HIV-negative and HIV-positive people. Positive cases are deaths of HIV-positive people rather than necessarily people who died of HIV/AIDS – as Mayanja and colleagues explicitly recognise. These might be understood as deaths "with HIV", rather than necessarily "from HIV".

Study	Metrics used	Definition of true negative cases	Definition of true positive cases
Kamali et al 1996[96]	Specificity PPV	HIV-negative status	HIV-positive status
Todd et al 1997[89]	Specificity Sensitivity PPV	HIV-negative status	HIV-positive status
Mayanja et al 2011*[19]	Specificity PPV	HIV-negative status or HIV status unknown	HIV-positive status
Byass et al 2013[102]	Specificity	HIV-negative status with no VA report of pre-mortem diagnosis of HIV/AIDS	–
Lopman et al 2010[87]	Specificity Sensitivity	HIV-negative status, or HIV-positive with VA evidence of obstetric cause of death or injuries	HIV-positive with no VA evidence of obstetric cause of death or injuries

Table 2: Metrics used to validate VA diagnoses of HIV/AIDS, in studies using HIV status for validation. PPV=positive predictive value. *This study was not a cause-of-death study, but aimed to estimate HIV prevalence among the dead using physician review of VA.

A reference standard such as that used by Lopman and colleagues attempts to discriminate among deaths of HIV-positive people, by identifying those that were unlikely to be due to underlying HIV infection. Using no information beyond the VA record and HIV-status data, this means finding indicators in the VA data that "override" a diagnosis of HIV/AIDS where such would otherwise be made. In practice, it may be difficult to determine which deaths of HIV-positive people this should include: in order to avoid misclassifying HIV/AIDS-related deaths as false-negatives, the VA indicators used should be highly specific, as well as being likely to be remembered and reported in a VA interview. Lopman and colleagues[87] use indicators of injuries, and obstetric indicators comprising death during childbirth, or death shortly before delivery with either excessive bleeding or severe headaches. It is not clear what this categorisation would mean for the level of specificity, compared to treating deaths of HIV-negative people only as true negatives: if specificity among the "injury/obstetric" deaths of HIV-positive people were greater than among all HIV-negative people, which seems possible, this categorisation would raise specificity – although modestly if the total number of "injury/obstetric" HIV-positive deaths were small. On the other hand, if HIV/AIDS-related comorbidities were reported among the injury/obstetric HIV-positive deaths, some might be diagnosed with HIV/AIDS, and specificity may be little affected.

However, it is not clear that even deaths due to injuries or obstetric causes should be considered non-HIV/AIDS-related for HIV-positive people: there is some evidence that even these causes are more likely for HIV-positive than HIV-negative people[102], although for obstetric causes among HIV-positive women there is stronger evidence of no excess risk among HIV-positive compared to HIV-negative women[48].

There are two reasons not to use a stricter, more specific definition of positive cases, akin to the strict definition of AIDS deaths used in the PHMRC validation dataset, when defining a reference standard based on VA data. First, the level of detail typically available in VA records is probably inadequate to discriminate between more moderate and more advanced HIV-related disease (e.g. stage 2 vs. stage 3/4). This is compounded if questions on the duration of symptoms (crucial to diagnose HIV wasting syndrome) are removed from VA instruments, as occurred with the instrument recently developed by the PHMRC[61, 84]. The second reason is related: evidence of higher mortality from most causes among HIV-positive compared to HIV-negative people[102] blurs the line between HIV/AIDS and other causes of death, which undermines confidence in which deaths are truly due to HIV/AIDS. The three possible approaches discussed here are summarised in Table 3.

With increasing availability of antiretroviral therapy, both levels and cause-distributions of mortality will change[29, 30, 32]. It may be more difficult to determine when deaths are attributable to HIV/AIDS if common presentations change[139] and HIV/AIDS-related disease occurs more frequently as a co-morbid, contributory condition[30]. Such changes potentially further complicate validation of VA findings.

In general, using HIV status as a reference standard should benefit data-derived methods over expert-opinion-based methods: data-derived methods will be trained using HIV status as a proxy for cause of death, while expert-opinion-based methods will have been designed to ascertain HIV/AIDS as the cause of death.

Approach to defining HIV/AIDS-related reference standard	HIV status of deceased	HIV/AIDS-related deaths in reference standard		"True positive" deaths "with HIV" or "from HIV"
		"True positive"	"True negative"	
"HIV-status-only" standard	HIV+	All	None	"with HIV"
	HIV–	None	All	
Lopman-style "injury/obstetric" standard	HIV+	With obstetric/ injury symptoms	No obstetric/ injury symptoms	Close to "with HIV"
	HIV–	None	All	
Strict PMHRC-style "AIDS-only" standard	HIV+	With AIDS-defining conditions	Without AIDS-defining conditions	"from HIV"
	HIV–	None	All	

Table 3: Impact of definition of HIV/AIDS-related deaths in a reference standard

IV. Validation: summary and implications

If there were valid measures for the causes of all deaths, there would be no need to conduct VA studies: validation serves to help assess the performance of VA methods (as in most applications to date), or can be used on, for example, a sample of deaths recorded in a vital registration system, as a means of quality control and to help assess accuracy of the findings. Authors have stated the importance of validation studies when VA is used in new contexts, particularly non-research settings[10, 12, 98].

Where validation is possible, it seems advisable to report a range of metrics including the PHMRC metrics (CCC and CSMF accuracy) as well as standard metrics – the exact choice will depend on the planned use of the data.

Given that innovation in VA methods is likely to continue, and that there are several credible methods in the field at present, it seems unlikely that validation in the large datasets currently available[111, 121] will lead to a settled consensus on which method is best, its universal adoption and the abandonment of all others. It is also not clear that settlement on a single interpretation method, based on median performance across multiple datasets, is desirable. For example, where HIV/AIDS-related mortality is known *prima facie* to cause a substantial proportion of adult deaths, it may be unwise to choose an interpretation method based on median performance across a wide range of CSMDs in most of which a small proportion of deaths are due to HIV/AIDS.

Large validation datasets will be able to provide valuable information about the performance of interpretation methods, and may be able to determine a "best-performing" method[130]. However, even in the event of consensus as to which method performed best across multiple CSMDs, that method would nonetheless have a range of performance. There may be fundamental problems with using facility-based data for validation, based on differences in symptom–cause relationships between facility-based and community-based deaths[111]. In practice, when applying VA to determine population CSMDs, validation will often not be possible. In such cases it would be impossible to know where in its range even the best method was performing. An alternative approach might be to apply a range of interpretation methods and combine their results, considering what is known about their respective performances.

Different reference standards imply different definitions of what constitutes a death due or not due to HIV/AIDS. The strictest standards used to date, reflecting the WHO clinical staging criteria for advanced HIV-related disease or AIDS, require high-quality facility-based records. In community-based studies, reference data derived from health facilities are unavailable, as few people receive facility-based care[64]. For investigating HIV/AIDS-related deaths, a simple reference standard only using HIV status may overestimate the number of deaths in which HIV/AIDS is the underlying cause, treating as "true positives" deaths in which HIV was contributory or incidental as well as those in which it was causative. Using VA data to classify as "true negatives" deaths in which HIV was incidental may come closer to estimating the true HIV/AIDS-related CSMF, but requires subjective assessment of which indicators ought to classify HIV-positive deaths as true negatives. In either case, a reference standard where HIV status is treated as a proxy for HIV being the underlying cause of death will benefit data-derived over expert-opinion-based interpretation methods. Nonetheless, true HIV status provides a degree of objectivity that other reference standards do not, particularly for assessing false-positive cases and specificity.

7. Paucity of estimates, poverty of definition

A major past investigation found that a large majority of epidemiological studies reporting cause-specific mortality data (80%) reported either child or maternal mortality, and that deaths of non-pregnant adults were relatively neglected (Adetunji et al 1996, cited in Rao et al 2006[9]). International estimates have been made without empirical inputs on general adult

mortality in sub-Saharan Africa[140]. The number of studies reporting cause-specific adult mortality in representative general-population samples or surveillance systems has increased more recently[13, 32, 34, 88, 101, 116, 136, 141-144]. These studies have overwhelmingly used verbal autopsy, with the proportions of deaths due to HIV/AIDS and other causes decided by the interpretation method, in all cases physician review and/or InterVA. Where reference standards have been used, these have been medical records[101, 116], with no description of criteria defining a death as being due to HIV/AIDS.

Understanding the role of HIV/AIDS-related disease in adult cause-of-death, and understanding the validity and accuracy of existing estimates, is limited by the severe lack of population-based data on cause-specific adult mortality by HIV status. Okongo and colleagues[132], in Masaka, found that among 57 deaths of HIV-positive people aged 13+ with a cause assigned, 88% (50/57) had AIDS and only 5% had non-HIV-related causes of death. Byass and colleagues[102] reported significantly higher mortality rates among HIV-positive than HIV-negative people across many causes of death, even those seemingly unrelated to HIV infection, with a rate ratio of 29 (95% confidence interval 27–31) for all-cause mortality. Mayanja and colleagues[19] asked physicians to use VA data to assess which deceased people had been HIV-positive prior to death which, while not a cause-of-death study, was validated using known HIV status.

Similarly, many estimates of cause-specific pregnancy-related mortality come from health facilities[53, 145-147], but the resulting CSMDs cannot be said to represent the population CSMD due to cause selection leading to bias for general adult mortality. As noted above, there is uncertainty over how HIV/AIDS should be treated as a cause of pregnancy-related death, and nothing is known about how HIV/AIDS is treated in practice in studies reporting cause-specific pregnancy-related mortality.

~

This thesis investigates cause-specific adult mortality in two locations in eastern and southern Africa, assigned using verbal autopsy (VA) data, and its variation by HIV status. I focus on HIV/AIDS-related mortality and differences in the assignment of HIV as a cause of death using three methods of interpreting VA data.

This thesis will investigate physician review, InterVA and the Lopman algorithm. I have selected these for the following reasons: physician review data are collected within the DSSes

as part of routine data collection, and should be included as they are a common default interpretation method; InterVA is publicly available and can be implemented using VA data only; and although it is a data-derived algorithm, the Lopman algorithm was designed to require only HIV status in addition to VA data, rather than data on true cause of death. These methods are described in more detail in their respective chapters. I do not have access to Tariff, SSP or Random Forests due to their proprietary nature. The King-Lu method is publicly available, but is designed for multiple-cause estimation and requires true causes of deaths, which I do not have.

The reference standard of known HIV status means that specificity is defined as the proportion of deaths of HIV-negative people assigned to non-HIV/AIDS-related causes, and other definitions are clearly stated in the thesis. Deaths are treated as having single causes for physician review and the Lopman algorithm, and multiple fractional causes for InterVA. Accuracy (as discussed above in section 5-III) is only used in presentation of the results of the Lopman algorithm, and is measured by the difference, in percentage points, between the estimated proportion of deaths due to HIV/AIDS and the proportion in the reference standard.

The data for this thesis come from routine data collection in two DSSes in the Alpha network: Kisesa, in Tanzania, and Manicaland in Zimbabwe. The Alpha network includes a further eight sites in Southern and Eastern Africa, each with its own research focuses. Participation in the Alpha network aims to achieve comparable analyses of demographic and epidemiological data, including VA data and data on HIV status. To this end, Alpha organises workshops around specific topics of interest,[‡] including one on the topic of verbal autopsy in Kisumu, Kenya, in October 2011. Data prepared for that workshop are central to this thesis. The analyses will use known HIV status to investigate cause-specific mortality by HIV status and allow partial validation of the findings.

The next chapter, General Methods, consists of two parts: Data Sources and Study Population, which describes the settings in Kisesa and Manicaland from which the data are drawn, the collection of the data within the DSSes, and the study populations; and Data Management and Analyses, which describes the definitions used to facilitate the analyses and the receipt and cleaning of the data, and outlines the analyses conducted. Three chapters then present the

[‡] See Alpha network website: http://www.lshtm.ac.uk/eph/dph/research/alpha/workshops/index.html

results of the substantive analysis for each of the three methods of interpreting VA data (Physician Review, InterVA and Lopman Algorithm). Chapter six, Assessment of Potential Bias, describes the potential presence and magnitude of selection bias in the dataset overall and for the analyses specific to methods of interpreting VA data. It goes on to discuss information biases the risk of which cannot easily be assessed. The thesis ends with a Discussion which draws together the findings of the results chapters and places them in the context of what is already known about using VA to estimate HIV/AIDS-related mortality.

8. Objectives

I. Overall objective

To assess the performance of methods of interpreting VA data for assigning HIV/AIDS as a cause of adult death.

II. Specific objectives

1. To describe the proportion of deaths assigned to HIV/AIDS by the interpretation methods.
2. To use known HIV status to:
    a. Further describe cause-specific mortality by HIV status;
    b. Partially assess the validity of the findings of the interpretation methods.
3. To draw conclusions about the role of VA in estimating the proportion of mortality due to HIV/AIDS.

2. General methods

A. Data sources and study population

This chapter describes the populations that provided the data used in this thesis, and the methods of collecting those data. Limitations of the data and the methods of collecting the data are discussed in the Discussion chapter.

1. Data sources and study population: Kisesa

I. Study setting and population

Tanzania is a country in eastern Africa. According to a census, the population in 2012 was 44,928,923, growing at 2.7% per year between 2002 and 2012[148] (Table 4). Three quarters of people live in rural areas[149]. The crude death rate in Tanzania was estimated to be 10.8 per 1000 across all ages for the period 2005–10[150]. The total fertility rate (TFR) among 15-49 year-old women nationally was 5.4 in 2010, and much higher in rural settings (6.1) than urban (3.7)[151]. HIV prevalence among 15-49 year-olds in Tanzania in 2011–12 was 5.1%, down from 5.7% in 2007–08[152]. Prevalence is higher among urban than rural residents, among women than men, and among older than younger people in the 15–49 age group.

Kisesa ward is located 20 kilometres east of Mwanza city in north-west Tanzania, along the main road to Kenya. A ward is the administrative unit above a village[153]. The Kisesa data are drawn from a cohort study based in Kisesa ward, which encompasses seven villages, five of which are rural, one of which forms an urban trading centre, and one of which takes in peri-urban areas on the main road[154]. The 2012 census found that the ward had a population of 30,486[148]. The demographic surveillance system (DSS) has operated continuously since its establishment in 1994, as an open, geographically defined cohort. The DSS data show that around 52% of the population live in rural areas, 22% in peri-urban areas and 26% in the urban trading centre. The crude death rate among 15–59 year-olds in Kisesa in 2005–2009 was 8.5 per 1000 in men and 6.5 per 1000 in women[155]. Fertility is higher than for Tanzania overall, with a TFR of 7.7 in 2006–07[156]. The trading centre hosts a government-run health centre and several private clinics, and there are small government-run dispensaries in three of the villages[157]. Farming and small-scale agricultural trade are the main sources of income[153].

HIV prevalence in Kisesa in people aged 15+ was 7% in 2008[158], down from a high of 8.3% in 1999–2000[159]. HIV incidence in 15-44 year-olds was 1.1 per 1000 person-years (PY) between 1999/2000 and 2003/04, up from 0.8 per 1000 PY between 1994/95 and 1996/97 but down from 1.2 per 1000 PY between 1996/97 and 1999/2000[160]. In 1994-1996, mortality rates among HIV-positive people aged 15-44 were almost 18 times higher than among HIV-negative people (72.8 vs 4.1 per 1000 PY)[161]. Both prevalence and incidence of HIV are higher in urban/peri-urban areas than in rural areas, but between the late 1990s and 2003/04 both decreased in urban/peri-urban areas and rose in rural areas, toward convergence[160]. Anti-retroviral therapy has been available in Kisesa since 2008 and from nearby Mwanza town since 2005[157]; prior to 2005 the population can be considered ART-naïve.

The study population is all people who died aged 15–59 and while resident in the DSS area. Determination of which deaths were eligible is described below under Data Management and Analyses.

Indicator	**Tanzania**	**Kisesa ward**
Population	44,928,923 (2012)	30,486 (2012)
Population growth	2.7% (average annual increase 2002-12)	*Not available*
Urban/rural population distribution	74% rural 26% urban	52% rural 22% peri-urban 26% urban
Crude death rate	10.8 per 1000 (all ages, 2005–10)	Men: 8.5 per 1000 (15–59, 2005–2009) Women: 6.5 per 1000 (15–59 2005–2009)
Total fertility rate, 15-49 year-olds	5.4 (urban 3.7, rural 6.1; 2010)	7.7 (2006–07)
Adult HIV prevalence	5.1% (15-49 year-olds, 2011–12) 5.7% (15-49 year-olds, 2007–08)	7% (15+ year-olds, 2008) 8.3% (15+ year-olds, 1999–2000)

Table 4: Characteristics of the populations of Tanzania and Kisesa (references in text)

II. Data sources and availability

i. Household enumeration data

The central activity of the DSS is regular household enumeration. Trained interviewers visit households approximately every six months, with the "entire population" of Kisesa ward enumerated[154]. Each round involves several months of data collection. In each round, interviewers collect data on births, deaths and migrations in and out of the households, and on the vital status of all previous household residents. The report of death triggers a VA interview (see below).

These data provide a history of all residency episodes within the DSS area. Each episode ends with a type of exit: continued presence in the DSS area, outmigration, death, or loss to follow-up. Household enumeration data also include the sex, age and residence type (urban/peri-urban/rural) of people resident in the DSS area.

Data from all twenty-six rounds of household enumeration were available for analysis, with the first household enumeration in 1994 and the last in December 2011. Vital status of people who have left the DSS area is asked about in the household enumeration, but the dates of death for people who left the DSS area with exit type "Outmigration" or "Loss to follow up" are only available from round 15 (2002–03) onward, meaning residency status at death of those people cannot be determined for purposes of determining eligibility for analysis (see Data Management and Analyses).

ii. Verbal autopsy data[4]

From the beginning of the Kisesa DSS, verbal autopsies have been conducted to investigate causes of mortality. Coverage of VA interviews for deaths recorded in Kisesa has been reported to be around 75% for deaths in 1999-2001[162].

[4] This section and the following section on physician-review data are heavily based on personal communications with Denna Michael, research scientist, and Lucas Ng'winamilla, VA interviewer, at the National Institute for Medical Research, Mwanza, Tanzania.

The VA process entails administration of a VA questionnaire, review of the interview results by physicians to assign cause of death, and entry of the VA interview results and causes of death assigned by physician review into a database to be stored in electronic format.

Three different questionnaires have been used to collect VA data in Kisesa:

- the DSS used its own VA questionnaire from 1994–2002 (the "Tanesa" questionnaire);
- from 2002–2007, it adopted the questionnaire used by the Indepth network (the "Indepth" questionnaire); and
- in 2007 it began using an adapted version of the WHO-recommended standard VA questionnaire (the "Tazama" questionnaire)[85].

These questionnaires all consist of an open narrative section where the respondent is asked to describe the illness that led to the death, followed by several pages of closed-answer questions recording the presence/absence of symptoms and their duration, and information on lifestyle factors and health-seeking behaviour. The three VA questionnaires are included in Appendix 1.

The three questionnaires record largely the same information; the important difference is that the Indepth questionnaire lacks several symptoms that are relatively specific for HIV disease (vaginal tumours, oral candidiasis, herpes zoster and abscesses/sores) and are included in the other questionnaires[87].

Following the report of a death in the household enumeration, a VA interview is conducted by a trained VA interviewer, who is a clinical officer (someone with three years' clinical training). The fieldwork protocol states that VA interviews should be completed within 4 months of the death, although a substantial proportion of VA interviews have been conducted more than a year after the death. The interview is conducted with a family member or other respondent who is best placed among those available to recall the symptoms suffered by the deceased during their final illness.

From 2004 until 2008, death reports were prepared once the DSS data had been processed at the end of the round of household enumeration, meaning VA interviews only happened after the end of the enumeration and data processing. From 2008 until 2012, DSS enumerators completed death registration and VA request forms during the enumeration, and VA interviews were arranged to take place concurrent with ongoing household enumeration. At

present (2012 onward), household enumeration is electronic and a VA request is automatically generated when a death is reported. It is unclear what the practice was before 2004.

Not all deaths received a VA interview: I assess potential selection bias in the chapter Assessment of Selection Bias. The available verbal autopsy data include only the responses to the closed questions: the narrative section of the VA interview is not entered in the database in electronic format and does not form part of the present dataset, though it was available to the physicians who reviewed the data to assign causes of death.

iii. Physician review data

Transcripts of VA interviews are reviewed independently by two trained physicians who conduct clinical and research work in the area and are trained in assigning causes of death according to ICD-10. These physicians are asked to assign cause of death according to a standard four-line death certificate, identifying the underlying cause of death, as well as the immediate and contributory causes. The physicians assign an ICD-10 code and a description for the underlying cause of each death.

There is formally a process whereby records for deaths with discrepant underlying causes assigned by the two reviewing physicians are sent to a third physician in Tanzania, and cause of death is recorded where two of the three agree; however, no third-physician-review of discrepant causes of death has been done, so the available data do not contain a single authoritative cause of death assigned by physician review.

Physician-review data were only available for VA records that used the Tazama questionnaire, comprising interviews conducted from 2007 onward but also covering deaths before that date. No physician-review data were available for VA records that used the Tanesa or Indepth questionnaires, and consequently no cause of death assigned by physician review is available for these deaths. I assess potential selection bias in the chapter Assessment of Selection Bias.

iv. HIV status data

Data on HIV status of people resident in the DSS area are available for a subset of people who opt to take part in anonymous HIV testing for research purposes ("sero-surveys"). Sero-surveys have been conducted by DSS staff every three years since 1994[157]. All people aged 15 or over at the time of the sero-survey and resident in the DSS area at the last round of household enumeration are eligible and invited to participate. There is a process of informed consent, separate to that governing participation in the household enumeration, which explains the purpose of the sero-survey and makes clear that participants will not be told the results of their HIV test. Testing is anonymous, with participants providing only their unique ID number to study administrators. Over time, participation in the sero-surveys has declined from 74% in 1994-95 to 61% in 2006-07[155, 159], but repeat testing between rounds for those remaining resident is over 90%[160].

Since the 2000–01 sero-survey, temporary confidential HIV testing services have been available alongside the research testing for those wishing to know their status[155], and a permanent centre for voluntary counselling and testing (VCT) has been available in Kisesa trading centre since 2005[157].

Blood samples are tested for HIV at the laboratory of the National Institute for Medical Research (NIMR) in Mwanza; HIV status is determined by two reactive enzyme-linked immunosorbent assay (ELISA) results, and discrepant results are re-tested using two ELISA tests[155]; samples that remain discrepant are excluded from analysis.

All data from six sero-surveys were available for analysis, with the last HIV status recorded in October 2010.

2. Data sources and study population: Manicaland

I. Study setting and population

Zimbabwe is a country in southern Africa. According to a census, the population in 2012 was 13,061,239, growing at 1.1% per year between 2002 and 2012[163]. Two thirds of people live in rural areas. The crude death rate in Zimbabwe was estimated to be 10.2 per 1000 across all ages for the period 2002–2012. The total fertility rate (TFR) nationally was 3.8 in 2011–2012[163]

(Table 5). HIV prevalence among 15-49 year-olds in Zimbabwe in 2010–11 was 15%, down from 18% in 2005–06[164]. Prevalence is higher among urban than rural residents, among women than men, and among older than younger people in the 15-49 age group.

Manicaland is a province in eastern Zimbabwe, bordering Mozambique. Its population in 2012 was 1,752,698[163]. According to the 2012 census, 83% of the population live in rural areas[163]. The Manicaland data are drawn from a cohort study based in twelve sites across the province. These comprise two small towns, four forestry tea and coffee estates and six rural areas (four subsistence farming areas and two roadside trading centres)[165]. The demographic surveillance system (DSS) has operated continuously since 1993, as an open, geographically defined cohort[166]. Baseline enumeration for the HIV/STD Prevention Project took place in 1998–2000[165]. The project is not a traditional DSS in that it does not conduct frequent rounds of household enumeration; rather, it was set up specifically to investigate the dynamics of the HIV epidemic. The enumerated population has varied over time: it was around 9,000 at baseline, 7,000 in round 2, 15,000 in round 3 and 12,000 in round 4[165] (Simon Gregson personal communication). The mortality rate among 15–59 year-olds in the Manicaland study in 2003–2005 was 31 per 1000 person years for men and 26 per 1000 person years for women[167]; the crude death rate was almost identical to that of Zimbabwe as a whole, at 10.3 per 1000 people of all ages in 2002–2012[163]. Fertility in Manicaland province is joint-highest in Zimbabwe, with a TFR of 4.3 in 2011–2012[163].

Between 1998–2000 and 2006–2008, HIV prevalence fell from 20.5% to 13.5% in men, and from 25.9% to 18.4% in women[168]. HIV incidence among men was 1.8% in 2001, 0.8% in 2003 and 1.1% in 2006; in women incidence was 1.6% in 2001, 1.1% in 2003 and 1.4% in 2006[168]. In 2003–2005, mortality rate ratios for HIV-positive compared to HIV-negative 15–59 year-olds were over 10 for men and over 12 for women (104.9 vs 8.2 per 1000 PY for men, 88.3 vs 6.8 per 1000 PY for women)[167]. Anti-retroviral therapy has been available in Manicaland since mid-2005[169], prior to which the population can be considered ART-naïve.

The study population is all people who died aged 15–59 and while resident in the DSS area. Determination of which deaths were eligible is described below in the section Data Management and Analyses.

Indicator	Zimbabwe	Manicaland
Population	13,061,239 (2012)	7,000–15,000
Population growth	1.1% (average annual increase 2002-12)	*Not available*
Urban/rural population distribution	67% rural 33% urban	83% rural (DSS sites) 17% urban (DSS sites)
Crude death rate	10.2 per 1000 (all ages, 2002–2012)	10.3 per 1000 (all ages, 2002–2012)
Total fertility rate, 15–49 year-olds	3.8 (2011–2012)	4.3 (2011–2012)
Adult HIV prevalence	15% (15–49 year-olds, 2010–11)	14%/18% (men/women 15–54 years old, 2006–2008)
	18% (15–49 year-olds, 2005–06)	21%/26% (men/women 15–54 years old, 1998–2000)

Table 5: Characteristics of the populations of Zimbabwe and Manicaland (references in text)

II. Data sources and availability

i. Household enumeration and HIV status data

In Manicaland, survey rounds take place every 2–3 years, with each round taking around two years to complete. Up to three attempts are made to visit each household at each round. Five rounds have been completed to date (1998–2000, 2001–03, 2003–05, 2006–08 and 2009–11). The household visits in each round include both household enumeration and blood sampling for determining HIV status. The information collected for household enumeration is similar to that in Kisesa, but without enquiry into the vital status of all previous household residents. The report of death triggers a VA interview (see below).

The household enumeration data provide a history of all residency episodes within the DSS area. Each episode ends with a type of exit: continued presence in the DSS area, outmigration, death, or loss to follow-up. Household enumeration data also include the sex, age and residence type of people resident in the DSS area (small towns/agricultural estates/roadside trading settlements/subsistence farming areas). The ages at which people are eligible for

enumeration have varied over the rounds of data collection: initially only women aged 15–44 and men aged 15–54 were eligible for enumeration, but from round three onward the eligible ages have been 15–54 for both men and women. People eligible at enumeration were not censored upon becoming older than the upper threshold for eligibility. The 55–59 year-old men and 45–59 year-old women do not represent entirely the same sample as the younger ages.

Blood samples are taken by finger-prick as dried blood spots on filter paper. HIV status is determined by a dipstick-dot immunoassay with sensitivity and specificity both equal to 99.6%[170].

Participation was 77–80% in the baseline enumeration (1998–2000) and first follow-up (2001–2003). In the latter round, 54% of men and 66% of women not known to have died were included in the follow-up enumeration[165]. Data from five rounds of household enumeration and blood sampling were available for analyses, with the last date in July 2011.

ii. Verbal autopsy data

Verbal autopsies have been conducted in the Manicaland DSS since 2001, with reported coverage of over 90% of deaths[87].

The VA process entails administration of a VA questionnaire, review of the interview results by physicians to assign cause of death, and entry of the VA interview results and causes of death assigned by physician review into a database to be stored in electronic format.

The questionnaire used to collect VA data in Manicaland consists of sections on the social circumstances and financial implications of the death, followed by an open narrative section where the respondent is asked to describe the cause of death, followed by several pages of closed-answer questions recording the presence/absence of symptoms and their duration, and information on lifestyle factors and health-seeking behaviour. The VA questionnaire is included in Appendix 1.

Following the report of a death in the household enumeration, a VA interview is conducted by a trained VA interviewer, who is a research nurse. As there is a long time between rounds of household enumeration, many VA interviews are conducted a year or more after the death,

but are almost all conducted within two weeks of the death being enumerated in the household data collection activities (Simon Gregson, personal communication). The interview is conducted with a family member or other respondent who is best placed among those available to recall the symptoms suffered by the deceased during their final illness.

Not all deaths received a VA interview: I assess potential selection bias in the chapter Assessment of Potential Bias. The available verbal autopsy data include only the responses to the closed questions: the narrative section of the VA interview is not entered in the database in electronic format and does not form part of the present dataset, though it was available to the physicians who reviewed the data to assign causes of death. VA data relating to all rounds of data collection were available for analysis.

iii. Physician review data

Two rounds of physician review of verbal autopsy have been conducted, for deaths occurring between rounds 1 and 2, and between rounds 2 and 3. The first set of VA transcripts were coded by two British medical students with no experience in southern Africa, and were considered by Manicaland researchers to be of very poor quality (Simon Gregson, personal communication). The second set of VA transcripts were coded by two Zimbabwean physicians.

These physicians were asked to assign a cause of death from one of 19 categories, and also to say whether they believed the deceased to have been HIV-positive. HIV/AIDS was not one of the cause of death groupings. Where the two reviewers assigned discrepant causes, they were asked to reconcile and agree a cause of death.

Only the physician-review data for the second set of physician reviews were available (N=227). Unfortunately, the categories of death assigned by reviewing physicians were inconsistent with the categories used in this thesis: HIV/AIDS was not one of the 19 cause-of-death categories; there was no valid way to determine which deaths, among those estimated to have been to be of HIV-positive people, were believed to be due to HIV/AIDS by the reviewers. Therefore even these available records could not be analysed, and no physician-review analysis was possible using data from Manicaland.

B. Data management and analysis

This chapter presents definitions of which deaths were eligible for analysis, the standard cause-of-death categories for cause-specific mortality distributions used in the physician-review and InterVA analyses, and the definition of the HIV status of people who died following a negative HIV test. It describes the processes of data cleaning, particularly for the data from Kisesa, the cleaning of which I was primarily responsible for. Finally, this chapter outlines the analyses conducted. Detailed descriptions of the processes used to put data in the necessary formats for analysis by the individual methods of interpreting VA data are described in the method-specific chapters.

3. Data management and analysis: Kisesa

I. Definitions necessary for analyses

i. Deaths eligible for analyses

To be eligible for inclusion in the analyses, records had to relate to people who were:

a) recorded as 15–59 years old at death according to the household enumeration; and
b) resident in the DSS area at death.

To determine which people had died aged 15–59, I used the age recorded in the DSS enumeration, as this having been recorded prospectively and while the person was living was more likely to be accurate than age retrospectively recorded in the VA interview.

The restriction to people resident at death was made to ensure the findings represented as accurately as possible the true situation in Kisesa. People were assumed to have probably been resident in the DSS at the time of their death if their type of exit from the DSS area, recorded in the DSS enumeration, was one of the following:

1) Death in their household in the DSS area;
2) Outmigration to another household in the DSS area before death;
3) Outmigration beyond the DSS area less than six months prior to death, as these people are likely to have left their homes to seek medical care; or

4) Loss to follow up less than two years prior to death, as this suggests nobody in the household was available for interview at one or more rounds of DSS enumeration, possibly due to a death-related crisis, but that the VA interview was later conducted successfully.

To calculate time between exit and death for outmigrants beyond the DSS area and people lost to follow up required dates of exit from the DSS area and dates of death. Dates of exit were those recorded in the DSS enumeration. For people whose type of exit was either "Outmigration" or "Loss to follow up", I used dates of death from the DSS enumeration where these were available – namely, the date of death recorded in response to enquiry about the deaths of no-longer-resident household members. Where dates of death from the DSS enumeration were unavailable, I used dates of death recorded in the VA interview.

Deaths in the DSS with all other types of exit (outmigration/loss to follow up outside the time constraints, or continuing presence in the study area) or lacking necessary information such as the date of death, and VA records linked to such deaths, were excluded from the analyses.

ii. Standard cause-of-death categories

The analyses using physician review and InterVA assigned cause-specific mortality distributions. To assign deaths to standard groups of causes, I used the 54 groupings in the WHO 2012 Verbal Autopsy Standards[171] (excluding neonatal causes), henceforth 'cause groups'. These are groups in which similar causes of death are brought together into a single group comprising several ICD-10 codes. For example, the "Meningitis/encephalitis" group encompasses codes A39 and G00–G05. Less prevalent causes of death are grouped together in "Other/unspecified" groups (such as "Other/unspecified infectious disease", encompassing codes A20–A38, A42–A89, B00–B19, B25–B49, B55–B99). Table 6 gives the full distribution of ICD-10 codes by cause groups, as published by WHO.

This classification of ICD-10 codes into groups in the original WHO publication contained some errors:

- several ICD-10 codes for infectious diseases were classified in the group "Other/unspecified non-communicable diseases";

- several ICD-10 codes for neoplasms were classified in the group “Other/unspecified neoplasms” when a more specific group (“Reproductive neoplasms” or “Digestive neoplasms”) was available.

Cause groups	**Associated ICD-10 codes**
Infectious diseases	
HIV/AIDS-related	B20-B24
Sepsis (non-obstetric)	A40-A41
Acute respiratory infection/pneumonia	J00-J22
Diarrhoeal diseases	A00-A09
Malaria	B50-B54
Measles	B05
Meningitis and encephalitis	A39; G00-G05
Tetanus	A33-A35
Pulmonary tuberculosis	A15-A16
Pertussis	A37
Haemorrhagic fever	A90-A99
Other/unspecified infectious diseases	A17-A19; A20-A38; A42-A89; B00-B19; B25-B49; B55-B99
External causes	
Road traffic collision	V01-V89
Other transport incident	V90-V99
Accidental fall	W00-W19
Accidental drowning	W65-W74
Exposure to smoke/fire	X00-X19
Venomous plant/animal	X20-X29
Accidental poisoning	X40-X49
Intentional self-harm	X60-X84
Assault	X85-Y09
Exposure to force of nature	X30-X39
Other/unspecified external causes	S00-T99; W20-W64; W75-W99; X50-X59; Y10-Y98
Non-communicable diseases	
Oral neoplasms	C00-C06
Digestive neoplasms	C15-C26
Respiratory neoplasms	C30-C39
Breast neoplasms	C50
Reproductive neoplasms	C51-C58; C60-C63
Other/unspecified neoplasms	C07-C14; C40-C49; C60-D48
Severe anaemia	D50-D64
Severe malnutrition	E40-E46

Diabetes mellitus	E10-E14
Acute cardiac disease	I20-I25
Stroke	I60-I69
Sickle cell with crisis	D57
Other/unspecified cardiac disease	I00-I09; I10-I15; I26-I52; I70-I99
Chronic obstructive pulmonary disease	J40-J44
Asthma	J45-J46
Acute abdomen	R10
Liver cirrhosis	K70-K76
Renal failure	N17-N19
Epilepsy	G40-G41
Other/unspecified non-communicable disease	D55-D89; E00-E07; E15-E35; E50-E90; F00-F99; G06-G09; G10-G37; G50-G99; H00-H95; J30-J39; J45-J99; K00-K31; K35-K38; K40-K93; L00-L99; M00-M99; N00-N16; N20-N99; R00-R09; R11-R94
Maternal causes	
Ectopic pregnancy	O00
Abortion-related death	O03-O08
Pregnancy-induced hypertension	O10-O16
Obstetric haemorrhage	O46; O67; O72
Obstructed labour	O63-O66
Pregnancy-related sepsis	O85; O75.3
Anaemia of pregnancy	O99.0
Ruptured uterus	O71
Other/unspecified maternal cause	O01-O02; O20-O45; O47-O62; O68-O70; O73-O84; O86-O99
Cause of death unknown	
Cause of death unknown	R95-R99

Table 6: Cause groups and associated ICD-10 codes according to the WHO 2012 Verbal Autopsy Standards

Table 7 shows the ICD-10 codes that were wrongly assigned, and the description attached to the ICD-10 code in the online ICD-10 directory (at http://apps.who.int/classifications/icd10/browse/2010/en). This is not an exhaustive list of erroneously grouped codes: it only includes errors among codes reported in my dataset. In order that the analyses included as much and accurate information as possible, and particularly to ensure that infectious causes were recognised as such, the groupings of ICD-10 codes used to create the cause groups for the analyses included these corrections.

ICD-10 code [description]		Cause group according to WHO VA standards	Cause group assigned
L03.9	[cellulitis]	Other/ unspecified non-communicable diseases	Other/ unspecified infectious diseases
J98.8	[unspecified respiratory disorder – most are infectious]		
J36	[peritonsillar abcess]		
K93	[tuberculous intestinal disorders]		
N34	[urethral abcess]		
L08.9	[skin infection]		
M60.0	[infective myositis]		
N10	[acute tubulo-interstitial nephritis]		
N15.9	[kidney infection]		
D26.1	[benign neoplasm of corpus uteri]	Other/ unspecified neoplasms	Reproductive neoplasms
D29.1	[benign neoplasm of prostate]		
D39.0	[neoplasm of uterus of uncertain behaviour]		
D39.9	[neoplasm of female genital organ of uncertain behaviour]		
D37.4	[neoplasm of colon of uncertain behaviour]		Digestive neoplasms

Table 7: ICD-10 codes that the WHO VA standards erroneously classify, and their corrected cause groups

I presented "HIV/AIDS-related" separate from other infectious cause groups. For summary presentation of causes of death, I used the broad cause categories "HIV/AIDS", "Non-HIV infections", "Non-communicable diseases", "Maternal causes", "External causes" and "Cause of death unknown". Deaths in the cause group "Pregnancy-related sepsis" were categorised under "Maternal causes" rather than "Non-HIV infections".

iii. Definition of HIV status

People were classified as HIV-positive at death if they had ever had a positive HIV test. People were classified as HIV-negative if they had had a negative last test within the last five years –

the period someone was assumed to be HIV-negative following a negative HIV test, consistent with other literature[155] although a shorter period of three years has also been used[87, 172]. Five years is the assumed period for Kisesa used within the Alpha network analyses. The period between the last HIV test and death was taken as the difference between the last recorded HIV-test date (which was assumed to be reliable) and the date of death. The date of death was that recorded in the household enumeration data where this was available, and otherwise was the date of death recorded in the VA interview.

People whose date of death was prior to their last reported HIV test result were treated as having unknown HIV status, because such incongruity of dates raised doubt about the accuracy of the record linking. People without known HIV status (due to an inconclusive test or to not having been tested), and people who had had a negative test result more than five years prior to death, were also classified as having unknown HIV status.

In addition, the specificity analyses were conducted twice, once using this HIV status and once treating as HIV-positive anyone whose VA interview included a report of a diagnosis of HIV prior to death. This was to allow comparison with another study investigating the same question, which used the latter definition of HIV status[102].

iv. Definition of symptom profiles

Where the symptom profile was investigated in order to elucidate the reasons for assignation of unknown cause of death or false-positive assignation of HIV, "symptoms" were defined as: a positive report of any of a list of signs, symptoms or other useful indicators including diagnoses, medication or treatments received; or any report of the length of the illness or the season in which death occurred. A full list of symptoms is given in Appendix 2.

II. Data management

i. Receipt

Table 8 shows the datasets I received to prepare the analyses. All datasets were Stata datasets in .dta format. I received raw VA and physician-review data in 11 files: six covering VA

interviews conducted using the Tanesa questionnaire (TanVA1–TanVA6); one covering VA interviews conducted using the Indepth questionnaire (IndVA); and four covering VA interviews conducted using the Tazama questionnaire. Two of these Tazama datasets contained both VA data and the results of the physician review (TazVA1 and TazVA2), while for the third set of VA data (TazVA3) the physician review results were sent separately (TazPR3).

I received the household enumeration data in two files: one containing all residency episodes recorded during the history of the DSS, which included data on sex, age, residence type, and date and type of exit from the household (Residency_Episodes); and one containing the vital status of everyone recorded in the household enumeration, and the dates of death for those who had died from round 15 onwards (Vital_Status).

I received the HIV-status data in one dataset (HIV_Status). The HIV-status data were provided with a unique individual identifier ("linking identifier") different to the identifier in the VA and household enumeration data ("DSS identifier"). I received a further dataset that enabled me to link the DSS identifier with the linking identifier.

ii. Conversion

Raw VA data were converted to a standard format outlined in the data specification developed for the Alpha Network Workshop 8 on mortality, held in Kisumu, Kenya in October 2011 ("Spec 8.1"). Spec 8.1 contains 168 variables relating to the symptoms and circumstances prior to death, as well as personal identifiers and the date of interview. Appendix 3 details Spec 8.1 and indicates where certain variables could not be created due to the absence of relevant questions in the VA questionnaires.

In addition to such absent information, the information received in the VA datasets meant that creating some variables in Spec 8.1 was not straightforward and I made a judgement about whether and how to create such variables. Appendix 4 presents nine variables for which I made such judgements.

iii. Data linking, deduplication and creating the dataset for analyses

Two variables served as unique identifiers: the DSS identifier, an eleven-digit number that is unique to each resident listed in the household enumeration; and the linking identifier, a number used to link the non-anonymous household enumeration data with the anonymous data on HIV status. I used the DSS identifier to link VA records to DSS residency episodes and to the linking dataset. I then used the linking identifier to link VA/DSS records to the HIV-status dataset.

Almost all deduplication concerned duplicates in the DSS identifier. Duplicates took two forms: identical records from a single VA interview that appeared in more than one dataset; and multiple records for a single death due to more than one VA interview having been conducted for that death.

Surplus identical records were dropped. Where two non-identical VA records had the same DSS identifier, I checked whether the records seemed to indicate the same individual (suggesting that two separate interviews were conducted for one death) or whether they seemed to indicate different individuals (suggesting an error in one of the DSS identifiers). Where the names of the deceased person matched in the duplicate records, I assumed they were for the same person. Where names could not be compared, I assumed the records pertained to the same person if the sex matched and the date of death and age at death were similar.

Where two non-identical VA records existed for the same person, I retained the record which had:

- the most yes/no responses to symptom questions (as opposed to "don't know" or blank answers), as this suggested greater familiarity with the terminal illness; or where this did not discriminate,
- an earlier VA interview, as this would have had a shorter recall; or where this did not discriminate,
- a more closely related respondent, as this person may have been closer to the deceased prior to death.

Where names differed, or where there were differences in sex, date of death and age at death, I assumed that one of the duplicate DSS identifiers was wrong. I used the Residency_Episode file to determine which VA record matched the household enumeration records on sex and age of the deceased, and assumed the DSS identifiers in such matching cases to be correct. For the remaining records with erroneous DSS identifiers, I used the household enumeration data to seek deaths in people with similar DSS identifiers and matching data on sex, date of death and age at death. Where a plausible death with a similar DSS identifier existed, I corrected the erroneous DSS identifiers. Where no correct DSS identifier could be found, I dropped the VA record.

There was substantial overlap between the six datasets containing VA records from the era of the Tanesa questionnaire, and there were far fewer duplicates between later records. Table 9 shows the reduction in the number of records through deduplication of the datasets originally received, and through failures in linking datasets. The total number of unique VA records for adults of all ages was 2148. I dropped VA records with DSS identifiers that did not link to the linking dataset, as this failure raised doubts about the accuracy of the DSS identifiers. Where there were records with duplicate linking identifiers, I dropped the records received in the most recent VA dataset as these had had least chance of errors being ascertained and were therefore most likely to contain erroneous linking identifiers. I reported on all duplicate and non-linking records to the data management team at NIMR. The total number of unique VA records linked to the household enumeration, covering adults of all ages, was 2115.

Dataset	Description	Function	Source	Unique record identifiers
TanVA1–TanVA6	Raw VA data	VA interview data from the era of the Tanesa VA questionnaire	LSHTM (Basia Zaba)	DSS identifier
IndVA	Raw VA data	VA interview data from the era of the Indepth VA questionnaire	NIMR (Raphael Isingo)	
TazVA1 & TazVA2	Raw VA data and physician-assigned causes of death	VA interview data and physician-assigned cause-of-death data from the era of the Tazama VA questionnaire	NIMR (Chifundo Kanjala)	
TazVA3	Raw VA data			
TazPR3	Physician-assigned COD			
Residency_Episodes	Data on residency episodes	Assessing errors in DSS identifiers in the VA datasets; Obtaining information on type and date of exit, to assess eligibility and selection bias; Determining sex and residential area of the deceased.	LSHTM (Milly Marston)	DSS identifier
Vital_Status	Data on vital status of all people enumerated	Determining eligibility for analysis of people who died after outmigration or loss to follow up Providing denominator for assessing selection bias		DSS identifier
HIV_Status	Data on HIV status	Calculation of specificity and analysis of COD by HIV status		Linking identifier
Tazama_ID_link	ID-linking variables	Facilitating linking HIV-status data with other datasets		DSS identifier; Linking identifier

Table 8: Starting datasets received from which I created the VA dataset for analyses. COD=cause of death; NIMR=National Institute for Medical Research, Mwanza, Tanzania; LSHTM=London School of Hygiene and Tropical Medicine.

To determine the total number of deaths that had occurred among the population ever recorded in the household enumeration, I linked the data on residency episodes (Residency_Episodes file) with the data on vital status of all people enumerated (Vital_Status) using the DSS identifier. There were 5539 deaths of ever-resident people recorded in the Vital_Status file, of which 215 (3.9%) did not link to the Residency_Episode file (Figure 1): I dropped these as they were likely to contain erroneous DSS identifiers, leaving 5324. A further 156 deaths were recorded in the Residency_Episodes file but not in the Vital_Status file, meaning 5480 deaths in the DSS area were identified.

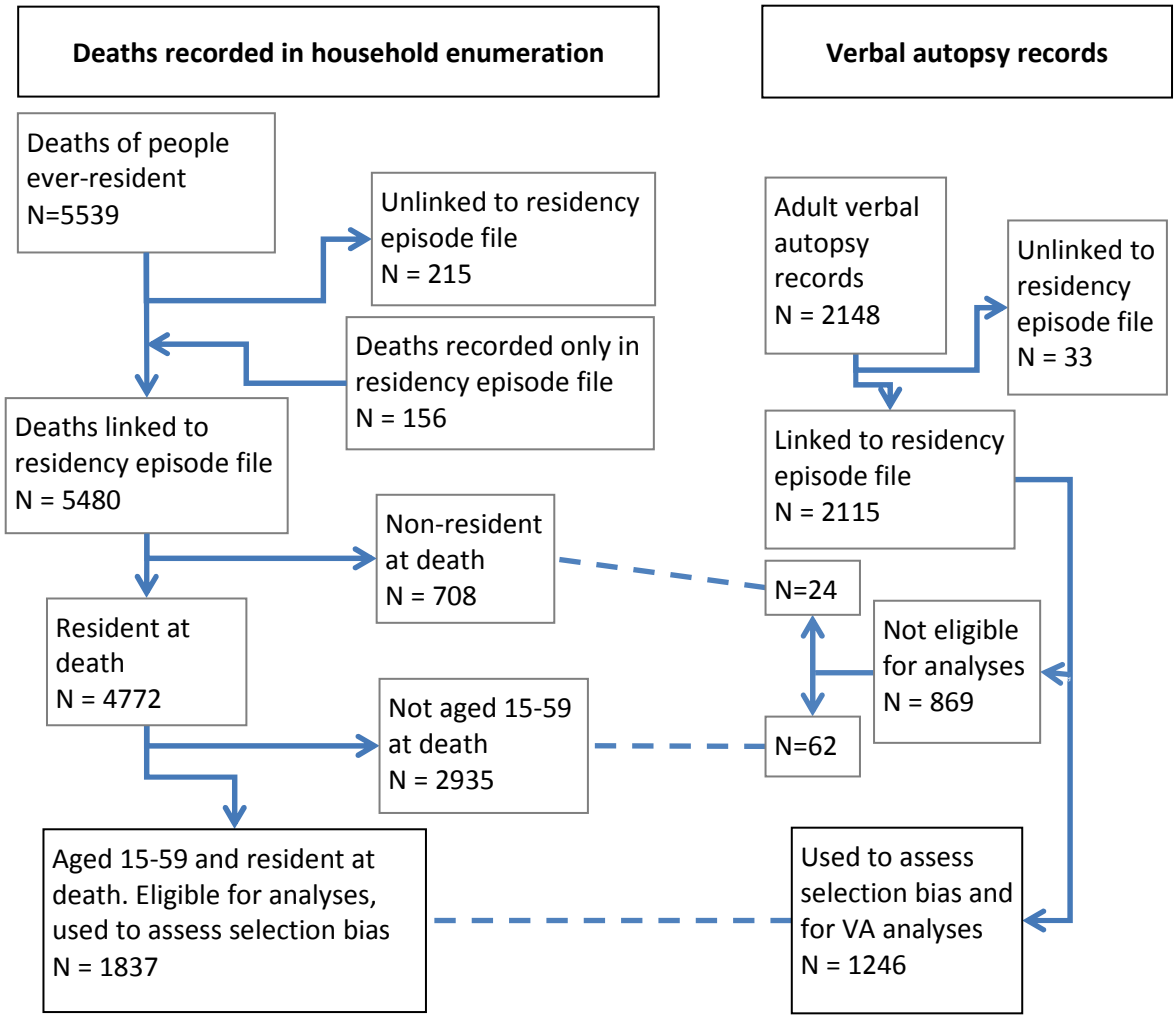

Figure 1: Numbers of deaths from household enumeration and linkage to VA records

To determine which deaths were eligible for analysis, I investigated the time interval between exit from the DSS area and death for those deaths with exit types "Outmigration" or "Loss to follow up", and retained those probably resident at death in line with the criteria outlined

above. A small number of deaths (41, 0.8%) were excluded as their exit type was “Present in study”.

Of the 5480 deaths recorded in the household enumeration data, 4772 were of people resident in the DSS area at death. 1837 of these were aged 15–59, and therefore eligible for analysis (Table 10). Of the 2115 unique VA records linked to deaths in the household enumeration data, 1246 linked to deaths eligible for the analyses (Figure 1).

Reason for loss of records	N records dropped for each reason	Total N records remaining	Raw verbal autopsy datasets & number of VA records									
			TanVA1	TanVA2	TanVA3	TanVA4	TanVA5	TanVA6	IndVA	TazVA1	TazVA2	TazVA3
Starting number of records in each verbal autopsy dataset		**2779**	**645**	**558**	**74**	**95**	**83**	**77**	**110**	**686**	**216**	**235**
Removing duplicate records												
Deduplication between Tanesa-questionnaire datasets: keeping those with more responses / earlier reports / more closely related respondent	613	2166	430	478	11	0	0	0				
Deduplication within Tazama-questionnaire datasets	4	2162								684	215	234
Deduplication between Tazama-questionnaire datasets: keeping those with more responses / earlier reports / more closely related respondent	7	2155								683		228
Deduplication between records from different questionnaires: keeping records with DSS identifier confirmed by comparison with the household enumeration data	6	2149								681		224
Drop one record that becomes a duplicate after DSS identifier is corrected	1	2148		477								
Total unique VA records		**2148**										
Removing non-linking records												
Drop records that do not link to the linking dataset	16	2132		475	10					680		212
Drop records that have duplicate linking identifiers	2	2130										210
Drop records that do not link to the household enumeration data	15	2115	423	473					109	678		207
Total VA records linked to household enumeration data		**2115**	**423**	**473**	**10**	**0**	**0**	**0**	**109**	**678**	**215**	**207**

Table 9: Deduplicating VA records from Kisesa: the change in numbers through combining multiple overlapping datasets

Type of exit from DSS area	Number of deaths N	%	Eligible for analyses?	Resident at death… N	%	…of whom aged 15–59 at death N	%
Present in Study	41	0.8	No	–	–	–	–
Death	4308	78.6	Yes	4308	90.3	1543	84.0
Out-migration	959	17.5	If exited <6 months before death	328	6.9	222	12.1
Lost to follow up	172	3.1	If exited <24 months before death	136	2.8	72	3.9
Total	**5480**	**100.0**		**4772**	**100.0**	**1837**	**100.0**

Table 10: Distribution of deaths recorded in the Kisesa DSS area by type of exit from the DSS area

iv. Records of people not aged 15–59 in the verbal autopsy data

Among the 1246 VA records for people recorded as aged 15–59 in the household enumeration, 75 (6.0%) were not aged 15–59 in the VA record itself (Table 11). Most (56/75, 74.7%) did not have an age recorded in the VA record; of the remaining 19 records, 12 (63.2%) were within one year of the eligible age range and the most outlying were two records of people recorded as 67 years old in the VA interview.

Age in VA record	N	%
14	1	1.3
60	11	14.5
61	2	2.6
63	1	1.3
64	1	1.3
65	1	1.3
67	2	2.6
No age in VA record	56	74.7
Total	**75**	**100.0**

Table 11: Ages recorded in the VA interview for people aged 15–59 in Kisesa household enumeration who were not aged 15–59 in the VA record

III. Analyses conducted

I conducted four main sets of analyses, each of which is described in detail in its respective chapter. Not all eligible VA records could be used for each substantive analysis: some had not received physician review; some lacked necessary information to allow the InterVA model to run; and some were not linked to HIV-status data.

i. Physician review analyses

Some deaths had a cause assigned by physician review (N=462). I assigned each physician review to a cause group, assessed the reliability between reviewing physicians, and presented the cause-specific mortality distributions assigned by physician review. I looked at causes assigned by HIV status, comparing the cause distributions among HIV-negative and HIV-positive people.

I calculated specificity for assignment of non-HIV causes among HIV-negative people. The length of time between negative HIV test and death was presented for HIV-negative people who were assigned to the cause group "HIV/AIDS-related", to determine the period in which those people would have to have seroconverted and died in order for "HIV/AIDS-related" to be an accurate cause of death. For comparison with the findings of another study (Byass et al 2013[102]), I calculated an alternative measure of specificity.

I determined which symptoms occurred frequently among HIV-negative and HIV-positive people assigned to the cause group "HIV/AIDS-related", and assessed the association between the occurrence of those symptoms and false-positive status. I conducted a sensitivity analysis to determine whether the assumed five years of HIV-negative status following a negative HIV test affected the specificity.

ii. InterVA analyses

Some deaths had a cause assigned by InterVA (N=1107). As with physician review, I: presented the cause-specific mortality distributions assigned; compared cause distributions by HIV status; calculated specificity for HIV, and calculated an alternative specificity for comparison with another study; investigated symptoms occurring in deaths assigned to the cause group "HIV/AIDS-related"; and conducted a sensitivity analysis on the assumed period for which someone was HIV-negative following a negative HIV test.

iii. Lopman algorithm analyses

All deaths with a linked HIV status (N=598) were used in the analysis of the Lopman algorithm. As the Lopman algorithm is data-derived and requires a reference standard to operate, I defined deaths as true negative and true positive. I applied the version of the algorithm originally published to the present data and calculated the following metrics: specificity, sensitivity, % correctly classified, % assigned to HIV/AIDS and absolute difference in % assigned to HIV/AIDS compared to the reference standard (measured in percentage points). I re-derived the Lopman algorithm in the present data, and investigated the random variation in its performance according to differences in which records were assigned to the training and testing datasets. I conducted a sensitivity analysis on the assumed period for which someone was HIV-negative following a negative HIV test.

iv. Risk of selection bias affecting the proportion of deaths assigned as "HIV/AIDS-related"

I investigated potential selection bias affecting the proportion of deaths assigned to the cause group "HIV/AIDS-related" by age, sex, residence, year of death and HIV status. Values for age, sex, residence and year of death were taken from the DSS enumeration data where available.

I assessed the risk of selection bias in whether deaths (N=1837) received a VA interview, applicable to the findings of all three interpretative methods. For those records relating to deaths that received a VA interview (N=1246), I assess the risk of selection bias regarding which records were used in analysis by physician review and InterVA, as not all records either received a physician review, or contained sufficient information to run in InterVA. For each interpretative method, overall assessment of the risk of selection bias was assisted by creating a table summarising the direction and magnitude of potential biases – assessment of the direction and magnitude was based on ad hoc criteria described in the chapter.

4. Data management and analysis: Manicaland

I. Definitions necessary for analyses

i. Deaths eligible for analyses

Criteria and methods for determining eligibility for analysis were the same as used for the Kisesa data.

ii. Standard cause-of-death categories

The standard cause-of-death categories, drawn from the WHO VA standards, were the same as those used in the Kisesa analyses.

iii. Definition of HIV status

HIV statuses were defined in the same way as in the Kisesa analyses, except that the time someone was assumed to be HIV-negative following a negative HIV test was 3.75 years, which is the assumed period for Manicaland used within the Alpha network analyses.

iv. Definition of symptom profiles

Symptom profiles were defined in the same way as in the Kisesa analyses.

II. Data management

i. Receipt

I received four datasets:

- Residency episodes (6.1)
- HIV test results (6.2)
- VA data (8.1)
- Physician review data

ii. Conversion

I received all data in standard Alpha Network data specification formats, which needed no conversion. Physician review records were in a specific format containing causes assigned by individual reviewers, and their consensus causes assigned in cases mutually reviewed after initial disagreement.

iii. Data linking, deduplication and creating the dataset for analyses

I linked all datasets using a common unique identifier provided by the data managers. The total number of unique VA records linked to the household enumeration, covering adults of all ages, was 1094. To determine which deaths were eligible for analysis, I investigated the time interval between exit from the DSS area and death for those deaths with exit types "Outmigration" or "Loss to follow up", and retained those probably resident at death in line with the criteria outlined above. In total there were 4563 people resident in the DSS area at death. 3155 of these were aged 15–59, and therefore eligible for analysis. Linking the VA records with the residency-episodes data showed that 1021/1094 VA records were for people aged 15–59 in the DSS record and therefore eligible for analysis. Forty VA records (3.9%) were excluded due to being linked to household enumeration records in which the exit type was "Present in study".

iv. Records of people not aged 15–59 in the verbal autopsy data

Among the 1021 VA records for people recorded as aged 15–59 in the household enumeration, three (0.3%) were not aged 15–59 at death according to the VA record: their VA ages were 60, 61 and 62.

III. Analyses conducted

As with Kisesa, not all eligible VA records could be used for each substantive analysis: some had not received physician review; some lacked necessary information to allow the InterVA model to run; and some were not linked to HIV-status data.

i. Physician review analyses

As noted, physician review analyses were ultimately not possible using the data from Manicaland.

ii. InterVA analyses

Some deaths had a cause assigned by InterVA (N=1016). Analyses were the same as those conducted using the data from Kisesa.

iii. Lopman algorithm analyses

All deaths with a linked HIV status (N=965) were used in the analysis of the Lopman algorithm. Analyses were the same as those conducted using the data from Kisesa.

iv. Risk of selection bias affecting the proportion of deaths assigned as "HIV/AIDS-related"

I investigated potential selection bias affecting the proportion of deaths assigned to the cause group "HIV/AIDS-related" by age, sex, residence, year of death and HIV status. Values for age, sex, residence and year of death were taken from the DSS enumeration data where available.

I assessed the risk of selection bias in whether deaths (N=3155) received a VA interview, applicable to the findings of all three interpretative methods. For those records relating to deaths that received a VA interview (N=1021), I assessed the risk of selection bias regarding which records were used in analysis by InterVA and the Lopman algorithm, as not all records contained sufficient information to run in InterVA or had linked HIV status necessary for use in the Lopman algorithm. Overall assessment of the risk of selection bias was the same as in the analysis of data from Kisesa.

3. Causes of death by physician review

1. Introduction

Physician review is the most widely used method for interpreting verbal autopsy (VA) data, and consists of having physicians read and interpret the transcripts of VA interviews and assign a cause of death based on this interpretation[65]. Stated disadvantages of physician review include the time and money required to have physicians review transcripts, as well as the unreliability of the method: physicians do not necessarily agree with one another when reviewing the same transcript, or with themselves in repeated reviews[107]. This method has most often used two or more physician reviewers to review each VA record, to minimise the effects of individual subjectivity[65], though the cost implications of doing this in large-scale VA activities used for sample death registration in India and Tanzania are high[173]. Recent innovations in VA instruments have assumed that these will be used with computerised methods[171], which are widely proposed as a means of reducing the cost of using VA at scale.

The performance of newly proposed methods is often compared with that of physician review [94, 97, 114, 120, 137]. As with other methods of interpreting VA data, validation of physician review tends to involve assessment of specificity against a reference standard, as well as sensitivity where the reference standard allows. Among studies that have sought to ascertain HIV/AIDS-related mortality from VA data using physician review, some have calculated specificity – the proportion of "true negative" deaths that are not assigned HIV/AIDS as cause. The definitions of a true negative death used by these studies have included: deaths in people with HIV-negative status[96]; deaths with non-HIV clinical diagnosis[100]; and, using clinical diagnosis and HIV status, allowing true negative cases to include HIV-positive people who did not have an AIDS-defining condition[174]. Where more than one physician is involved, as is often the case, validation of physician review often further entails assessing the inter-rater reliability between reviewing physicians.

This chapter uses data from routine activities in the demographic surveillance systems in Kisesa to assess the specificity and reliability of physician review in diagnosing HIV/AIDS-related deaths, and to investigate causes of death by HIV status. It begins by presenting the objectives and methods used in the physician-review analyses. It then presents the results for Kisesa, covering the causes assigned by the respective physicians, the agreement between them and reliability of the method, the final cause-specific mortality distribution assigned, and the causes of death assigned by HIV status, including specificity.

2. Objectives

I. Overall objective

To assess the reliability and specificity of physician review for the assignment of HIV/AIDS as cause of death from verbal autopsy data.

II. Specific objectives

1. To assess the reliability of physician review in assigning causes of death, with emphasis on HIV/AIDS.
2. To describe the cause-specific mortality distributions in the population under investigation.
3. To use known HIV status to:
    a. calculate the specificity of physician review for assigning HIV/AIDS as cause of death; and
    b. investigate causes of death by HIV status.

3. Methods: Kisesa

I. Assigning individual physician reviews to cause groups

Two physicians independently reviewed the VA transcripts and assigned a cause to each death. The outputs from the physician reviews consisted of an ICD-10 code[175] and a written description of the cause of death. I first examined the reviews provided by each physician separately, whether or not the two physicians agreed on the cause of death. Since there were a large number of different ICD-10 codes, I pooled related causes of death into the cause groups derived from the WHO VA standards[171], as described in the Data Management and Analysis section of the General Methods chapter.

When the ICD-10 code and the description provided by the physician were consistent, I assigned the review to the cause group containing that code. When the ICD-10 code and the description were inconsistent, I assigned the review to the cause group indicated by the ICD-10 code, except where the code was a probable data-entry error. For example, where the

description indicated malaria, but the assigned code was B24 ("Unspecified HIV") rather than B54 ("Unspecified malaria"), I assigned the review to the cause group "Malaria".

If the physician provided an ICD-10 code without giving a description, I assigned the review to a cause group based on the ICD-10 code alone. Where the physician provided only a description and no ICD-10 code, I assigned the death to a cause group based on the description alone.

I present summaries of the cause-distributions assigned by each physician, and investigated discordance in their assignments, using the broad cause categories described in the Data Management chapter.

II. Assessing the reliability of physician review

Reliability of physician review as to the causes of deaths was assessed for records where both physicians had assigned a cause of death. Reliability was assessed separately for three-character ICD-10 codes, cause groups, and the cause group "HIV/AIDS-related".

In assessing reliability in assigning cause groups, all deaths that were assigned to different cause groups by the two physicians were scrutinised for consistency, and corrected where relevant. For example, where both physicians described a snake bite, one physician might assign the ICD-10 code T63.0 ("Bite – snake"), which is in cause group "Other/unspecified external causes", and the other might assign X20 ("Bite – bitten by snake") which is in cause group "Venomous plant/animal". I assigned such reviews to consistent cause groups.

To assess reliability, I report the percent agreement between the two physicians across ICD-10 codes, across cause groups, and for assigning deaths to the cause group "HIV/AIDS-related" versus all other cause groups. I also report the Kappa statistic, describing the degree of observed agreement using the scale proposed by Landis and Koch[176].

III. Cause-specific mortality distribution

Cause-specific mortality distributions (CSMD) were based on the cause groups. As discrepant physician reviews were not reconciled, I investigated the respective physician reviews in two ways: looking at the CSMD assigned by each of the respective physicians, and looking at the CSMD assigned using both physician reviews. I compared the CSMDs assigned by the

respective physicians and using both reviews, using the Z-test to ascertain differences in the proportions of deaths assigned to the cause groups using the different definitions.

i. CSMDs assigned by single physicians

I looked at the overall cause-specific mortality distribution assigned by each physician, and compared the proportion assigned to each cause group by the respective physicians to ascertain whether identity of the physician made a difference to the CSMD assigned.

ii. CSMDs assigned using both physician reviews

Where two physicians agreed on the cause group, the death was assigned to that group. Deaths for which the VA record was reviewed by one physician only, and deaths for which the physicians did not agree on the cause group, were assigned as having "Cause of death unknown".

The exception to this procedure was that, when one of the physicians had assigned the death to "HIV/AIDS-related" and the description in the other physician review described a condition related to HIV, the death was assigned to "HIV/AIDS-related". Conditions related to HIV are those indicating stage 3 or 4 HIV disease ("advanced HIV disease" and "AIDS" respectively) in the WHO clinical staging criteria[14]. This was limited to those conditions with signs/symptoms that enable diagnoses to be made clinically, as the reviewing physicians would not have access to the laboratory data necessary for diagnosing conditions for which no presumptive clinical diagnosis is possible (Table 12). No statistical tests were performed.

The CSMD presents the cause groups assigned, and the summary broad cause categories. I considered presenting the CSMD excluding the deaths that were assigned to "Cause of death unknown" due to a discrepancy between the physicians; this approach would have reduced the denominator and increased the proportion of deaths assigned to each named cause group and broad cause category. As the analysis of reliability showed that agreement between physicians was higher for HIV/AIDS than for overall cause groups, deaths assigned to HIV/AIDS would have been disproportionately likely to have been retained in an analysis that excluded deaths with unknown cause due to discrepancy between reviews. I therefore decided to

retain the "Cause of death unknown" in the CSMD to avoid inflating the proportion of deaths due to HIV/AIDS. I also considered assigning discrepant reviews a fractional weight of 0.5 to each cause group. I decided against this as it would have meant there was no distinction in terms of contribution to the CSMD between deaths where the physicians agreed and those where they did not, and physician agreement is an important component part of physician review with multiple reviewers[65].

Condition	**ICD-10 codes indicating conditions (assigned by author)**
Stage 3	
Unexplained severe weight loss	R63
Unexplained chronic diarrhoea for longer than one month	No ICD-10 code
Unexplained persistent fever	No ICD-10 code
Persistent oral candidiasis	B37
Oral hairy leukoplakia	K13
Pulmonary tuberculosis (current)	A15-A16
Severe bacterial infections (such as pneumonia, empyema, pyomyositis, bone or joint infection, meningitis or bacteraemia)	J15, J18; J86; M60; M00-M01, M86; G00-G01; A49
Acute necrotizing ulcerative stomatitis, gingivitis or periodontitis	A69; K05
Stage 4	
HIV wasting syndrome	No ICD-10 code
Pneumocystis pneumonia	B59; J18
Recurrent severe bacterial pneumonia	J15; J18
Chronic herpes simplex infection (orolabial, genital or anorectal of more than one month's duration or visceral at any site)	B00; A60
Oesophageal candidiasis (or candidiasis of trachea, bronchi, lungs)	B37
Extrapulmonary tuberculosis	A17-A19
Kaposi's sarcoma	C46
Cytomegalovirus infection (retinitis or infection of other organs)	B25
Central nervous system toxoplasmosis	B58; G05
HIV encephalopathy	G93.4
Extrapulmonary cryptococcosis including meningitis	B45

Table 12: Conditions defining stage 3/4 HIV disease, limited to selected signs/symptoms and diagnoses that can be made clinically

IV. Assessing causes of death against known HIV status

I present the broad cause categories assigned under III above by HIV status (HIV-negative, HIV-positive and unknown HIV status). To determine associations between broad cause categories

and HIV status, I compared the proportion assigned to each broad cause category among HIV-negative and HIV-positive people, using the chi-squared test or the Fisher's exact test where the chi-squared test was not valid (Cochran 1954, cited in Kirkwood and Sterne 2003[177]). For each broad cause category, the chi-squared test was conducted on a two-by-two table showing the binary assignment to the cause category or not, for all HIV-negative and HIV-positive people. I calculated the specificity – the key indicator of validity measured in this study – as the proportion of deaths of HIV-negative people assigned to cause groups other than "HIV/AIDS-related". The Wilson method without continuity correction was used to create 95% confidence intervals[178]. I conducted a sensitivity analysis of the effect on specificity of the assumed length of the post-negative period, using the comparison of two proportions[177] with the command -prtesti- in Stata 12.1 (StataCorp 2011).

To understand whether certain symptoms were driving false-positive assignment of deaths of HIV-negative people to the cause group "HIV/AIDS-related", I investigated the symptoms reported in the VA transcripts for deaths assigned to that cause group among HIV-negative and HIV-positive people. For any symptom that occurred in at least 50% of deaths assigned to "HIV/AIDS-related" (an arbitrary cut-off), the chi-squared test was used to determine whether that symptom was associated with assignment to the "HIV/AIDS-related" cause group, among all deaths of HIV-negative and HIV-positive people.

4. Results: Kisesa

I. Assigning physician reviews to cause groups

Across 462 deaths there were 922 reviews (462 by one physician, 460 by the other), and these were assigned to cause groups (Figure 2).

Five reviews (0.5%) had ICD-10 codes and descriptions that were inconsistent for which the ICD-10 code was likely to be a data-entry error and was corrected (Appendix 5). Ten reviews (1.1%) were incomplete and were assigned to a cause group based solely on the ICD-10 code or the description (Appendix 6).

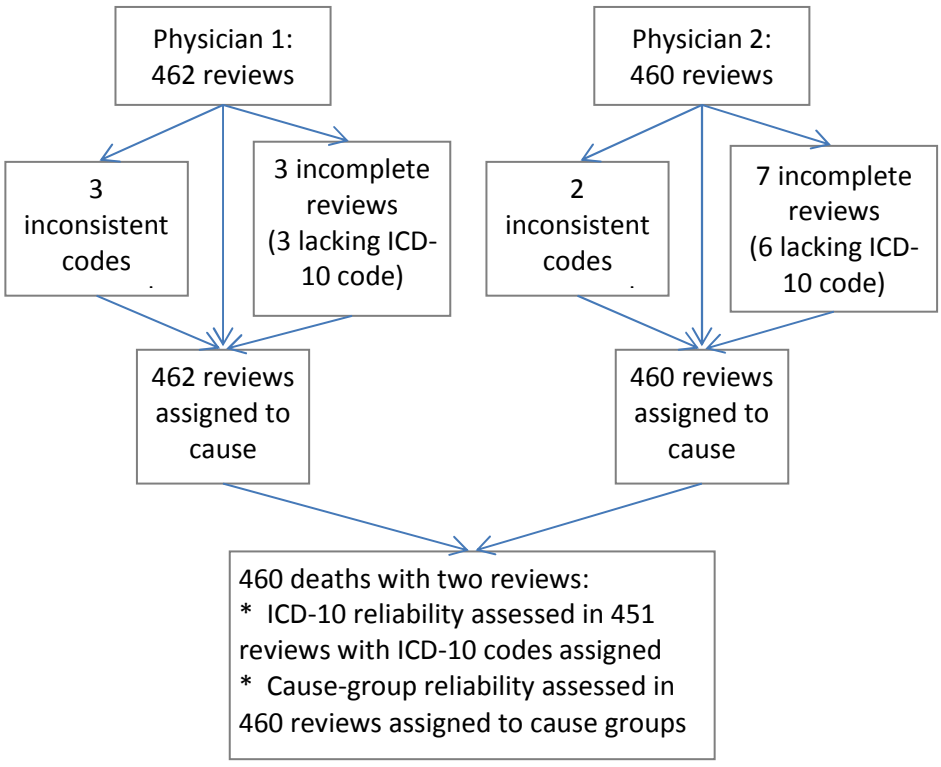

Figure 2: Flow diagram showing physician reviews available for reliability analyses

Overall the 922 reviews contained 176 unique three-character ICD-10 codes assigned by one or other physician. Cross tabulation of these codes is not presented.

Appendix 7 shows the cause groups to which the deaths were assigned by each physician in the 460 records with two reviews, and Figure 3 shows the distribution of deaths between the cause groups as assigned by the two reviewing physicians. The overall cause distributions

were similar, with the leading cause groups assigned by both physicians being “HIV/AIDS-related” and “Other/unspecified NCD” (non-communicable diseases), followed by “Assault”, “Pulmonary tuberculosis”, “Malaria” and “Other/unspecified infectious diseases”.

Figure 4 shows the distribution of broad cause categories assigned, and Table 13 shows the distribution of deaths into broad cause categories by the respective physicians. Perhaps surprisingly, the largest discordance was between non-communicable diseases and infectious diseases: in 38 cases (8.3%), the first physician assigned the broad cause category “Non-communicable diseases” and the second physician assigned an infectious cause (16 “HIV/AIDS” and 22 “Non-HIV infections”); in 22 cases (4.8%), the second physician assigned “Non-communicable diseases” and the first physician assigned an infectious cause (11 “HIV/AIDS” and 11 “Non-HIV infections”).

At least one physician assigned the broad cause category “HIV/AIDS” in 172 deaths. In 122 of these deaths (70.9%), the other physician also assigned “HIV/AIDS” (Figure 5). In a further 50 deaths, one physician review was assigned to “HIV/AIDS”, and the other review to another category: in 16 deaths the second physician assigned HIV/AIDS and the first assigned a non-communicable disease; in 14 deaths the second physician assigned HIV/AIDS and the first assigned a non-HIV infection; in 11 deaths the first physician assigned HIV/AIDS and the second assigned a non-communicable disease; and in five deaths the first physician assigned HIV/AIDS and the second assigned a non-HIV infection.

Table 14 shows the descriptions and ICD-10 codes from the physician reviews for the 50 deaths for which only one of the physicians diagnosed the death as from “HIV/AIDS”. The category “Non-communicable diseases” is disaggregated to highlight the leading causes. Among these 50 reviews, the most prevalent condition was liver diseases (10), followed by neoplasms (nine, including two of the cervix and one of the uterus) and tuberculosis (seven).

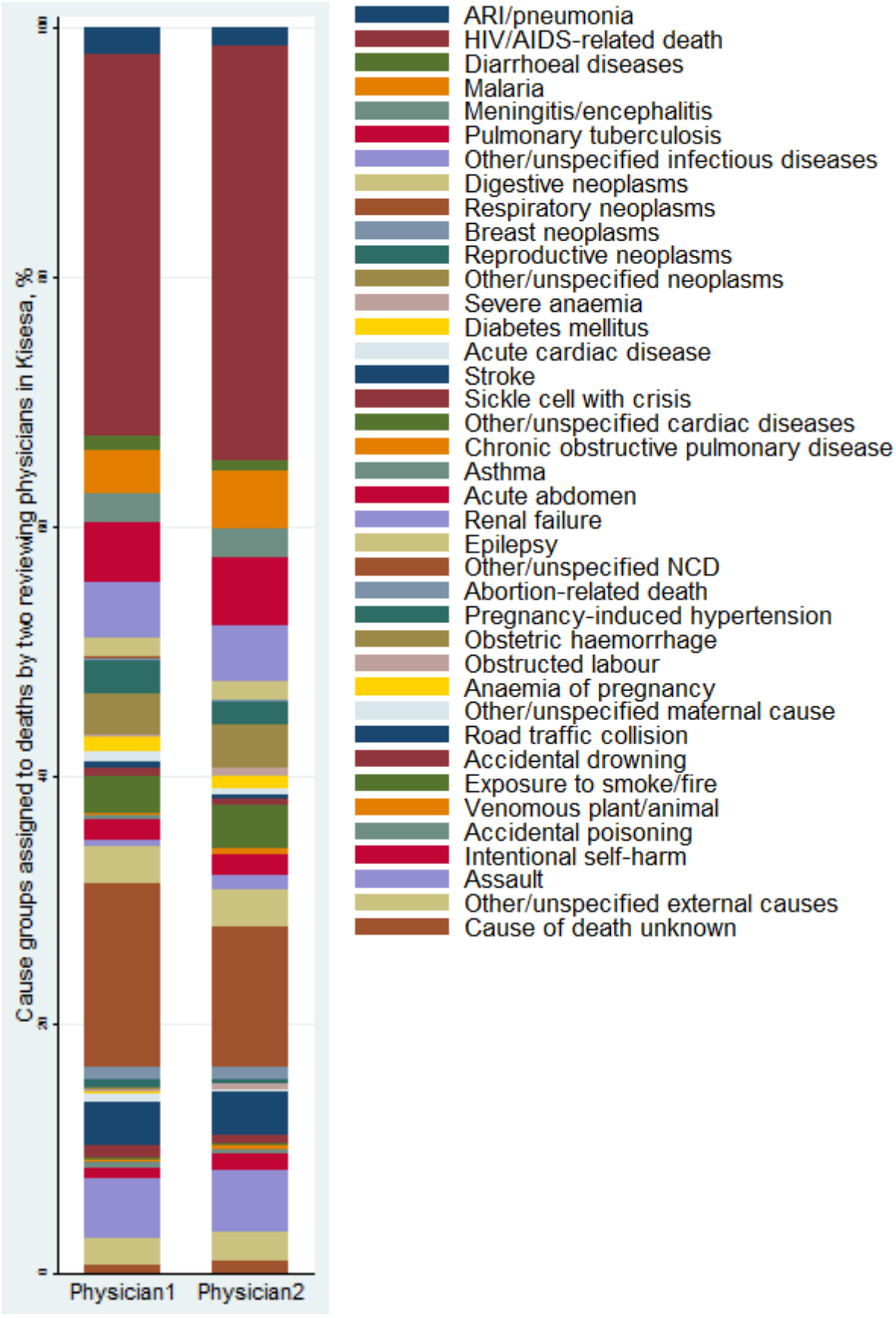

Figure 3: Distribution of cause groups assigned by two reviewing physicians in Kisesa

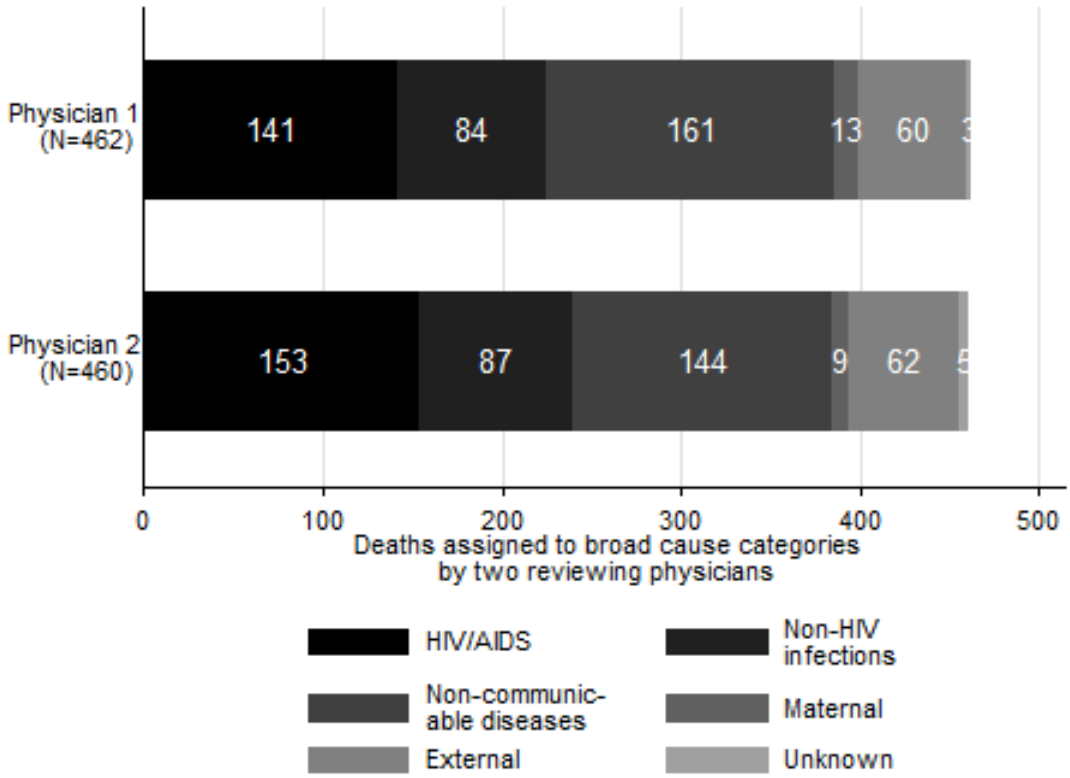

Figure 4: Distribution of respective physician reviews by broad cause category

First physician / Second physician	HIV/AIDS	Non-HIV infectious	Non-communicable diseases	Maternal causes	External causes	Cause of death unknown	Total
HIV/AIDS	122	14	16	1	0	0	153
Non-HIV infectious	5	59	22	1	0	0	87
Non-communicable diseases	11	11	119	1	1	1	144
Maternal causes	0	0	0	9	0	0	9
External causes	2	0	1	0	58	1	62
Cause of death unknown	1	0	1	1	1	1	5
Total	141	84	159	13	60	3	460

Table 13: Classification of 460 deaths in Kisesa by two physicians. NCD=non-communicable disease

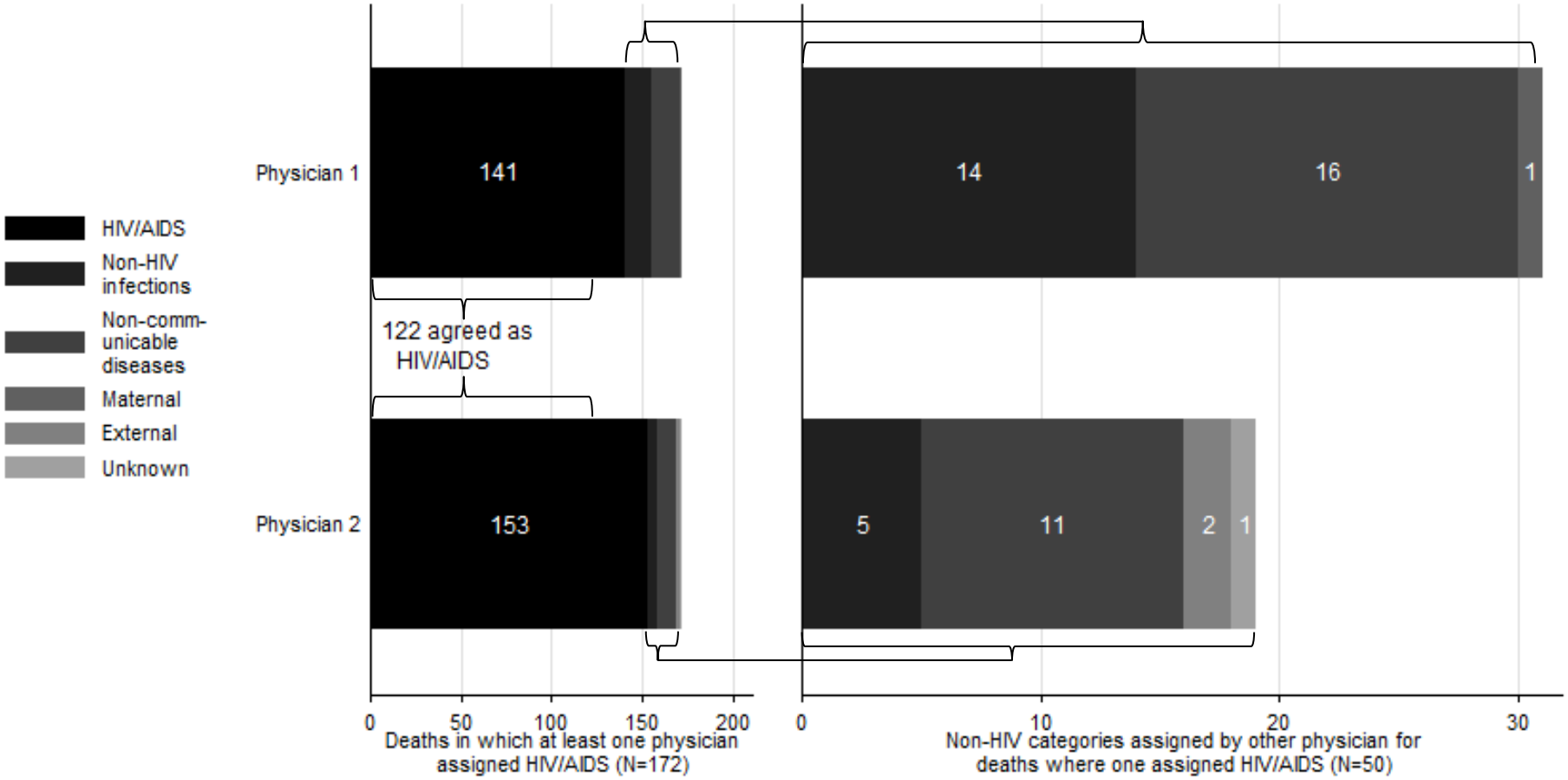

Figure 5: Distribution of physician reviews in 172 deaths where one physician assigned the death to HIV/AIDS across broad cause categories

Non-HIV cause category	Non-HIV causes of death (in black) assigned where only one physician assigned HIV/AIDS (in grey)			
	First physician description and ICD-10 code		Second physician description and ICD-10 code	
Non-HIV infections				
Tuberculosis	Pulmonary tuberculosis*	A16	B20	HIV resulting tuberculosis
	Pulmonary tuberculosis*	A169	B200	HIV disease resulting into tuberculosis
	Pulmonary tuberculosis*	A169	B24	AIDS
	Pulmonary tuberclulsis*	A169	B24	HIV disease
	HIV disease	B20	A169	Tuberculosis*
	HIV resulting into pulmonary TB	B200	A18	Tuberculosis lymphoadenopathy*
	HIV disease	B24	A169	Pulmonary tuberculosis*
Other infections	Chronic lung infection*	J188	B20	HIV AIDS
	Pneumonia*	J189	B24	HIV
	Cutaneous abcess of the trunk*	L02	B207	HIV resulting into bacteria infection
	Multiple abscesses*	L02	B209	HIV disease with unspecified infection
	Multiple abscess*	L02	B209	HIV disease with unspecified in fection
	Pyomyositis*	M600	B207	HIV disease with multiple infection
	Renal disease	N159	B209	HIV disease with unspecified infection
	Viral hepatitis unspecified	B199	B24	HIV disease
	Severe malaria	B54	B24	HIV disease
	Cholera	A009	B24	AIDS
	Tentative slim disease	B222	J988	Tentative chronic lung disease
	HIV resulting into neoplasia	B210	L039	Cellulitis*
Non-communicable diseases (NCD)				
Liver conditions	Liver cirrhosis unspecified	K746	B200	HIV resulting into TB
	Liver disease	K769	B207	HIV disease with multiple infection
	Liver cirrhosis	K746	B227	HIV resulting into multiple infection
	Hepatocellular disease	K769	B20	AIDS
	Chronic liver disease	K769	B24	HIV disease
	Liver disease	K76	B20	HIV disease with unspecified infection
	Chronic liver disease	K769	B24	HIV disese
	AIDS	B24	K769	Liver disease
	HIV disease with unspecified infection	B209	K703	Alcoholic liver disease
	HIV disese	B24	K769	Liver diseses
Neoplasms	Ca cenix	C53	B209	HIV with unspecified infection
	Neoplasia of the utlrus	C559	B238	HIV with unspecified infector
	Malignancy neoplasia involving eye	C69	B24	HIV disease
	Malignant melanonma	C43	B21	HIV disease
	Brain tumor	C71	B24	HIV disease with encephalopathy
	HIV disease with mycosis	B205	C539	Malignant neoplasia of the cervix
	HIV disease with unspecified	B209	C798	Secondary malignant neoplasia of

	infection or par			breast
	HIV disease	B20	C76	Intra abdomina malignancy
	HIV disease resulting into mycobacterial dise	B20	C679	Malignacy neoplasia of urinary bladder
Other NCDs	Chronic respiratory	J989	B200	HIV with mycobacterial disease
	Chronic	J989	B20	HIV disease with unspecified infection
	Chest pain	R07	B24	HIV disease
	Angina pectoria with ischaemic heart disease	I259	B24	HIV disease
	HIV disease with mycobacterial disease	B200	I519	Heart disease
	HIV disease with unspecified infection	B209	N819	Genital prolapse
	HIV disease	B22	G934	Encephalopathy*
	HIV disease resulting into	B205	J449	Obstructive lung disease
External causes	HIV disease resulting into encephalopathy	B22	X70	Self hanging
	AIDS	B24	W20	[Struck by object – no description given]
Maternal causes	Gonodotrophoblastic disease (GTDS)	O019	B20	HIV disease with mycobacterial disease
Cause of death unknown	HIV disease with encephalopathy	B22	R97	Undermined

Table 14: Descriptions and ICD-10 codes from reviews where only one physician assigned HIV, arranged by cause-category of the non-HIV cause. Descriptions as provided. * indicates stage 3/4 HIV disease

II. Assessing the reliability of physician review

Of 462 deaths, 460 (99.6%) received two physician reviews. In nine of these (2.0%), one or the other physician only provided a description and assigned no ICD-10 code. These were excluded from assessment of reliability in assigning ICD-10 codes, which used 451 records. Assessment of reliability in assigning cause groups used 460 records (Figure 2).

In 10 cases (2.2%%) where deaths were assigned to different cause groups by the two physicians, I corrected this discrepant assignment as the reviews were consistent regarding the cause of death (Appendix 8).

Reliability was much lower in assigning three-character ICD-10 codes (agreement 175/451, 38.8%; kappa=0.350) than in assigning cause groups (agreement 326/460, 70.9%; kappa=0.663) (Table 15). Reliability for broad cause categories was higher again (agreement 368/460, 80.0%; kappa=0.729). The kappa values were equivalent to 'fair' agreement at the

level of three-character ICD-10 codes, and 'substantial' agreement at the level of cause group and broad cause categories, on Landis and Koch's scale.

	Level at which agreement is assessed		
	Three-character ICD-10 codes	Cause groups	Broad cause categories
Agreed	38.8% (175/451)	70.9% (326/460)	80.0% (368/460)
Kappa	0.350	0.663	0.729

Table 15: Proportion of deaths assigned to the same cause by two reviewers by the level at which agreement is assessed, and kappa values

I assigned the respective reviews to different cause groups in 134/460 deaths with two reviews (29.1%). A full list of these 134 deaths with the respective descriptions and ICD-10 codes assigned by the two physicians is given in Appendix 9.

Table 16 shows how many deaths were assigned by each physician to the cause group "HIV/AIDS-related" and to all other cause groups. Reliability in assigning HIV/AIDS or any other cause was higher than reliability for assigning overall cause groups: the physicians agreed in 89.1% of deaths (410/460) and Kappa was 0.750, the high end of 'substantial' agreement.

	First physician		
Second physician	HIV/AIDS-related death	Other cause group	Total
HIV/AIDS-related	122	31	153
Other cause group	19	288	307
Total	141	319	460
Agreement = 89.1%, (122+288)/460; Kappa = 0.750			

Table 16: Distribution of deaths assigned by two reviewers to the cause group "HIV/AIDS-related" and to all other cause groups

III. Cause-specific mortality distribution

The cause-specific mortality distributions (CSMD) assigned by the respective physicians were virtually identical (consider Figure 3): only for the cause group "Other/unspecified non-communicable diseases" was the *p*-value of the Z-test for difference in proportions below 0.1 (p=0.084), and there were no differences between the proportions assigned to the broad cause categories (data not shown). To obtain a single CSMD by physician review, I considered the CSMDs assigned using both physician reviews, assigning each death to a single cause group.

There were 15 deaths in which one physician review was assigned to the cause group "HIV/AIDS-related", and the other physician review indicated a condition that defines stage 3/4 HIV disease: the seven tuberculosis cases, seven severe bacterial infections (three abscesses, two pneumonias, one cellulitis and one pyomyositis) and one case of unspecified encephalopathy (Table 14). These 15 deaths were assigned to the cause group "HIV/AIDS-related", meaning the total number of deaths in that cause group, including the 122 deaths with an agreed cause, was 137.

The reliability analysis showed that reliability for HIV/AIDS was higher than for all cause groups, meaning a CSMD excluding the discrepant reviews would give an inflated estimate of the proportion of deaths assigned to HIV/AIDS. In addition to this, the CSMDs of the respective physicians were virtually identical, and the proportion of deaths assigned to HIV/AIDS by each was similar to that by both reviewers including discrepant reviews. For these reasons, further analyses used the CSMDs assigned using both reviews including discrepant reviews as "Cause of death unknown".

Table 17 shows the cause-specific mortality distribution assigned by physician review across 462 deaths in Kisesa. "HIV/AIDS-related" was the leading cause group, assigned to 137 deaths (29.7%). The cause was unknown for 122 deaths (26.4%), almost entirely due to physician disagreement (121/122). The broad cause category "Non-communicable diseases" was next most common (19.0%), with "Other/unspecified non-communicable diseases" the most common cause group within it (6.7%). There were more deaths due to "External causes" (12.3%, primarily the cause groups "Assault" (4.8%) and "Road traffic collision" (3.5%)) than there were due to "Non-HIV infections" (10.8%, primarily "Pulmonary tuberculosis" (3.7%)). "Maternal causes" were assigned in 3.8% of deaths of women (data not shown, 1.7% overall).

	Deaths assigned by physician review	
Cause group	**N**	**%**
HIV/AIDS-related	*137*	*29.7*
Non-HIV infections	*50*	*10.8*
Acute respiratory infections/pneumonia	6	1.3
Diarrhoeal diseases	3	0.6
Malaria	10	2.2
Meningitis/encephalitis	8	1.7
Pulmonary tuberculosis	17	3.7
Other/unspecified infectious diseases	6	1.3
Non-communicable diseases	*88*	*19.0*
Digestive neoplasms	4	0.9
Breast neoplasms	1	0.2
Reproductive neoplasms	7	1.5
Other/unspecified neoplasms	6	1.3
Severe anaemia	1	0.2
Diabetes mellitus	5	1.1
Stroke	2	0.4
Sickle cell with crisis	2	0.4
Other/unspecified cardiac diseases	10	2.2
Acute abdomen	4	0.9
Renal failure	2	0.4
Epilepsy	13	2.8
Other/unspecified non-communicable diseases	31	6.7
Maternal causes	*8*	*1.7*
Abortion-related death	4	0.9
Pregnancy-induced hypertension	2	0.4
Obstructed labour	1	0.2
Other/unspecified maternal causes	1	0.2
External causes	*57*	*12.3*
Road traffic collision	16	3.5
Accidental drowning	3	0.6
Exposure to smoke/fire	1	0.2
Venomous plant/animal	1	0.2
Accidental poisoning	2	0.4
Intentional self-harm	4	0.9
Assault	22	4.8
Other/unspecified external causes	8	1.7
Cause of death unknown	*122*	*26.4*
Total	**462**	**100.0**

Table 17: Distribution of 462 deaths in Kisesa by cause group, as assigned by physician review

Broad cause category	HIV-negative		HIV-positive		Status unknown		Total		Chi² *p*-value HIV-neg v. HIV-pos
	N	%	N	%	N	%	N	%	
HIV/AIDS	18	11.7	59	56.2	60	29.6	137	29.7	<0.001
Non-HIV infections	16	10.4	10	9.5	23	11.3	50	10.8	0.820
Non-communicable diseases	37	24.0	13	12.4	36	17.7	88	19.0	0.015
Maternal causes	4	2.6	0	0.0	5	2.5	8	1.7	0.274*
External causes	26	16.9	5	4.8	25	12.3	57	12.3	0.002
Cause of death unknown	53	34.4	18	17.1	54	26.6	122	26.4	0.003
Total	**154**	**100.0**	**105**	**100.0**	**203**	**100.0**	**462**	**100.0**	

Table 18: Distribution of deaths assigned by physician review in Kisesa, by broad cause category and HIV status. *Fisher's exact test *p*-value

IV. Assessing causes of death against known HIV status

i. Associations between causes of death and HIV status

Of 462 VA records that received physician review, 259 (56.1%) had a linked HIV test result (154 negative, 33.3%; 105 positive, 22.7%). The HIV status of the remaining 203 people who died (43.9%) was unknown. The largest proportion of deaths of HIV-negative people – 34.4% – was those assigned to "Cause of death unknown". The next largest broad cause category was "Non-communicable diseases" (24.0%) followed by "External causes" (16.9%), "HIV/AIDS" (11.7%), "Non-HIV infections" (10.4%) and "Maternal causes" (2.6%, 6.3% in women). Two thirds of deaths of HIV-positive people were assigned to infectious causes (56.2% to "HIV/AIDS", 9.5% to "Non-HIV infections"), with 17.1% being assigned "Cause of death unknown". "Non-communicable diseases" were assigned in 12.4% of deaths, and "External causes" in 4.8%.

Deaths of HIV-negative people were far less likely to be assigned to the broad cause category "HIV/AIDS-related" than deaths of HIV-positive people (11.7% vs 56.2%, $p<0.001$) (Table 18). Deaths of HIV-negative people were more likely than deaths of HIV-positive people to be assigned to the broad cause categories "Non-communicable diseases" (24.0% vs 12.4%, $p=0.015$), "External causes" (16.9% vs 4.8%, $p=0.002$) and "Cause of death unknown" (34.4% vs 17.1%, $p=0.003$). There was no difference between HIV-negative and HIV-positive people in the proportion of deaths assigned to the broad cause categories "Non-HIV infections" (10.4%

vs 9.5%, *p*=0.820) or "Maternal causes" (1.9% vs 0.0%, *p*=0.274), in which there were very few deaths.

For those broad cause categories to which different proportions of HIV-negative and HIV-positive people were assigned, the proportion of people with unknown HIV status assigned to that broad cause category was in all cases between the proportion in the HIV-negatives and that in the HIV-positives.

ii. Specificity

Among deaths of HIV-negative people, 136/154 were assigned to cause groups other than "HIV/AIDS-related": specificity was 88.3% (95% confidence interval: 82.3–92.5%). Allowing reported pre-mortem diagnosis of HIV to serve as an indicator of positive HIV status, five HIV-negative people were reclassified as HIV-positive. These included 4/18 of the previously HIV-negative people assigned to the cause group "HIV/AIDS-related". Following this reclassification, 135/149 HIV-negative people were assigned to non-HIV cause groups, giving specificity of 90.6% for physician review diagnosing HIV/AIDS as a cause of death in adults in Kisesa (95% confidence interval: 84.8–94.3%) (data not shown). Among HIV-negative people for whom "HIV/AIDS-related" was assigned by physician review, the median period between their last HIV test and death was 24 months for those with a report of a pre-mortem diagnosis of HIV, and 25 months for those without (Figure 6).

Specificity by the respective individual physicians, without the reclassification described, was 85.1% and 84.2%, which was insignificantly different from the 88.3% found (*p*=0.259 and *p*=0.299 respectively) (data not shown). Specificity using only those reviews where the two physicians agreed on the broad cause category, excluding those deaths where these were discrepant, was 82.3%, which was also insignificantly different (*p*=0.170) (data not shown).

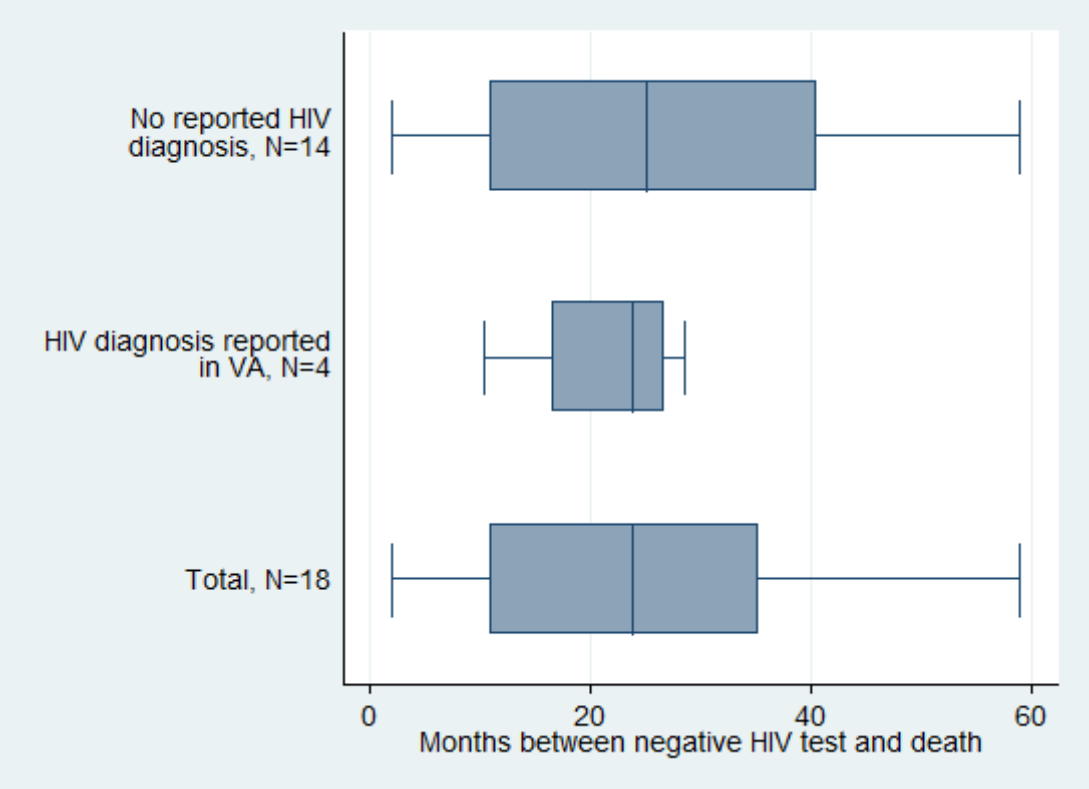

Figure 6: Time from negative HIV test to death for HIV-negative people assigned "HIV/AIDS-related" as cause of death

iii. Sensitivity analysis

Table 19 shows the proportion of deaths of HIV-negative people assigned to cause groups other than "HIV/AIDS-related" according to the length of time they were assumed to be HIV-negative following a negative HIV test result. The differences in this proportion comparing the assumed five-year negative period with alternative periods of one, three and seven years were not significant (p>0.3 in all cases).

Assumed HIV-negative period following negative HIV test	**Number of HIV-negative deaths assigned to cause groups other than "HIV/AIDS-related" (specificity)**		***p*-value compared with an assumed HIV-negative period of five years following a negative HIV test result**
One year	35/41	(85.4%)	0.611
Three years	112/126	(88.9%)	0.879
Five years	136/154	(88.3%)	–
Seven years	148/174	(85.1%)	0.389

Table 19: Specificity of physician review by HIV-negative period following a negative HIV test result

iv. Symptom profile of deaths of HIV-negative people assigned to the cause group "HIV/AIDS-related"

Among 154 deaths of HIV-negative people, 18 (11.7%) were wrongly assigned to the cause group "HIV/AIDS-related". With respect to the diagnosis of HIV-related deaths, these are false positives. Ten symptoms were reported in at least half of these 18 VAs (Table 20), of which eight were associated with false-positive assignment at the 5% level: receipt of treatment for their final illness from a health facility (not necessarily antiretroviral therapy), weight loss, final illness lasting longer than three weeks, chest pain, wasting, anaemia, abdominal pain lasting longer than two weeks and fever lasting longer than two weeks or of unknown duration. An additional three symptoms occurred in at least half of deaths of HIV-positive people assigned to the cause group "HIV/AIDS-related", all of which were strongly associated with being assigned to that cause group: swollen mouth, oral candidiasis and productive cough. The symptoms "swollen mouth" and "oral candidiasis" were interchangeable: all cases of oral candidiasis also reported swollen mouth, and only two cases of swollen mouth occurred without a report of oral candidiasis. The symptoms that occurred frequently in deaths assigned to "HIV/AIDS-related" occurred with similar frequency among HIV-negative and HIV-positive people, and with very similar strength of association with whether HIV/AIDS was assigned as the cause of death. Frequently reported symptoms often occurred in combination in the false-positive cases: 14/18 had at least five of the eight associated symptoms that occurred in at least half of false-positive cases (Table 21).

5. Summary

Overall, the results from Kisesa show specificity for HIV/AIDS of 88%, and greater reliability for HIV/AIDS than for cause groups overall. Over one quarter of deaths were assigned to "Cause of death unknown", rising to over one third among HIV-negative people. The leading cause of death was HIV/AIDS, accounting for 30% of all deaths and 56% of deaths of HIV-positive people. Several symptoms were associated with the assignment of "HIV/AIDS-related" as the cause of death among HIV-negative people.

	154 HIV-negative people					105 HIV-positive people				
	Assigned HIV		Not assigned HIV		Chi² *p*-value for association with	Assigned HIV		Not assigned HIV		Chi² *p*-value for association with
Symptom	N	%	N	%	**assignment of HIV**	N	%	N	%	**assignment of HIV**
Symptoms occurring in ≥50% of HIV-negative deaths assigned to "HIV/AIDS-related"										
Received treatment	18/18	100.0	98/136	72.1	0.010	58/59	98.3	40/46	87.0	0.021
Weight loss	16/18	88.9	36/136	26.5	<0.001	55/59	93.2	18/46	39.1	<0.001
Final illness lasted ≥3 weeks	14/18	77.8	54/136	39.7	0.002	47/59	79.7	18/46	39.1	<0.001
Chest pain	12/18	66.7	32/136	23.5	<0.001	40/59	67.8	11/46	23.9	<0.001
Wasting	12/18	66.7	18/136	13.2	<0.001	38/59	64.4	10/46	21.7	<0.001
Anaemia	12/18	66.7	29/136	21.3	<0.001	34/59	57.6	9/46	19.6	<0.001
Fever >2 weeks/unknown duration	11/18	61.1	16/136	11.8	<0.001	46/59	78.0	9/46	19.6	<0.001
Abdominal pain >2 weeks	10/18	55.6	24/136	17.6	<0.001	33/59	55.9	8/46	17.4	<0.001
Headache	10/18	55.6	45/136	33.1	0.062	29/59	49.2	18/46	39.1	0.306
Used alcohol	9/18	50.0	49/136	36.0	0.250	28/59	47.5	19/46	41.3	0.529
Symptoms additionally occurring in ≥50% of HIV-positive deaths assigned to "HIV/AIDS-related										
Swollen mouth	8/18	44.4	5/136	3.7	<0.001	40/59	67.8	2/46	4.3	<0.001
Oral candidiasis	8/18	44.4	5/136	3.7	<0.001	40/59	67.8	2/46	4.3	<0.001
Productive cough	6/18	33.3	16/136	11.8	0.014	35/59	59.3	7/46	15.2	<0.001

Table 20: Symptoms reported in ≥50% of deaths in Kisesa assigned by physician review as HIV/AIDS-related among HIV-negative and HIV-positive people, by whether deaths were assigned to the cause group "HIV/AIDS-related"

Symptom	Distribution of symptoms associated with false-positive HIV diagnosis occurring in at least half of 18 deaths of HIV-negative people assigned to the cause group "HIV/AIDS-related"																	
Received treatment	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•
Weight loss	•	•	•	•	•	•	•	•	•		•		•	•	•	•	•	•
Final illness lasted ≥3 weeks	•	•	•	•	•	•	•	•	•	•	•	•		•	•			
Chest pain	•	•	•	•	•	•	•		•	•		•	•				•	
Wasting	•	•	•	•	•	•	•	•			•		•	•		•		
Anaemia	•	•	•	•	•		•	•		•	•	•	•	•				
Abdominal pain >2 weeks	•	•	•		•	•		•	•	•	•				•			
Fever >2 weeks/unknown duration	•	•	•	•		•	•	•	•	•		•						
Total number of symptoms	**8**	**8**	**8**	**7**	**7**	**7**	**7**	**7**	**6**	**6**	**6**	**5**	**5**	**5**	**4**	**3**	**3**	**2**

Table 21: Symptoms reported for ≥50% of HIV-negative people assigned as "HIV/AIDS-related" and associated with false-positive assignment

6. Discussion

The present analyses give an indication of the main causes of death in the population under surveillance, and relate those causes of death to HIV status. The results also offer some indication of the performance of physician review for assigning HIV/AIDS as a cause of death. Specificity of physician review for HIV/AIDS was moderately high (88%). Reliability of physician review was moderately high for HIV/AIDS (89% agreement, kappa=0.75), although it was lower for all causes of death (71%, 0.66). No clear symptom pattern existed driving the false-positive assignment of deaths to HIV/AIDS. HIV/AIDS was the leading cause of death in Kisesa, followed by non-communicable causes, external causes and non-HIV infections.

I. Findings of other studies

Only two other studies, using data from a cohort in Masaka, south-west Uganda, have assessed the specificity of physician review for HIV/AIDS against known HIV status alone[19, 96]. These reported specificities of 92% and 90% among deaths of people aged 13 and older, in 1990–1993 and 2006–2008 respectively. These values are within the 95% confidence intervals of the present estimate for Kisesa. However, when restricted to a more comparable range of ages, the specificities in those studies are lower than in the present work (85% among 13–44 year-olds who died in 1990–1993, and 82% among 13–64 year-olds who died in 2006–2008). The other pertinent difference was that the study using the deaths from 2006–2008 assessed the specificity of physicians' views on whether the deceased was HIV-positive, rather than whether they died of HIV/AIDS. Unfortunately, the authors did not report the causes assigned to those deaths.

Other estimates of the specificity of physician review for HIV/AIDS have used different reference standards. The present values are at the lower end of the specificities reported for diagnosing "TB/AIDS" in Tanzania, Ethiopia and Ghana (89%–100%), where the reference standard was hospital records supplemented in some cases by postmortem or laboratory findings[97]; the confidence intervals in the present study contain the 90% minimum threshold for specificity suggested by those authors. The present specificities are slightly lower than the 94% reported for "TB/AIDS" where the reference diagnoses were "obtained from a combination of hospital records and death certificates by one of the authors" but criteria for diagnosing "TB/AIDS" were not stated (Boulle et al 2001: 516[98, 120]). Setel and colleagues[100] reported specificity of physician review for HIV/AIDS of 86% (95%CI: 84%–88%) validated

against medical records. Bauni and colleagues[101] reported physician-review specificity of 94% (95%CI: 88%–98%) for "HIV/AIDS-related death", validated against hospital records. Araya and colleagues[174] used a strict reference standard of HIV-positive status and clinical records showing AIDS-defining conditions, and reported a lower specificity of 78% (95%CI: 65%–87%). A stricter reference standard means that more deaths of HIV-positive people will be included in the denominator for specificity. Coding physicians may nonetheless be more likely to assign HIV as the cause of these deaths than they would for those deaths that would have been in that denominator using a more permissive reference standard for positive cases (such as HIV-positive status, or HIV/AIDS-related disease not meeting the level of an AIDS-defining condition). Nonetheless, Murray and colleagues used a strict reference standard of medical records showing AIDS, and reported specificity of physician review of 97%[121].

The cause-specific reliability of physician review is rarely reported, but there are two comparable studies: the 89% agreement in diagnosing HIV/AIDS found in Kisesa is similar to the 91% agreement reported by Kamali and colleagues[96] and higher than the 80% agreement reported by Mayanja and colleagues[19]. All-cause agreement between physicians at the level of cause groups was 71%. This is below the 79% reported for all causes of death in Ghana[179], and at the lower end of the range reported in Tanzania, Ethiopia and Ghana (70%–78%)[97]; both those studies used similarly detailed cause-of-death groupings to the cause groups used in the present study[97, 179]. The present estimate is similar to the 69% achieved by physician review for deaths at all ages in rural Ethiopia[64], although the detail in the cause-of-death groupings used in that study in unclear. Joshi and colleagues reported reliability of 94% between physicians coding to the broader level of ICD-10 chapter heading[173], which is not comparable with the present results. Similarly, those authors reported an exceptionally high kappa value of 0.96 (95%CI: 90–98) for inter-physician reliability in assigning deaths to the category "infectious diseases", which is not comparable to the present estimate for HIV/AIDS. No other studies have reported kappa values for assigning HIV/AIDS by physician review. Similarity of the CSMDs assigned by respective physicians, despite less-than-perfect reliability with regard to individual deaths, has also been seen in South Africa[137].

II. Discrepant reviews

The physicians in Kisesa disagreed on whether 50 of the deaths were due to HIV/AIDS, among which 15 of the non-HIV/AIDS diagnoses were AIDS-defining conditions. In many of the remaining 35 deaths, the non-HIV/AIDS diagnosis was of a condition that often entails some

degree of wasting, such as liver disease or neoplasm[180]. Non-communicable diseases and HIV/AIDS also tend to cause illness of longer duration than other infections or external causes[18].

The absence of reconciliation of discordant physician reviews in Kisesa meant that deaths with an unknown cause were more common than they would probably otherwise have been (26%); this large proportion with unknown cause gives reason for caution in interpreting the proportions assigned to named causes. Had a single physician review been used in Kisesa, the proportion of deaths with unknown cause would have been 0.7% and 1.1% by the two respective physicians (data not shown).

III. Proportion of deaths due to HIV/AIDS

It is notable that the present estimate of the proportion of adult deaths due to HIV/AIDS in Kisesa, using physician review of deaths largely from 2004–2011, is insignificantly different from the estimate for 1994–1996 made using similar methods (30% vs 35%, p=0.257)[18]. Conversely, the proportion of deaths assigned to non-HIV infections was much lower in the present analysis than in the previous analysis (11% vs 24%, p<0.001). However, the present proportion assigned an unknown cause is also much larger (26% vs 15%, p=0.006), which should prompt caution in making comparisons, as the cause-mix of deaths with and without a named cause assigned through physician review of VA could differ systematically.

Reports of deaths of adults in the general population in sub-Saharan Africa, disaggregated by cause, are scarce. A study of a rural cohort in Karonga, Malawi in 2002–2006[142] reported a greater proportion of deaths due to HIV/AIDS (40%) than in our populations, which is unsurprising given the greater burden of HIV in Malawi compared to Tanzania[181]. That study also reported a greater proportion of deaths due to non-communicable diseases (28%) than seen in Kisesa, and fewer deaths with unknown cause (11%). A study in three sites in Tanzania in 1992–1995[182] reported from 22% to 40% of deaths due to HIV – consistent with the present findings – and from 50% to 67% of deaths overall due to infectious causes – higher than the present findings. Non-communicable diseases were responsible for 15–29% of adult deaths in that study, consistent with the present estimate for Kisesa.

IV. Cause-specific mortality by HIV status

Assignment to the broad cause category "HIV/AIDS" was significantly higher among HIV-positive people, as would be hoped. Assignment to non-HIV/AIDS, non-infectious broad cause categories was significantly higher for HIV-negative people, with the exception of "Maternal causes" in Kisesa (where there were very few deaths). The proportion assigned to "Non-HIV infections" was the same among HIV-negative and HIV-positive people. The absence of elevated mortality from non-HIV infections among HIV-negative compared with HIV-positive people may suggest a role for HIV infection in causing infectious diseases that are not assigned to "HIV/AIDS" by physician review.

Among HIV-negative people in Kisesa, over half of deaths assigned to infectious causes were assigned to "HIV/AIDS". This raises the suspicion that reviewing physicians in Kisesa may tend to assign HIV/AIDS as the cause of death in the presence of symptoms of infection without adequately considering whether HIV/AIDS is truly indicated. Concern that reviewing physicians are potentially biased toward expected causes of death in their area has been raised before[58, 65].

V. False-positive symptoms

No clear pattern of symptoms was visible in the false-positive diagnosis of HIV/AIDS for HIV-negative people. The only symptoms common to false-positive assignment in both sites were fever and a final illness lasting more than three weeks. Weight loss and wasting are classic symptoms of HIV-related disease[36], and were prominent among false-positives in Kisesa. This somewhat contradicts the findings of Lopman and colleagues[87], who found weight loss and wasting to have specificities, respectively, of 99% and 95% in 15–59 year-olds in Manicaland, Zimbabwe[87]. Literature on the occurrence of specific symptoms in false-positive diagnoses of HIV/AIDS is scarce. The analysis of symptoms that occur in false-positive cases suggests a strong influence of the occurrence of common symptoms in combination with one another – this is perhaps unsurprising, given the tendency of physicians to assign a diagnosis that explains as many symptoms as possible[183]. The analysis of false-positive symptoms shows that among the symptoms associated with people being assigned to the cause group "HIV/AIDS-related", all symptoms occurring among HIV-negative people also occur frequently among HIV-positive people – these data do not suggest any "rogue" symptoms driving false-positive assignment of HIV/AIDS. The symptoms that were associated with false-positive HIV diagnosis

in both sites were several common constitutional symptoms of infection – fever, cough and diarrhea – as part of a lengthy final illness.

VI. Limitations

The main limitation to the present results arises from the quality of the data. Because no reconciliation between discrepant physician reviews occurs in this study, these data are not typical of physician-review data using multiple reviewers; more deaths have unknown cause than might be the case had there been reconciliation. I decided to conduct the analyses by HIV-status using the CSMD in broad cause categories assigned using both reviews, with the discrepant cases classified as "Cause of death unknown". As noted, this was in order to avoid inflating the proportion of deaths assigned to HIV/AIDS, as the higher agreement between physicians on HIV/AIDS as a cause compared to other causes meant probable selection bias for other causes out of the group of deaths with agreed cause. One corollary of this decision was that the proportions of deaths in the categories "Non-HIV infections" and "Non-communicable diseases" were significantly lower than as assigned by either individual physician (data not shown), while the proportion assigned to "Cause of death unknown" was, predictably, higher. A further corollary was that specificity for physician review was higher than it would have been excluding the discrepant cases, or for either individual physician.

Assigning deaths to the cause groups using ICD-10 codes assigned by the reviewing physician introduced a slight limitation: it was possible for reviews that agreed on the cause of death to be assigned to different cause groups and vice versa, due to the way ICD-10 codes are grouped in the original publication[171]. I minimised underestimation of agreement by correcting cases where the physician reviews were assigned to different cause groups despite physicians agreeing on the cause of death. But it is possible that there is minor overestimation of reliability, as dissimilar diagnoses are classified as agreed if both physicians assign ICD-10 codes that appear in the same "Other/unspecified" group. It is a limitation of this investigation of physician agreement that because it is not entered electronically with the rest of the VA data, I could not use the narrative section of the VA interview. The narrative section is often viewed as key to determination of cause of death by physician review[65], and its absence from datasets used for analysis of physician-review data is a limitation that others have encountered[120].

The proportion of deaths that are assigned unknown cause in VA studies using physician review varies, with physicians in some studies unable to assign a named cause to over 20% of deaths, as in the present analysis in Kisesa[184-186], while some studies achieve an intermediate 10–20%[49, 142], and some studies achieve <10% of deaths with unknown cause[30, 34, 144, 187, 188]. The high proportion of deaths in cause-of-death studies without a cause assigned, using both VA data and vital registration data, has been a concern since at least the mid-1990s[189].

VII. Conclusion

The present specificity achieved by physician review was in the middle of the distribution of specificity for HIV/AIDS reported in the literature. It was similar to other reports of the specificity of physician review, and compared favourably to reported specificities of other methods of interpreting VA data validated against the same and different reference standards. All-cause agreement between physicians in Kisesa was at the lower end of the range reported in the literature, though by no means outlying. Physician agreement on HIV/AIDS was higher, and respectively very similar to and much higher than the two comparable estimates.

The distribution of causes of death assigned by the reviewing physicians for HIV-negative and HIV-positive people, respectively, was plausible, with the distributions for people of unknown HIV status located between those for HIV-negatives and HIV-positives. Symptoms occurring among HIV-negative people assigned HIV/AIDS as cause of death suggest that occurrence of multiple symptoms is important, but there is little literature in which to situate this finding. In terms of its quality as a method of interpreting VA data, these analyses suggest that physician review should remain a candidate alongside newer methods.

4. Causes of death by InterVA

1. Introduction

InterVA is an algorithm used to interpret verbal autopsy data, in order to estimate the cause-specific distribution of deaths at population level[5]. It uses Bayesian statistics to estimate the probability of causes of death from indicators comprising symptoms suffered prior to death, and demographic and circumstantial information. InterVA considers each death in a population independently of each other death. For each death of an adult, it assigns a probability that the death was due to each of a set of causes that account for almost all deaths in low- and middle-income countries. Across all causes these probabilities sum to one, and the probability of any given cause may be zero. The proportion of deaths due to each cause in the population is the summed total of the probabilities of each individual death being due to that cause. The current version, InterVA-4, has been developed to be used alongside the WHO recommended VA questionnaire[171], and the causes for which it assigns probabilities are the same as those that comprise the cause groups described in Data Management and Analyses and used in this thesis.

InterVA has been developed over the last decade in several iterations[80, 114, 115], with substantial differences in performance between InterVA-4 and previous versions that have been widely used[80]. It has been used in diverse settings around the world to estimate all-cause mortality in people of all ages[32, 64, 88, 93, 117, 118], in datasets of adults only[101, 116], in women aged 15–49[190, 191], in neonates[192] and in people aged over 65[193]. It has also been used to investigate individual causes of mortality in depth, including HIV[137], tuberculosis[119] and malaria[194].

I. Development and working of InterVA

InterVA is based on conditional probability, as expounded in Bayes's theorem[114]. Its use for derivation of causes of death from verbal autopsy data is based on the proposition that a) there is a relationship between causes of death and the indicators that can be observed prior to death; and b) the likelihood of such indicators being experienced by someone suffering a given cause of death can be estimated. The basic form of Bayes's theorem for any given cause C_i and indicator I_j is:

$$P(C_i|I_j) = \frac{P(I_j|C_i) \times P(C_i)}{P(I_j|C_i) \times P(C_i) + P(I_j|!\,C_i) \times P(!C_i)} \quad (1)$$

[5] InterVA can be downloaded from www.interva.net.

Where $P(!\,C_i)$ is the probability of not-C_i and is equal to $\left(1 - P(C_i)\right)$,

InterVA is based on two sets of estimated values:

- A set of prior conditional probabilities describing the likelihood of indicators occurring prior to death for each given cause of death. These take the form $P\left(I_j \middle| C_i\right)$, and form an $n \times m$ matrix for values I_1 to I_n and C_1 to C_m.

- A set of prior unconditional probabilities for each cause of death in the population – that is, the approximate proportion of all deaths in the population that are due to each cause, where the population is all deaths across low- and middle-income countries (Peter Byass, personal communication). These prior unconditional probabilities can be thought of as having the form $(C_i|I_0)$ [80].

Both these sets of prior probabilities are based on the expert consensus of experienced physicians working in a range of low- and middle-income countries and clinical specialisms, according to a qualitative scale (Table 22). The prior probabilities were initially decided by a panel of five physicians[115], and subsequently revisited by further panels of physicians[80].

Label	Value	Interpretation
I	1.0	Always
A+	0.8	Almost always
A	0.5	Common
A–	0.2	
B+	0.1	Often
B	0.05	
B–	0.02	
C+	0.01	Unusual
C	0.005	
C–	0.002	
D+	0.001	Rare
D	0.0005	
D–	0.0001	
E	0.00001	Hardly ever
N	0	Never

Table 22: Qualitative probability scale used for eliciting expert opinions on probabilities (from Byass et al 2012[80])

The starting distribution of the probabilities of each cause is equal to the prior unconditional probability distribution. The processing of the indicators generally makes the probabilities of most causes decline, and increases the probabilities of the causes most associated with the reported indicators[80].

For the first indicator, I_1, the probability of a given cause C_i from the range of causes of death C_1 to C_m in the presence of that indicator is:

$$P(C_i|I_1) = \frac{P(I_1|C_i)\times P(C_i|I_0)}{\sum_{j=1}^{j=m}\left[P\left(I_1\middle|C_j\right)\times P\left(C_j\middle|I_0\right)\right]} \quad (2)$$

where the denominator is a normalisation that approximates the denominator from (1) for all causes of death C_1 to C_m and ensures the summed probabilities of all causes equal one[80]. This normalisation occurs after each indicator is processed. For indicators I_1 to I_n, InterVA takes the form:

$$P(C_i|I_1, I_2, \dots, I_n) = \frac{P(I_n|C_i)\times P(C_i|I_0, I_1, \dots, I_{n-1})}{\sum_{j=1}^{j=m}\left[P\left(I_n\middle|C_j\right)\times P\left(C_j\middle|I_0, I_1, \dots, I_{n-1}\right)\right]} \quad (3)$$

InterVA produces a set of probabilities for each death independently of each other, using the indicators reported in the VA data. The user must set the level of malaria and HIV as 'high', 'low' or 'very low' to reflect local prevalence (InterVA-4 user guide[81]), which changes the prior probabilities for those causes of death. This recognises the potential for misclassification which is so important in interpreting VA data, acting as an aid to distinguishing similar presentations with different causes; the authors of InterVA have described it as "analogous to a coding physician knowing that HIV or malaria represent more-common or less-common public health problems in a particular population" (Byass et al 2011: 3)[137].

Once all indicators have been processed, InterVA outputs the probability of up to three causes for each death. InterVA assigns the cause of death as "indeterminate" if VA records have "insufficient VA data to decisively determine the cause probabilities" (Fottrell et al 2011: 5[195]), and in InterVA-4 the threshold for an indeterminate cause of death is one where the probability of the most likely cause is below 0.4; second- and third-most-likely causes of death are reported where their probabilities are at least equal to half the probability of the most likely cause[80]. Used as intended, InterVA is for population-level cause-of-death estimation only, and is not meant to be used to estimate causes of individual deaths. The combined probability of the causes for which InterVA assigns a probability does not always sum to 100%. The residual probabilities – the difference between the sum of all assigned probabilities and 100% – are accessible in the output; the authors of InterVA encourage users to treat these residual probabilities as additional "indeterminate" fractions when calculating population mortality distributions (InterVA-4 User Guide 2012: 7[81]).

To run, InterVA requires a VA record to contain data on age, sex and at least one indicator. Records not meeting these requirements are not processed at all, do not receive a cause of death and do not contribute to the denominator.

One of the causes of death is "HIV/AIDS-related death". The indicators given a conditional probability for HIV-related death $P(I_k|C_{HIV}) \geq 2\%$ are given in
Appendix 10.

II. Validation of InterVA

Attempts have been made to validate InterVA, primarily using medical records as a reference standard[93, 101, 116] though also using physician review results to validate[136]. Metrics used for validation have primarily been sensitivity and specificity[101, 116, 119, 136] as well as the kappa statistic to assess agreement with medical records[116]. One study[93] has used the "robust" metrics proposed by Murray and colleagues[129].

For calculating measures of validity, most studies treat individual deaths as due to the most likely cause assigned by InterVA[93, 116, 136], though this is not always clear[101]. Tensou and colleagues proposed a method of allowing variable specificity by using a cut-off for the likelihood of the most-likely cause of death as the means of deciding whether a death was assigned to that cause[116].

To enable assessment of the validity of InterVA, one study created a reference standard using records from hospital admissions and known HIV status, and defined deaths as "AIDS-related" if the deceased was HIV-positive and an opportunistic infection was indicated in their admission record[116]. The authors used receiver-operator characteristic (ROC) curves to investigate which cut-off for the likelihood assigned by InterVA optimally identified AIDS deaths. Another study defined as HIV-positive anyone with a positive HIV test result or a report of a diagnosis of HIV prior to death made in the VA interview[102].

i. Non-validation comparison with other methods

Several studies have compared the findings of InterVA with other methods, while not regarding this as a validation exercise. InterVA has primarily been compared with physician review[32, 64, 88, 114, 118, 119, 136, 190, 192], though one study has compared InterVA to another proposed automated method, the "simplified symptom pattern" method[93]. The measures used to compare methods have primarily been to simply report percentages of deaths assigned to

given causes by the respective methods, percentage agreement between methods, or kappa statistics, though other methods have also been used[88].

2. Objectives

I. Overall objective

To assess the specificity of InterVA for the assignment of HIV/AIDS as cause of death from verbal autopsy data.

II. Specific objectives

1. To describe the cause-specific mortality distributions in the populations under investigation.
2. To use known HIV status to:
    a. calculate the specificity of InterVA for assigning HIV/AIDS as cause of death; and
    b. investigate causes of death by HIV status.

3. Methods

I. Data conversion

To attain the results of the InterVA algorithm, I converted the VA data from Spec 8.1 into the InterVA input format. The indicators in the InterVA input format are very similar to the items in Spec 8.1, as both were modelled on the WHO recommended VA items[80, 171]. The InterVA input specification is detailed in the InterVA User Guide[81].

II. Assessing causes of death against known HIV status

These analyses followed the same structure as was used in the Physician Review chapter. Proportions of deaths assigned to cause groups and broad cause categories were calculated

using summed fractional probabilities, in line with the intended use of InterVA. Likewise, specificity was calculated as the summed fractional probabilities assigned to non-HIV cause groups among the deaths of HIV-negative people. The significance of differences between the proportion of deaths assigned to the broad cause categories among HIV-negative and HIV-positive people was assessed using the Z test – the chi-squared test cannot be used as the summed fractional probabilities do not offer integer values for the number of deaths in a category and therefore do not allow the calculation of a chi-squared statistic.

To investigate the symptoms associated with assignment of deaths of HIV-negative people to HIV/AIDS, it was necessary to use individual deaths rather than summed fractional probabilities: this analysis was conducted on all deaths of HIV-negative people for which "HIV/AIDS-related" was the most likely cause assigned by InterVA, with no minimum likelihood cut-off used.

4. Results: Kisesa

I. Cause-specific mortality distribution

InterVA assigned causes to 1107 deaths from Kisesa. HIV/AIDS was the leading cause of death, assigned as the cause of 34.4% of deaths (Table 23). The next most common cause groups were pulmonary tuberculosis (11.0%), acute abdomen (6.2%) and digestive neoplasms (4.7%). InterVA assigned nearly half of all deaths in Kisesa to HIV/AIDS or tuberculosis.

Other infectious causes were assigned to a further 9.5% of deaths, meaning overall infectious diseases were assigned to 54.9% of deaths overall. Non-communicable diseases were assigned to 23.1% of deaths – primarily acute abdomen (6.2%) and digestive neoplasms (4.7%) – followed by cause of death unknown (10.0%), injuries (7.5%) and maternal conditions (4.5% overall, 9.1% in women).

II. Assessing causes of death against known HIV status

i. Associations between causes of death and HIV status

Of the 1107 VA records that InterVA assigned a cause of death, 291 (26.3%) were HIV-negative within five years of death, and 250 (22.6%) were HIV-positive. The remaining 566 (51.1%) had unknown HIV status. The largest proportion of deaths of HIV-negative people was assigned to the broad cause category "Non-communicable diseases" (31.4%) (Table 24). This was followed by deaths due to "Non-HIV infections" (20.4%), "HIV/AIDS" (18.6%), "External causes" (11.5%), "Cause of death unknown" (11.0%) and "Maternal causes" (7.2%). Over three quarters of deaths of HIV-positive people were assigned to infectious causes (57.6% to "HIV/AIDS" and 19.8% to "Non-HIV infections"), followed by "Non-communicable diseases" (13.3%), "Cause of death unknown" (6.4%), "External causes" (0.4%) and "Maternal causes" (2.5%).

Deaths of HIV-negative people were far less likely to be assigned to the broad cause category "HIV/AIDS-related" than deaths of HIV-positive people (18.6% vs 57.6%, $p<0.001$) (Table 24). Deaths of HIV-negative people were more likely than deaths of HIV-positive people to be assigned to the broad cause categories "Non-communicable diseases" (31.4% vs 13.3%, $p<0.001$), "External causes" (11.5% vs 0.4%, $p<0.001$) and "Maternal causes" (7.2% vs 2.5%,

p=0.013). There was a smaller difference between HIV-negative and HIV-positive people in the proportion of deaths assigned to the broad cause category “Cause of death unknown” (11.0% vs 6.4%, *p*=0.060), and no difference in the proportion assigned to “Non-HIV infections” (20.3% vs 19.8%, *p*=0.874).

For those broad cause categories to which different proportions of HIV-negative and HIV-positive people were assigned, apart from “Cause of death unknown”, the proportion of people with unknown HIV status assigned to that broad cause category was between the proportion of HIV-negative people and that of HIV-positive people.

Cause group	% deaths assigned by InterVA
HIV/AIDS-related	*34.4*
Non-HIV infections	*20.5*
Acute respiratory infection/pneumonia	1.8
Diarrhoeal diseases	1.5
Malaria	1.8
Meningitis and encephalitis	2.8
Pulmonary tuberculosis	11.0
Other/unspecified infectious diseases	1.6
Non-communicable diseases	*23.1*
Oral neoplasms	2.1
Digestive neoplasms	4.7
Respiratory neoplasms	0.4
Breast neoplasms	0.2
Reproductive neoplasms	2.2
Other and unspecified neoplasms	0.4
Severe malnutrition	0.4
Diabetes mellitus	1.4
Acute cardiac disease	0.5
Stroke	0.5
Other/unspecified cardiac diseases	1.6
Chronic obstructive pulmonary disease	0.2
Acute abdomen	6.2
Liver cirrhosis	0.7
Renal failure	0.5
Epilepsy	1.0
Maternal conditions	*4.5*
Ectopic pregnancy	0.3
Abortion-related death	0.7
Pregnancy-induced hypertension	0.8
Obstetric haemorrhage	1.7
Obstructed labour	0.1
Pregnancy-related sepsis	0.5
Anaemia of pregnancy	0.1
Other/unspecified maternal causes	0.3
Injuries and external causes	*7.5*
Road traffic collision	1.8
Accidental fall	0.1
Accidental drowning	0.8
Exposure to smoke/fire	0.1
Venomous plant/animal	0.2
Accidental poisoning	0.3
Intentional self-harm	0.8
Assault	3.3
Other/ unspecified external causes	0.2
Cause of death unknown	*10.0*
Total	**100.0**

Table 23: Distribution of causes of death assigned by InterVA in Kisesa

Broad cause category	HIV-negative	HIV-positive	Status unknown	Total	Z test *p*-value, HIV-neg v. HIV-pos
HIV/AIDS	18.6	57.6	32.3	34.4	<0.001
Non-HIV infections	20.3	19.8	20.9	20.5	0.874
Non-communicable diseases	31.4	13.3	23.2	23.1	<0.001
Maternal causes	7.2	2.5	4.0	4.5	0.013
External causes	11.5	0.4	8.6	7.5	<0.001
Cause of death unknown	11.0	6.4	11.0	10.0	0.060
Total	**100.0**	**100.0**	**100.0**	**100.0**	
Specificity	**81.4**				

Table 24: Distribution of deaths in Kisesa assigned by InterVA, by broad cause category and HIV status, %

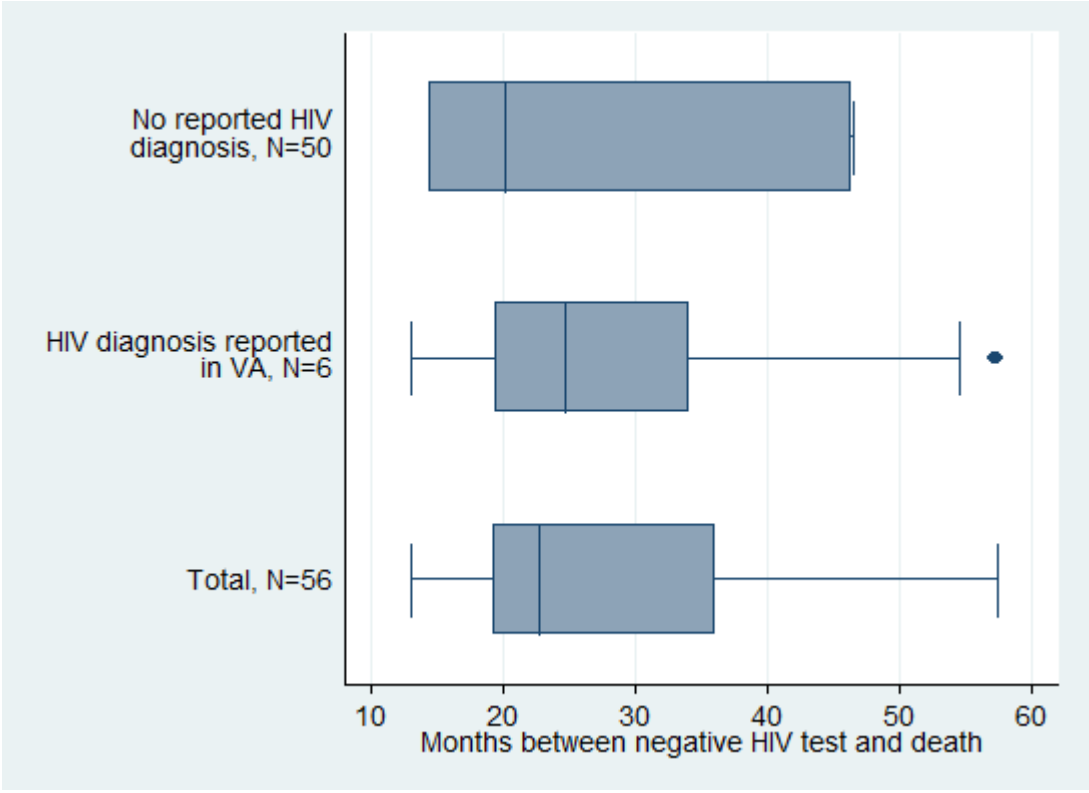

Figure 7: Time from negative HIV test to death for HIV-negative people assigned "HIV/AIDS-related" as their most likely cause of death

ii. Specificity

The summed fractional probability assigned to "HIV/AIDS-related" among deaths of HIV-negative people was 18.6%, and specificity was 81.4% (95% confidence interval: 76.6–85.5%). Allowing reported pre-mortem diagnosis of HIV to serve as an indicator of positive HIV status, eight HIV-negative people were reclassified as HIV-positive. These included 6/56 of the previously HIV-negative people for whom "HIV/AIDS-related" was the most likely cause group assigned by InterVA. Following this reclassification, specificity of InterVA for diagnosing HIV/AIDS as a cause of deaths of adults in Kisesa was 83.0% (95% confidence interval: 78.2–87.0%) (data not shown).

Among HIV-negative people for whom InterVA assigned "HIV/AIDS-related" as the most likely cause group, the median period between their last HIV test and death was 27 months for those with a report of a pre-mortem diagnosis of HIV (range: 10 months–41 months), and 17 months for those without (range: 1 month–51 months) (Figure 7).

iii. Sensitivity analysis of the period of HIV-negative status following a negative HIV test

Table 25 shows the proportion of deaths of HIV-negative people assigned to cause groups other than "HIV/AIDS-related" according to the length of time people were assumed to be HIV-negative following a negative HIV test result. Differences in this proportion comparing the five years used with periods of one, three and seven years were insignificant (p>0.6 in all cases).

Assumed HIV-negative period following negative HIV test	**Proportion of HIV-negative deaths assigned to cause groups other than "HIV/AIDS-related" (specificity), %**	**Z-test *p*-value compared with an assumed HIV-negative period of five years following a negative HIV test result**
One year	79.4	0.684
Three years	80.7	0.831
Five years	81.4	Reference category
Seven years	80.5	0.771

Table 25: Specificity of InterVA in Kisesa by HIV-negative period following a negative HIV test

iv. Symptom profile of deaths of HIV-negative people assigned "HIV/AIDS-related" as the most likely cause

Among 291 deaths of HIV-negative people, 56 were assigned "HIV/AIDS-related" as their most likely cause. With respect to the diagnosis of HIV/AIDS-related deaths, these are false positives. They were assigned strong likelihoods of being due to HIV/AIDS: the median likelihood assigned to these deaths being due to HIV/AIDS was 100%, with an inter-quartile range of 98–100% (Figure 8). Eleven symptoms were reported in at least half of these 56 VAs (Table 26), of which nine were associated at the 5% level with false-positive assignment: fever, weight loss, final illness lasting longer than three weeks, cough, headache, wasting, diarrhea, abdominal pain and lumps. Among the 71% of people with reported cough, three quarters had that cough for more than three weeks. An additional three symptoms occurred in at least half of deaths of HIV-positive people assigned to the cause group "HIV/AIDS-related", all of which were strongly associated with being assigned to that cause group: diarrhea lasting longer than four weeks, skin problems and oral candidiasis.

The symptoms that occurred frequently in deaths assigned to "HIV/AIDS-related" occurred with similar frequency among HIV-negative and HIV-positive people, and with very similar strength of association with whether HIV/AIDS was assigned as the cause of death. Frequently reported symptoms often occurred in combination in the false-positive cases: 45/56 had at least five of the nine most common associated symptoms (Table 27).

III. Summary

Overall, the results from Kisesa show specificity for HIV/AIDS of 81%, and a different pattern of mortality assigned among HIV-negative and HIV-positive people. The leading cause of death was HIV/AIDS, accounting for 34% of all deaths and 58% of deaths of HIV-positive people. Several symptoms were associated with the assignment of "HIV/AIDS-related" as the cause of deaths of HIV-negative people.

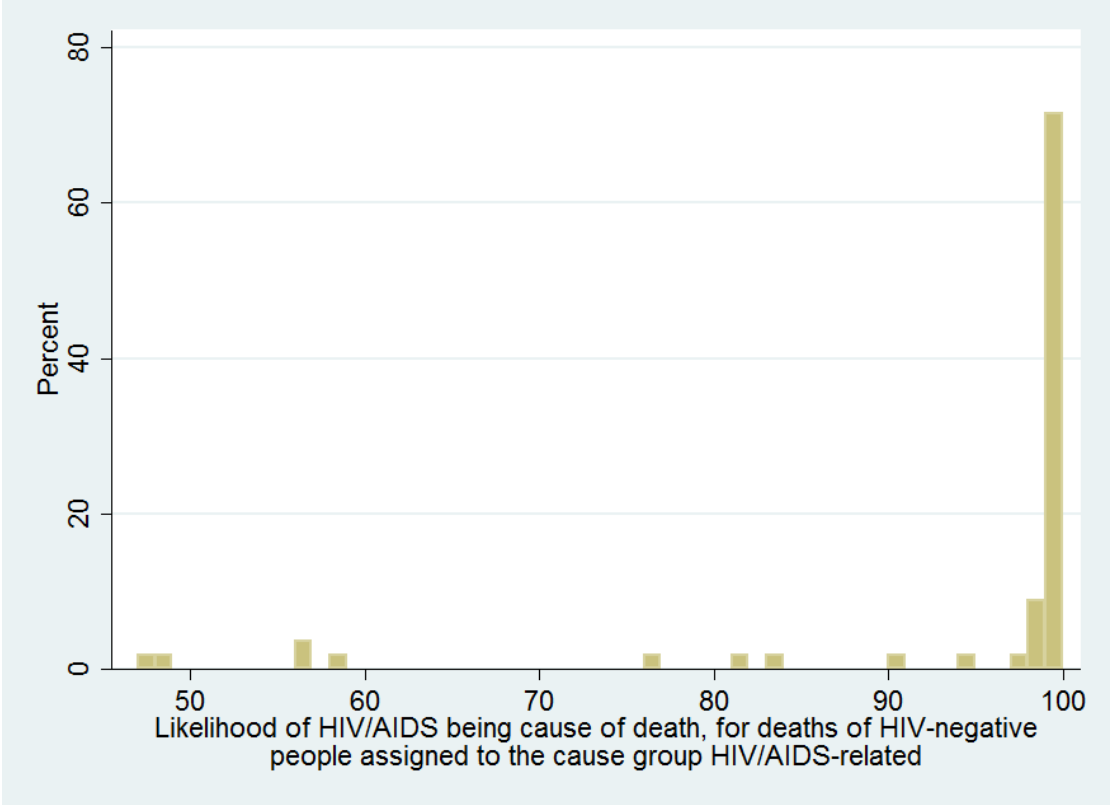

Figure 8: InterVA-assigned likelihood of "HIV/AIDS-related" for HIV-negative people, where that was the most likely cause group, in Kisesa

	291 HIV-negative people					250 HIV-positive people				
	Assigned HIV		Not assigned HIV		Chi² *p*-value for association with	Assigned HIV		Not assigned HIV		Chi² *p*-value for association with
Symptom	N	%	N	%	**assignment of HIV**	N	%	N	%	**assignment of HIV**
Symptoms occurring in ≥50% of HIV-negative deaths assigned to "HIV/AIDS-related"										
Fever	51/56	91.1	87/235	37.0	<0.001	137/147	93.2	70/103	68.0	<0.001
Weight loss	45/56	80.4	59/235	25.1	<0.001	133/147	90.5	55/103	53.4	<0.001
Final illness lasted ≥3 weeks	44/56	78.6	87/235	37.0	<0.001	118/147	80.3	57/103	55.3	<0.001
Cough	39/56	69.6	47/235	20.0	<0.001	121/147	82.3	41/103	39.8	<0.001
Cough lasting ≥3 weeks	*29/56*	*51.8*	*23/235*	*9.8*	*<0.001*	*105/147*	*71.4*	*27/103*	*26.2*	*<0.001*
Headache	39/56	69.6	78/235	33.2	<0.001	101/147	68.7	42/103	40.8	<0.001
Wasting	36/56	64.3	33/235	14.0	<0.001	104/147	70.8	31/103	30.1	<0.001
Death in dry season	34/56	60.7	160/235	68.1	0.293	97/147	66.0	78/103	75.7	0.098
Diarrhea	32/56	57.1	32/235	13.6	<0.001	112/147	76.2	25/103	24.3	<0.001
Any abdominal problem	32/56	57.1	103/235	43.8	0.073	55/147	37.4	53/103	51.5	0.027
Abdominal pain	28/56	50.0	72/235	30.6	0.006	50/147	34.0	47/103	45.6	0.064
Any lumps	28/56	50.0	31/235	13.1	<0.001	71/147	48.3	20/103	19.4	<0.001
Symptoms additionally occurring in ≥50% of HIV-positive deaths assigned to "HIV/AIDS-related"										
Diarrhea lasting ≥4 weeks	22/56	39.3	6/235	2.6	<0.001	87/147	59.2	13/103	12.6	<0.001
Skin problems	26/56	46.4	11/235	4.7	<0.001	82/147	55.8	21/103	20.4	<0.001
Oral candidiasis	16/56	28.6	4/235	1.7	<0.001	80/147	54.4	10/103	9.7	<0.001

Table 26: Symptoms reported in ≥50% of deaths assigned to the cause group "HIV/AIDS-related", by whether deaths were assigned to that cause group and by HIV status, in Kisesa

Any fever	Weight loss	Final illness ≥3 weeks	Cough	Head-ache	Wasting	Diarr-hea	Abdomi-nal pain	Lumps	# symptoms per death
•	•	•	•	•	•	•	•	•	9
•	•	•	•	•	•	•	•	•	9
•	•	•	•	•	•	•	•	•	9
•	•	•	•	•	•	•	•	•	9
•	•	•	•	•	•	•	•		8
•	•	•	•	•	•	•		•	8
•	•	•	•	•	•	•		•	8
•	•	•	•	•	•	•		•	8
•	•	•	•	•	•	•	•		8
•	•	•	•	•	•	•	•		8
•	•	•	•		•	•	•	•	8
•	•	•		•	•	•	•	•	8
•	•	•	•	•	•		•	•	8
•	•	•	•		•	•	•	•	8
•	•	•	•	•	•	•		•	8
•	•	•	•	•	•	•			7
•	•	•	•	•	•	•			7
•	•	•	•	•	•	•			7
•	•	•	•	•	•		•		7
•	•	•	•	•		•	•		7
•	•	•	•		•		•	•	7
•	•	•	•	•	•	•			7
•	•	•		•	•	•	•		7
•	•		•	•	•	•	•		7
•	•	•	•	•	•			•	7
•	•	•		•	•		•	•	7
•	•	•			•	•		•	6
	•	•	•		•	•		•	6
•	•		•	•		•	•		6
•	•		•			•	•	•	6
	•	•	•	•		•	•		6
•	•	•	•		•		•		6
•	•	•		•	•			•	6
	•		•	•	•	•	•		6
•		•	•	•			•	•	6
•	•	•			•			•	5
•		•		•		•		•	5
•	•	•	•				•		5
•		•	•	•		•			5
•	•				•		•	•	5
•	•		•		•			•	5
•			•	•		•		•	5
•	•	•	•	•					5
•	•	•			•			•	5
•	•		•	•				•	5
	•	•			•		•		4
•			•	•			•		4
•	•	•			•				4
•			•	•				•	4
	•		•	•			•		4
•	•	•				•			4
•		•		•					3
•		•		•					3
•		•		•					3
•						•			2
•		•							2

Table 27: Symptoms reported in ≥50%of 56 deaths of HIV-negative people in Kisesa assigned HIV as cause of death, and associated with false-positive assignment of HIV

5. Results: Manicaland

I. Cause-specific mortality distribution

InterVA assigned causes to 1016 deaths from Manicaland. HIV/AIDS was the leading cause of death, assigned as the cause of 57.8% of deaths (Table 28). The next most common cause groups were pulmonary tuberculosis (11.7%), digestive neoplasms (3.3%) and reproductive neoplasms (2.4%). InterVA assigned over two thirds of all deaths in Manicaland to HIV/AIDS or tuberculosis.

Non-HIV, non-tuberculosis infectious causes were assigned to a further 6.1% of deaths, meaning infectious diseases were assigned to 75.6% of deaths overall. Non-communicable diseases were assigned to 11.9% of deaths – primarily digestive neoplasms (3.3%) and reproductive neoplasms (2.4%) – followed by cause of death unknown (7.8%), injuries (4.0%) and maternal conditions (0.7% overall, 1.2% in women).

II. Assessing causes of death against known HIV status

i. Associations between causes of death and HIV status

Of the 1016 VA records that InterVA assigned a cause of death, 186 (18.3%) were HIV-negative within 3.75 years of death, and 777 (76.48%) were HIV-positive. The remaining 53 (5.2%) had unknown HIV status. The largest proportion of deaths of HIV-negative people was assigned to the broad cause category “HIV/AIDS” (37.2%) (Table 29). Similar numbers of deaths were assigned to each of “Non-HIV infections” (17.3%), “Non-communicable diseases” (14.8%), “External causes” (13.7%) and “Cause of death unknown” (15.2%), while the smallest proportion was assigned to “Maternal causes” (1.9%). Over four fifths of deaths of HIV-positive people were assigned to infectious causes (63.2% to “HIV/AIDS” and 18.0% to “Non-HIV infections”), followed by “Non-communicable diseases” (11.2%), “Cause of death unknown” (5.7%), “External causes” (1.4%) and “Maternal causes” (0.5%).

A significantly smaller proportion of deaths of HIV-negative people than HIV-positive people were assigned to the broad cause category “HIV/AIDS” (37.2% vs 63.2%, $p<0.001$) (Table 29). A significantly larger proportion of deaths of HIV-negative people than HIV-positive people

were assigned to the broad cause categories “External causes” (13.7% vs 1.4%, $p<0.001$) and “Cause of death unknown” (15.2% vs 5.7%, $p<0.001$). There was a smaller difference between HIV-negative and HIV-positive people in the proportion of deaths assigned to the broad cause category “Maternal causes” (1.9% vs 0.5%, $p=0.050$), and no difference in the proportion assigned to “Non-communicable diseases” (14.8% vs 11.2%, $p=0.173$) or “Non-HIV infections” (17.3% vs 18.0%, $p=0.823$).

For those broad cause categories to which different proportions of HIV-negative and HIV-positive people were assigned, apart from “Maternal causes”, the proportion of people with unknown HIV status assigned to that broad cause category was between the proportion among HIV-negative people and that among HIV-positive people.

Cause group	% deaths assigned by InterVA
HIV/AIDS-related	*57.8*
Non-HIV infections	*17.8*
Acute respiratory infection/pneumonia	1.1
Diarrhoeal diseases	1.3
Malaria	1.5
Meningitis and encephalitis	0.8
Pulmonary tuberculosis	11.7
Other/unspecified infectious diseases	1.5
Non-communicable diseases	*11.9*
Oral neoplasms	0.1
Digestive neoplasms	3.3
Respiratory neoplasms	0.9
Breast neoplasms	0.1
Reproductive neoplasms	2.4
Other and unspecified neoplasms	0.2
Severe malnutrition	0.5
Diabetes mellitus	1.0
Acute cardiac disease	0.1
Stroke	0.6
Other/unspecified cardiac diseases	0.1
Chronic obstructive pulmonary disease	0.2
Asthma	0.4
Acute abdomen	1.8
Renal failure	0.1
Epilepsy	0.1
Maternal conditions	*0.7*
Abortion-related death*	0.0
Pregnancy-induced hypertension	0.3
Obstetric haemorrhage	0.2
Pregnancy-related sepsis	0.1
Injuries and external causes	*4.0*
Road traffic collision	0.8
Accidental fall**	0.0
Accidental drowning	0.3
Exposure to smoke/fire	0.1
Accidental poisoning	0.1
Intentional self-harm	1.0
Assault	1.0
Other/ unspecified external causes	0.6
Cause of death unknown	*7.8*
Total	**100.0**

Table 28: Distribution of summed fractional probabilities of causes of death assigned by InterVA, Manicaland. *0.02% **0.04%

Broad cause category	HIV-negative	HIV-positive	Status unknown	Total	Z test *p*-value, HIV-neg v. HIV-pos
HIV/AIDS	37.2	63.2	50.2	57.8	<0.001
Non-HIV infections	17.3	18.0	16.2	17.8	0.823
Non-communicable diseases	14.8	11.2	12.6	11.9	0.173
Maternal causes	1.9	0.5	0.0	0.7	0.050
External causes	13.7	1.4	7.1	4.0	<0.001
Cause of death unknown	15.2	5.7	13.9	7.8	<0.001
Total	100.0	100.0	100.0	100.0	
Specificity	**62.8**				

Table 29: Distribution of deaths assigned by InterVA in Manicaland, by broad cause category and HIV status

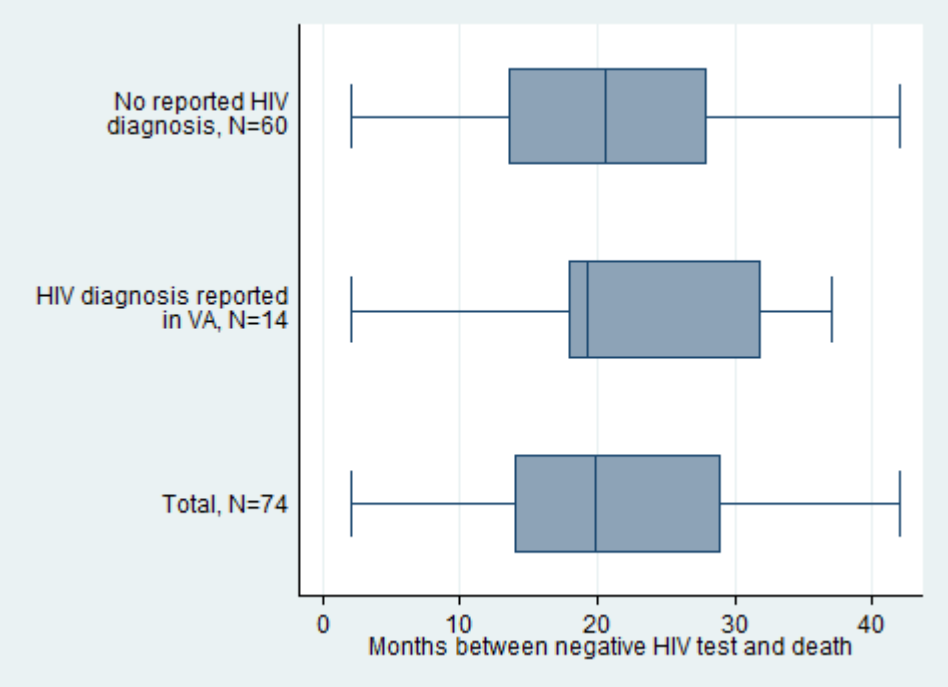

Figure 9: Time between negative HIV test and death for HIV-negative people assigned "HIV/AIDS-related" as their most likely cause of death

ii. Specificity

The summed fractional probability assigned to "HIV/AIDS-related" among deaths of HIV-negative people was 37.2%, and specificity was 62.8% (95% confidence interval: 55.8–69.5%). Allowing reported pre-mortem diagnosis of HIV to serve as an indicator of positive HIV status,

14 HIV-negative people were reclassified as HIV-positive. All these 14 were among the 74 previously HIV-negative people for whom "HIV/AIDS-related" was the most likely cause group assigned by InterVA. Following this reclassification, specificity of InterVA for diagnosing HIV/AIDS as a cause of deaths of adults in Manicaland was 67.9% (95% confidence interval: 60.7–74.5%) (data not shown). The difference in specificity between the two definitions was insignificant (Z test p=0.311, data not shown).

Among HIV-negative people for whom InterVA assigned "HIV/AIDS-related" as the most likely cause group, the median period between their last HIV test and death was 19 months for those with a report of a pre-mortem diagnosis of HIV (range: 2 months–37 months), and 21 months for those without (range: 2 months–42 months) (Figure 9).

iii. Sensitivity analysis of the period of HIV-negative status following a negative HIV test

Table 30 shows the proportion of deaths of HIV-negative people assigned to cause groups other than "HIV/AIDS-related" according to the length of time people were assumed to be HIV-negative following a negative HIV test result. Differences in this proportion comparing the 3.75 years used with periods of one, five and seven years were insignificant (p>0.3 in all cases).

Assumed HIV-negative period following negative HIV test	**Proportion of HIV-negative deaths assigned to cause groups other than "HIV/AIDS-related" (specificity), %**	**Z-test *p*-value compared with an assumed HIV-negative period of 3.75 years following a negative HIV test result**
1 year	70.0	0.349
3.75 years	62.8	Reference category
5 years	63.0	0.968
7 years	63.0	0.968

Table 30: Specificity of InterVA in Manicaland by assumed HIV-negative period following a negative HIV test

iv. Symptom profile of deaths of HIV-negative people assigned "HIV/AIDS-related" as the most likely cause

Among 186 deaths of HIV-negative people, 74 (39.8%) were assigned "HIV/AIDS-related" as their most likely cause. With respect to the diagnosis of HIV/AIDS-related deaths, these are false positives. They were assigned strong likelihoods of being due to HIV/AIDS: the median

likelihood assigned to these deaths being due to HIV/AIDS was 100%, with an inter-quartile range of 93–100% (Table 10). Twelve symptoms were reported in at least half of these 74 VAs (Table 31), of which ten were associated at the 5% level with false-positive assignment: Final illness lasted ≥3 weeks, weight loss, abdominal pain, severe headache, fever, cough, diarrhea, anaemia, drink too much water, vomiting. Among the 65% of people with reported cough, 4/5 had that cough for more than three weeks. An additional five symptoms occurred in at least half of deaths of HIV-positive people assigned to the cause group "HIV/AIDS-related", all of which were strongly associated with being assigned to that cause group: skin problems, productive cough, night sweats, fever lasting ≥2 weeks, difficulty swallowing liquids.

The symptoms that occurred frequently in deaths assigned to "HIV/AIDS-related" occurred with similar frequency among HIV-negative and HIV-positive people, and with very similar strength of association with whether HIV/AIDS was assigned as the cause of death. Frequently reported symptoms often occurred in combination in the false-positive cases: 53/74 had at least six of the ten most common associated symptoms (Table 32).

III. Summary

Overall, the results from Manicaland show specificity for HIV/AIDS of 63%, and a different pattern of mortality assigned among HIV-negative and HIV-positive people. The leading cause of death was HIV/AIDS, accounting for 58% of all deaths and 63% of deaths of HIV-positive people. Several symptoms were associated with the assignment of "HIV/AIDS-related" as the cause of deaths of HIV-negative people.

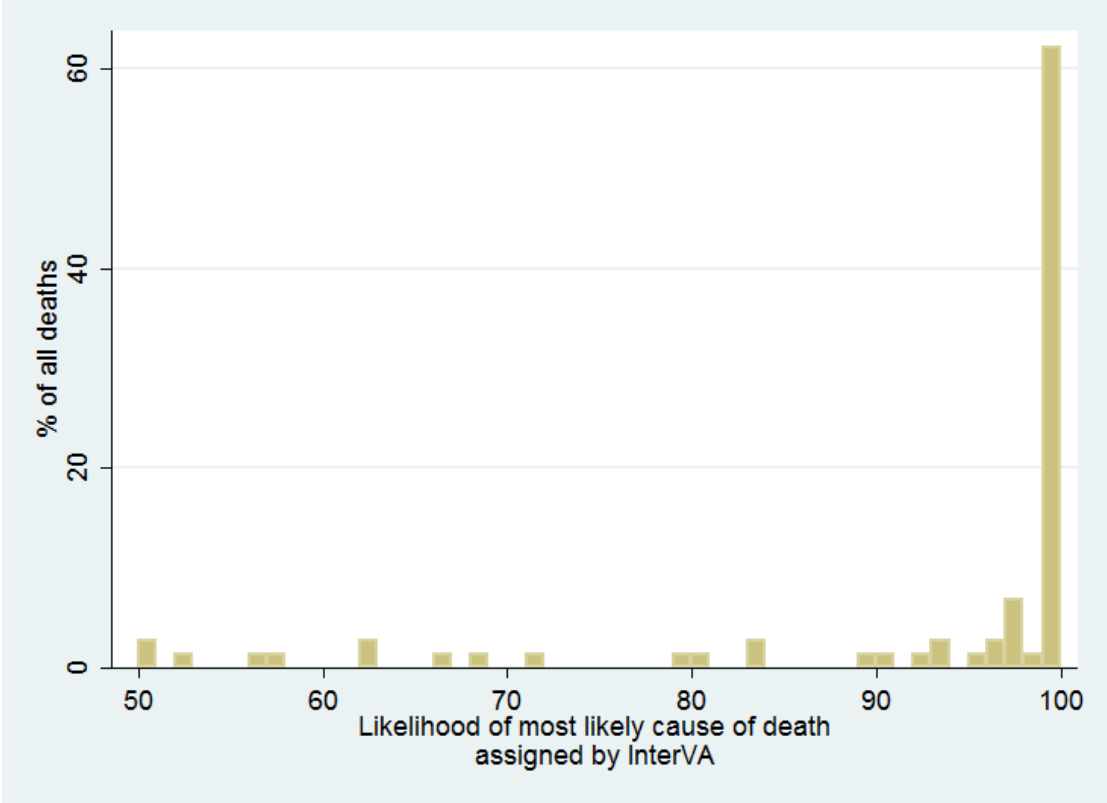

Figure 10: InterVA-assigned likelihood attached to cause group "HIV/AIDS-related" in deaths of HIV-negative people for whom that was the most likely cause group, in Manicaland

	186 HIV-negative people					777 HIV-positive people				
	Assigned HIV		Not assigned HIV		Chi² *p*-value for association with	Assigned HIV		Not assigned HIV		Chi² *p*-value for association with
Symptom	N	%	N	%	**assignment of HIV**	N	%	N	%	**assignment of HIV**
Symptoms occurring in ≥50% of HIV-negative deaths assigned to "HIV/AIDS-related"										
Rigidity	73/74	98.7	105/112	93.8	0.107	490/500	98.0	267/277	96.4	0.175
Final illness lasted ≥3 weeks	72/74	97.3	50/112	44.6	<0.001	485/500	97.0	198/277	71.5	<0.001
Death in wet season	59/74	79.7	84/112	75.0	0.454	380/500	76.0	207/277	74.7	0.693
Weight loss	59/74	79.7	33/112	29.5	<0.001	444/500	88.8	169/277	61.0	<0.001
Abdominal pain	50/74	67.6	40/112	35.7	<0.001	342/500	68.4	159/277	57.4	0.002
Severe headache	50/74	67.6	37/112	33.0	<0.001	361/500	72.2	142/277	51.3	<0.001
Fever	49/74	66.2	39/112	34.8	<0.001	385/500	77.0	163/277	58.8	<0.001
Cough	48/74	64.9	32/112	28.6	<0.001	406/500	81.2	117/277	42.2	<0.001
Cough lasting ≥3 weeks	*40/74*	*54.1*	*15/112*	*13.4*	*<0.001*	*345/500*	*69.0*	*88/277*	*31.8*	*<0.001*
Diarrhea	47/74	63.5	19/112	17.0	<0.001	340/500	68.0	96/277	34.7	<0.001
Anaemia	43/74	58.1	28/112	25.0	<0.001	342/500	68.4	116/277	41.9	<0.001
Drink too much water	43/74	58.1	34/112	30.4	<0.001	288/500	57.6	129/277	46.6	0.003
Vomiting	40/74	54.1	27/112	24.1	<0.001	287/500	57.4	129/277	46.6	0.004
Symptoms additionally occurring in ≥50% of HIV-positive deaths assigned to "HIV/AIDS-related"										
Skin problems	29/74	39.2	3/112	2.7	<0.001	326/500	65.2	69/277	24.9	<0.001
Productive cough	32/74	43.2	12/112	10.7	<0.001	322/500	64.4	82/277	29.6	<0.001
Night sweats	33/74	44.6	23/112	20.5	<0.001	307/500	61.4	126/277	45.5	<0.001
Fever lasting ≥2 weeks	34/74	45.9	12/112	10.7	<0.001	277/500	55.4	80/277	28.9	<0.001
Difficulty swallowing liquids	25/74	33.8	15/112	13.4	0.001	258/500	51.6	82/277	29.6	<0.001

Table 31: Symptoms reported for ≥50% of deaths in Manicaland assigned by InterVA as "HIV/AIDS-related", by HIV status and whether deaths were assigned to the cause group "HIV/AIDS-related"

Final illness ≥3 weeks	Weight loss	Abdominal pain	Severe headache	Fever	Cough	Diarrhea	Anaemia	Drink too much water	Vomiting	symptoms per death
•	•	•	•	•	•	•	•	•	•	10
•	•	•	•	•	•	•	•	•	•	10
•	•	•	•	•	•	•	•	•	•	10
•	•	•	•	•	•	•	•	•	•	10
•	•	•	•	•	•	•	•	•	•	10
•	•	•	•	•	•	•	•	•	•	10
•	•	•	•	•	•	•	•	•		9
•	•	•	•	•	•	•		•	•	9
•	•	•	•	•	•	•	•	•		9
•	•	•	•	•	•	•	•		•	9
•	•	•	•	•	•	•	•	•		9
•	•	•	•	•	•	•	•	•		9
•	•	•		•	•	•	•	•	•	9
	•	•	•	•	•	•	•	•	•	9
•	•	•	•	•	•	•		•	•	9
•	•	•	•	•	•	•		•	•	9
•	•	•		•	•	•	•	•	•	9
•	•	•		•	•	•	•	•	•	9
•	•		•	•	•	•	•	•	•	9
•	•	•	•	•		•	•	•	•	9
•	•	•	•	•		•	•	•	•	9
•	•	•	•	•		•		•	•	8
•	•	•	•	•	•	•	•			8
•	•	•	•	•	•	•		•		8
•	•	•	•	•	•		•	•		8
•	•	•	•		•	•	•		•	8
•	•	•		•	•	•	•		•	8
•	•	•			•	•	•	•	•	8
•	•			•	•	•	•	•	•	8
•		•	•	•		•	•	•	•	8
•		•	•	•	•			•	•	7
•	•	•	•		•	•	•			7
•	•	•	•		•	•	•			7
•	•	•			•	•	•		•	7
•	•	•		•	•	•		•		7
•	•			•	•		•	•	•	7
•	•	•	•	•		•	•			7
•	•	•	•	•	•	•				7
•	•	•	•		•	•	•			7
•	•			•		•	•	•	•	7
•	•		•	•	•			•	•	7
•			•	•		•	•	•	•	7
•	•	•			•	•		•		6
•	•	•			•	•	•			6
•	•	•		•	•			•		6
•	•	•		•	•				•	6
•	•			•	•	•		•		6
•	•	•	•	•		•		•		6
•	•	•		•			•	•	•	6
•	•	•	•	•		•				6
•	•			•		•	•	•		6
•	•			•		•	•		•	6
•	•	•			•				•	5
•			•	•	•		•			5
•	•	•			•		•			5
•	•		•				•		•	5
•	•			•				•	•	5
•	•		•					•	•	5
•	•		•					•	•	5
•		•	•				•			4
•		•			•				•	4
•		•	•					•		4
•	•		•				•			4
•			•	•		•				4
•			•	•					•	4
•	•				•					3
•		•	•							3
•			•	•						3
•			•					•		3
	•		•		•					3
•			•	•						3
•			•							3
•										1

Table 32: Symptoms reported for ≥50% of 74 deaths of HIV-negative people in Manicaland assigned HIV as cause of death, and associated with false-positive assignment of HIV

6. Discussion

Similar to the findings of physician review of verbal autopsy reported in the previous chapter, this analysis describes the cause-specific mortality distribution among the adult populations in Kisesa and Manicaland, and the performance of InterVA for assigning HIV/AIDS as a cause of death. Specificity of InterVA for HIV/AIDS was fairly low in Kisesa (81%) and very low in Manicaland (63%). Deaths assigned to the cause group "HIV/AIDS-related" had similar symptoms profiles among HIV-positive and HIV-negative people. In Kisesa, HIV/AIDS was the most important cause of death, while among other causes non-communicable diseases were predominant. In Manicaland, infectious causes were assigned to the large majority of deaths.

I. Findings of other studies

One study validating the specificity of InterVA for AIDS against medical records, in Addis Ababa, used as its reference standard positive HIV status and evidence of an opportunistic infection[116]. That study found specificity of InterVA-3 to be 80% (95%CI: 67–89%), identical to the present finding for Kisesa. Another validation study in Kilifi, Kenya, using medical records as the reference standard, found specificity somewhat higher than the present study, at 87% (79–92%)[101]. One study in a largely rural population in Ethiopia used the findings of physician review of VA records to validate InterVA-3, finding a specificity of 96% for HIV/AIDS[136]. The author of that study acknowledged the inherent limitation in using physician review as a reference standard, though they did not explicitly state that the same factors in the VA record may lead to false-positive assignment of HIV/AIDS by both the physician review and InterVA. The findings of Lozano and colleagues[93] cannot be compared to the present work, because the validation measures they used, while potentially representing an important methodological advance[130], cannot be calculated with the present datasets due to the absence of a "true" cause for each death. The present finding is lower than the 97% specificity for AIDS achieved by InterVA-4 validated using the medical records in the PHMRC validation dataset[121]. The present specificity is also lower than the specificity of the same version of InterVA found in five African populations, which found 90% (89–91%) overall and 83% (81–86%) among 15–49 year-olds; that study also found specificity of 87% in Kisesa and 66% in Manicaland[102].

II. Specificity and the definition of HIV-positive status

There was an important methodological difference between the latter cited study and the present work: those authors treated diagnoses of HIV reported by relatives of the deceased as evidence of HIV-positive status in the reference standard, compared to the use of positive HIV test result only in the present work. In the present study, such a "diagnosis" was reported for 11% (6/56) of people in Kisesa and 20% (14/74) of people in Manicaland given "HIV/AIDS-related" as their most likely cause group despite having had a recent negative HIV test result. Making the same assumption as Byass and colleagues, the present data give specificity of 83% (78–87%) in Kisesa and 68% (61–75%) in Manicaland. This is closer to the findings of Byass and colleagues – which is unsurprising as there is substantial overlap in the deaths contained in the datasets. People HIV-positive by this definition constituted 2.6% of HIV-negative people in Kisesa (6/235). The seroconversion of 2.6% of people between their last HIV test and death is relatively plausible: data from Kisesa show that around 2.1% of people with at least one negative HIV test seroconverted at some point following that test, with an incidence of 4.5 per 1000 person-years (data not shown). Seroconversion and progression to death in the same period is less plausible: the median time from seroconversion to death is estimated to be over a decade in other sub-Saharan African populations[196]; those estimates are for ART-naïve populations, and it is likely that the time from seroconversion to death is even longer in populations that have some ART availability, such as those in the present study. There were two further HIV-negative people with reported premortem HIV diagnosis, who were assigned to non-HIV cause groups. In Manicaland, 7.5% of HIV-negative people (14/186) were HIV-positive by this definition. Incidence in Manicaland averaged around 14 per 1000 person-years from 2001 to 2006[168]. Applying that incidence to 186 HIV-negative people, assigning each person maximum possible person-time between test and death (3.75 years each) suggests that 9.8 of the 186 (5.3%) would seroconvert in the available time. This is crude, but combined with the point noted above regarding time from seroconversion to death it suggests that the assumption made by Byass and colleagues is implausible applied to the Manicaland data.

The people who "seroconverted" in the present dataset had a median test-to-death interval of 27 months in Kisesa and 19 months in Manicaland, with respective maximum intervals of 41 months and 37 months. Unfortunately, no comparison can be made with studies on the time-to-death of seroconverters as I only have data on people who are deemed seroconverters due to a "diagnosis" reported in the VA interview: such a diagnosis is by definition not possible for

people who do not die, so the data is biased toward those who die. The data are too few to make any assessment of the validity of the assumption made by Byass and colleagues, beyond noting that it equates two very different standards of evidence over HIV status.

There are possible alternative explanations for the present finding regarding seroconversion: first, the HIV-status of deceased people could have been erroneously estimated, due to false test results, data entry errors or to records being wrongly linked. It is important to recall also that a diagnosis of HIV for oneself or one's family member gives rise to stigma and practical concerns[42], which ought to make it relatively unlikely that respondents would volunteer information on a diagnosis of HIV unless they were certain.

III. Cause-specific mortality by HIV status

In contrast to the finding in Kisesa (by both InterVA-4 and physician review), the proportion of mortality assigned to non-communicable diseases in Manicaland was not higher among HIV-negative than HIV-positive deaths. The analysis of false-positive symptoms shows that among the symptoms associated with people being assigned to the cause group "HIV/AIDS-related", all symptoms occurring among HIV-negative people also occur frequently among HIV-positive people – these data do not suggest any "rogue" symptoms driving false-positive assignment of HIV/AIDS. The symptoms that were associated with false-positive HIV diagnosis were several common constitutional symptoms of infection – fever, cough and diarrhea – as part of a lengthy final illness.

IV. Limitations

It is arguably a limitation of InterVA that unlike physician review, it cannot easily take advantage of information recorded in narrative sections of VA interviews. However, other studies have found mixed impacts of coding free text answers on the results given by InterVA, including negligible impact[93], little impact relative to other differences between methods of interpreting VA data[194], and importance for certain causes such as malnutrition and injuries[118].

Though the authors of InterVA suggest considering any of the top three most-likely causes when assessing causes assigned to a death, I excluded from analysis of symptoms occurring in false-positive cases those deaths for which InterVA assigned HIV/AIDS the second- or third-highest probability. This meant that deaths assigned generally lower probabilities of HIV/AIDS-related cause were not considered as potential false-positives. This approach was taken

because I worried that differences between the symptoms of HIV-negative people assigned HIV/AIDS as most-likely and second-most-likely cause might obscure associations with false-positive assignment among the deaths most confidently assigned false-positive HIV/AIDS.

In fact, the probabilities attached to the diagnosis of HIV/AIDS were over 90% for almost all the false-positive deaths (48/56 in Kisesa, 59/74 in Manicaland) for which HIV/AIDS was the most likely cause; the probabilities assigned to HIV/AIDS where it was the second-most-likely cause (4 cases in Kisesa, 2 in Manicaland) were below the lowest probability at which it was assigned as most-likely cause (data not shown).

This exclusion of deaths assigned false-positive HIV/AIDS as second- or third-most-likely cause did not affect the estimates of cause-specific mortality fractions or specificity, as those were calculated using summed fractional probabilities across all deaths (or all deaths of people of a given HIV status).

V. Conclusion

The distribution of causes of death assigned by InterVA to HIV-negative and HIV-positive people was plausible, with the distributions for people of unknown HIV status located between those for HIV-negatives and HIV-positives. Symptoms occurring in people assigned HIV/AIDS as cause of death suggest that occurrence of multiple symptoms is important and that HIV-negative people given false-positive diagnoses have very similar presentations to their HIV-positive counterparts. The specificity for HIV/AIDS was low, albeit similar to previously reported specificity for this method. This prompts caution toward InterVA as a tool for estimating cause-specific mortality. However, this inaccuracy at the individual level may not exclude InterVA from consideration as a useful tool: there is a trade-off between specificity and sensitivity, and authors have suggested that high specificity (≥90%) is most important for those causes of death with relatively low cause-specific mortality fractions[97, 124]. Where a cause is responsible for larger proportions of mortality, as with HIV in much of sub-Saharan Africa, lower specificity may be offset by lower sensitivity, meaning the cause-specific fraction assigned can be accurate. Indeed, in such settings low specificity is needed to offset the effects of sensitivity substantially lower than 100%. The absence of elevated mortality from non-HIV infections among HIV-negative compared with HIV-positive people may suggest a role for HIV infection in causing infectious diseases that are not assigned to “HIV/AIDS” by InterVA.

Combined with the high proportion of deaths of HIV-positive people assigned non-HIV causes, these data suggest that sensitivity is imperfect.

5. The Lopman algorithm

1. Introduction

The Lopman algorithm is a method of ascertaining HIV/AIDS-related mortality[††] from the verbal autopsy data to which it is applied – a "data-derived method". Presented by Ben Lopman and colleagues in two papers[87, 172], it was developed and tested using VA data from two demographic surveillance systems: Manicaland in Zimbabwe and Kisesa in Tanzania.

I. Constructing the algorithm[‡‡]

The algorithm consists of a set of symptoms derived from the larger set of symptoms recorded in the VA questionnaire. The report of any of this subset of symptoms assigns HIV/AIDS as the cause of death. The symptoms in this subset are those that are associated with optimal specificity and sensitivity in categorising deaths as HIV/AIDS-related. The specificity and sensitivity of these symptoms are determined by using a reference standard for an "HIV/AIDS-related" death. Lopman and colleagues defined as HIV/AIDS-related any death in an HIV-positive person who was not reported to have suffered a major injury or a direct obstetric death – strictly speaking, therefore, the algorithm does not aim to ascertain deaths due to HIV/AIDS, but rather to ascertain deaths of HIV-positive people not due to injuries or direct obstetric causes. Major injury was described as "motor vehicle accident, injury that was self-inflicted (suicide), or that was accidentally (accident) or intentionally inflicted by another person (homicide) in the two weeks prior to death". Direct obstetric deaths were "defined as (a) death shortly before delivery, with excessive bleeding and/or severe headaches, or (b) death during childbirth" (Lopman et al 2006: 1274[172]). Such a reference standard, in which all deaths are either HIV/AIDS-related or not HIV/AIDS-related, allows the calculation of both sensitivity and specificity – this is in contrast to the specificity-only reference standard used thus far in this thesis. The algorithm is trained on a dataset containing VA symptoms and known HIV-negative or HIV-positive status for each death. People with unknown HIV status are excluded. In its original development, "symptoms" considered for inclusion in the

[††] Although Lopman and colleagues use the term "AIDS-associated" rather than "HIV/AIDS-related", I will use the latter term for consistency with the rest of the thesis.

[‡‡] This section draws heavily on the methods outlined in Lopman et al 2006.

algorithm did not necessarily consist of single symptoms from the VA transcript: several of the "symptoms" included in the Lopman algorithm as originally derived were in fact combinations of items recorded in the VA. The process of creating the Lopman algorithm consists of four steps:

1. **Likelihood ratios and eligibility**

   All symptoms in a VA dataset have a likelihood ratio (LR) calculated for their association with HIV/AIDS-related deaths in the reference standard. This likelihood ratio prioritises specificity over sensitivity: LR = sensitivity / (1–specificity). Those VA symptoms with LR ≥ 1.92 (corresponding to $p<0.05$ in a chi-squared test on one degree of freedom) are eligible for inclusion in construction of the algorithm.

2. **Iterative addition to the algorithm**

   Symptoms are added to the algorithm one at a time, beginning with that with the highest likelihood ratio and specificity for AIDS deaths from step 1. All deaths for which that symptom is reported are categorised as HIV/AIDS-related. Likelihood ratios and specificity are then recalculated for the remaining symptoms, using the deaths not already classified as HIV/AIDS-related, and the process repeated until all eligible symptoms have been included or all HIV/AIDS-related deaths in the reference standard have been processed.

3. **Graphing symptoms in the ROC curve**

   The symptoms are then graphed in a modified ROC curve in the order of their inclusion in the previous step – that is, in descending order of likelihood ratio. This presentation visualises the trade-off between sensitivity and specificity: the upper-leftmost point of the plot area represents perfect sensitivity and specificity; the curve achieved by graphing the symptoms – as in Figure 11 – shows the cumulative increase in sensitivity and decrease in specificity resulting from iteratively adding symptoms.

4. **Selecting the cut-off point and defining the algorithm**

   In their 2006 paper describing the development of the method and its initial application, Lopman and colleagues stated "The best cut-off value from the modified ROC curve was selected by choosing the point closest to the upper left hand-corner of the plot" (Lopman

et al 2006: 1275[172]). The algorithm is applied non-hierarchically: any death with one or more of the symptoms included up to and including this point is HIV/AIDS-related. Having reported the sensitivity and specificity achieved by the eight most-sensitive symptoms, they stated that "Adding a ninth criteria (tuberculosis) resulted in a drop in specificity that was greater than the gain in sensitivity. … Thus the algorithm that included eight criteria optimised the trade-off between sensitivity and specificity." (Lopman et al 2006: 1275[172]).

The upper left-hand corner of the ROC plot is where specificity and sensitivity are both a perfect 100%. The "point closest to the upper left-hand corner of the plot" is that point of the curve that has the shortest line linking it to the upper left-hand corner. Figure 11 shows a hypothetical ROC curve with four symptoms plotted. It is clear that neither Symptom 1 nor Symptom 4 is closest to the upper left-hand corner, but it is not clear which of Symptom 2 and 3 is closest. The distance from the corner is the hypotenuse of a triangle with base (1 – Specificity) and height (1 – Sensitivity), and Pythagoras' theorem means that this distance is

$$\sqrt{(1 - Sensitivity)^2 + (1 - Specificity)^2}$$

The symptom with the minimum value of this expression is closest to the upper left-hand corner. In the example in Figure 11, this value is 0.408 for Symptom 2 and 0.412 for Symptom 3.

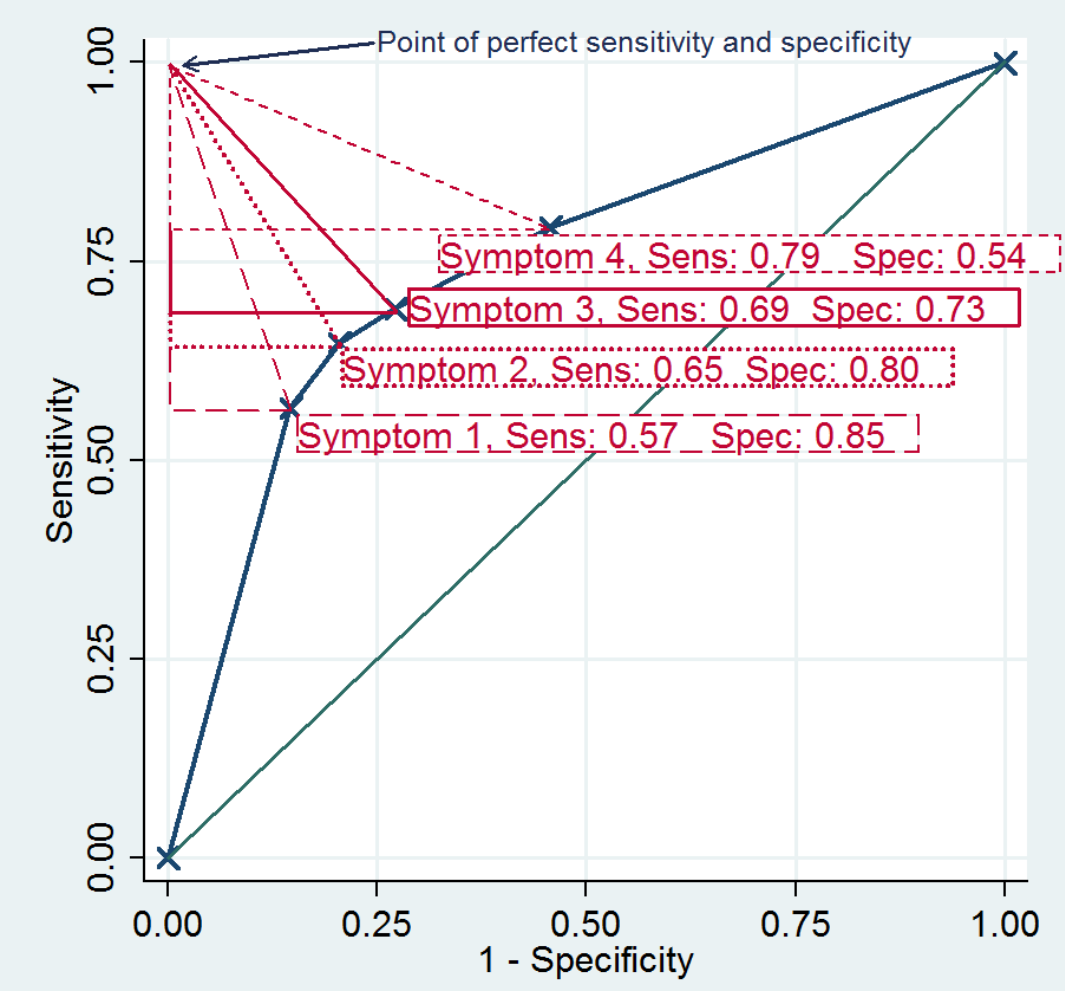

Figure 11: Hypothetical ROC curve showing distance of three points from the upper left-hand corner

II. Application of the algorithm by Lopman and colleagues

The original creation and validation of the algorithm used data from the Manicaland DSS. The deaths in that dataset were divided into a training dataset and a testing dataset in the ratio 3:1, with each death receiving a random number between 0 and 1 and those with numbers ≤0.75 comprising the training dataset[172].

The algorithm originally published consisted of eight symptoms that were included before the cut-off: weight loss, wasting, jaundice, herpes zoster, abscesses or sores, oral candidiasis, acute respiratory tract infections (ARTI) and vaginal tumours. Inclusion of the symptom following vaginal tumours (recent tuberculosis) meant a greater loss in specificity than was gained in sensitivity, and the vaginal tumours point on the ROC curve was closest to the top left of the plot. Vaginal tumours was therefore chosen as the cut-off (Figure 12). This

algorithm achieved specificity of 78% (95% CI: 69–88%) and sensitivity of 71% (65–77%) in the training dataset, and specificity of 76% (59–93%) and sensitivity of 66% (56–77%) in the testing dataset. Note that in Figure 12B, there is no point on the curve for the symptom vaginal tumours, as this symptom was not reported in the testing dataset, meaning the cut-off is effectively at ARTI. Lopman and colleagues present two symptoms beyond the cut-off, recent TB and diarrhoea: they do not report whether there were symptoms beyond these two that had LR≥1.92, nor which other symptoms comprised their VA dataset but had LR<1.92

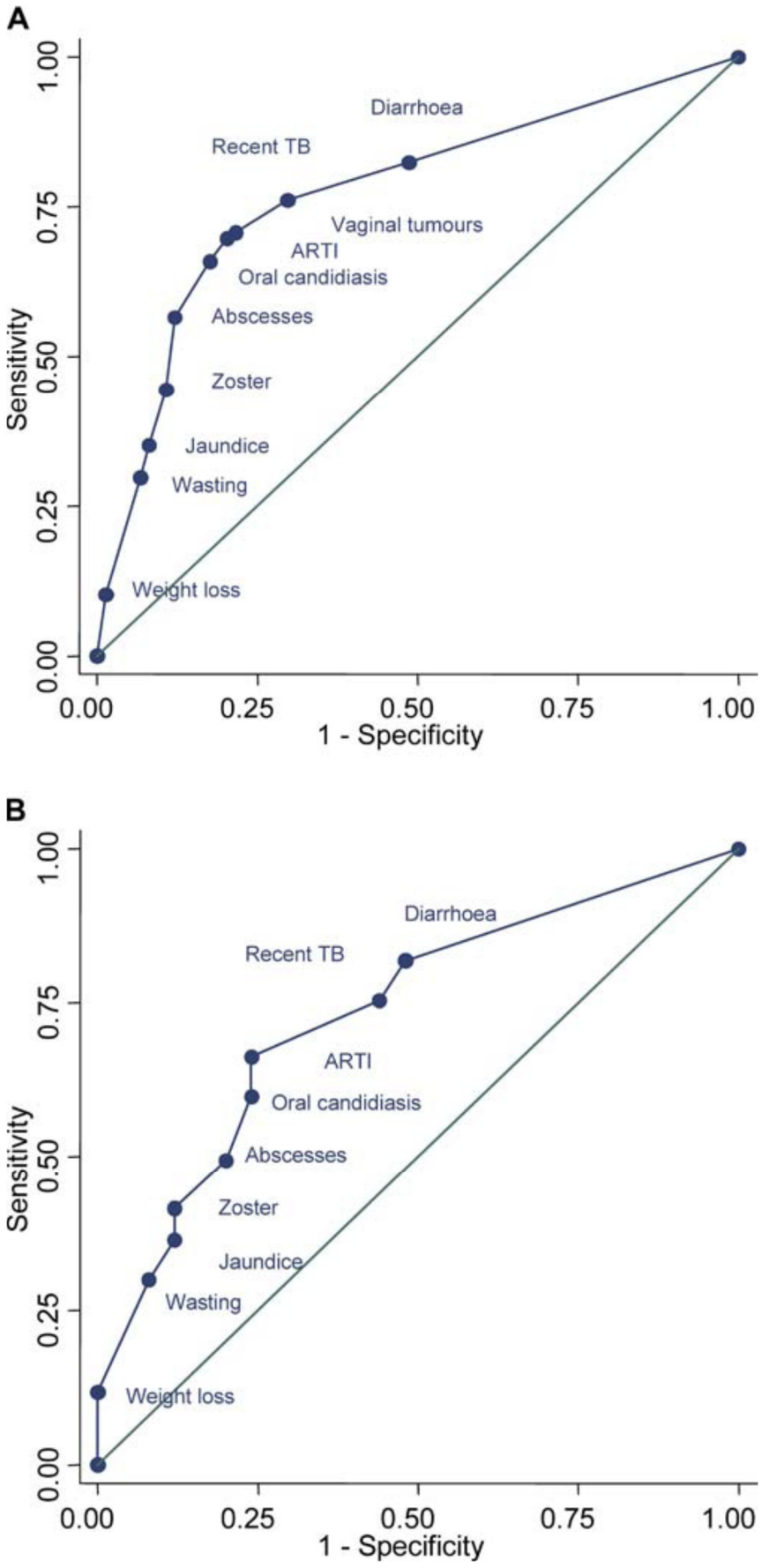

Figure 12: ROC curves showing the performance of the original Lopman algorithm in the training (A) and testing (B) datasets. From Lopman et al 2006.

Later, the algorithm was applied to two Manicaland testing datasets and a testing dataset from Kisesa in Tanzania. The authors found low sensitivity in the Kisesa data (67%) due to very low sensitivity among 45–59 year-olds (46%). Although this was not lower than the 66% sensitivity found in the Manicaland testing dataset in the original application of the algorithm, the authors retrained the algorithm on the Manicaland training data, restricted to 15–44 year-olds. This retrained algorithm consisted of a similar set of symptoms to the original algorithm, and one symptom (diarrhea) was added that had occurred after the cut-off in the original version (Table 33). Only the additional symptom affects the application of the algorithm: although the symptoms in the retrained version were in a different order, the algorithm follows a simple non-hierarchical "or" logic, classifying as HIV/AIDS-related all deaths with any of the symptoms reported.

Original	**Retrained**
Weight loss	Weight loss
Herpes zoster	Vaginal tumours
Jaundice	Wasting
Vaginal tumours	Herpes zoster
Wasting	Abscesses/sores
ARTI	ARTI
Abscesses/sores	Jaundice
Oral candidiasis	Oral candidiasis
	Diarrhoea (additional)

Table 33: Symptoms included in the original Lopman algorithm and that retrained on 15–44 year-olds

The retrained algorithm achieved a specificity of 75% and a sensitivity of 75% in the training dataset of 15–44 year-olds. In the testing datasets, among 15–44 year-old people specificity ranged from 74–79% and sensitivity from 75–83%; among people aged 45–59, specificity ranged from 62–80% and sensitivity from 54–73%.

2. Objectives

I. To assess the specificity and sensitivity of the original Lopman algorithm for HIV/AIDS when applied to the present datasets.

II. To re-derive the Lopman algorithm using the present data, and compare the composition and performance of that algorithm with the original.

III. To investigate variability in the performance of the Lopman algorithm according to the composition of the training dataset.

3. Methods

Analyses were conducted on people aged 15–59.

I. Defining the reference standard

To reproduce the reference standard used in the original Lopman algorithm, I included as "non-HIV-associated" those deaths of HIV-positive people with VA reports suggesting injuries or direct obstetric causes. These I defined as VA reports containing a positive response to any of the following questions:

- Was s/he in a transport accident?
- Did s/he drown?
- Had s/he fallen recently?
- Any suggestion of homicide?
- Any suggestion of suicide?
- Was s/he in a road transport accident?
- Was s/he in a non-road transport accident?
- Was s/he burnt by heat, steam or fire?
- Any poisoning (not by an animal)?
- Was s/he intentionally injured by another person or people?
- Was s/he injured by a force of nature?
- Injured in some kind of violence or assault by another person?

- Did she die in labour undelivered?
- Did she die within 24 hours of delivery?
- Did she have major bleeding in late pregnancy/delivery?
- Did she have major bleeding shortly before labour?

- Did she have major bleeding during labour, before delivering the baby?
- Did she have major bleeding after delivering the baby?
- Mother had excessive vaginal bleeding in pregnancy/postpartum period.
- Were fits only pregnancy-related? **AND** Any blurred vision during the last 3 months of pregnancy?

These questions do not precisely correspond to the description given by Lopman and colleagues, but they satisfy the intention to capture deaths due to direct obstetric causes and injuries. People who were HIV-positive at death but for whom any of these symptoms were reported, as well as HIV-negative people, were treated as "non-HIV-associated" and included in the denominator for specificity calculations. The number of people affected by this categorisation is reported, for information on the scale of the difference entailed by using this definition. The remaining deaths of HIV-positive people are described as "HIV-associated".

II. Applying the original Lopman algorithm to the data in the present study

I created approximately the same symptoms using my dataset as were described by the authors in their original paper. Several of the symptoms could not be precisely recreated using the present VA data specification, although differences between the definitions used by Lopman and colleagues and those used here were not large (Table 34) and were similar to differences deemed acceptably small in Lopman et al 2010.

I constructed the algorithm using the symptoms in the order shown in Figure 12A, and applied it to my data to obtain a ROC curve and calculate the specificity, sensitivity and proportion of deaths correctly classified, as well as the proportion of deaths assigned to HIV/AIDS.

i. Sensitivity analysis of the length of the post-negative period

Lopman and colleagues assumed people to be HIV-negative for three years following a negative HIV test, rather than five years as assumed in this thesis; I investigated whether the specificity estimates varied according to the length of the assumed period of HIV-negative status.

ii. Investigating Lopman's assumption about the composition of the reference standard

To help assess the validity of the assumption that HIV-positive people with reported symptoms of obstetric causes/injuries ought to be treated as non-HIV-associated, I investigated the other symptoms reported in those VA records. Further, to aid comparison with other chapters of this thesis I calculated the specificity using a reference standard of HIV-status alone, and compared this to the specificity using the method used by Lopman and colleagues.

I used the Z-test for all comparisons of specificity, as elsewhere in the thesis.

Symptom	Definition in Lopman et al 2006	Definition from present data specification	Limitations in defining the symptoms
Weight loss	Moderate or severe weight loss with no other symptoms of malnutrition	Had weight loss but not paleness, swollen ankles or hair changing colour	No indication of severity of weight loss
Wasting	Moderate or severe weight loss with at least four of the following symptoms: paleness, changing hair colour, oedema of legs, burning sensations of the feet, dry scaly skin	Had weight loss and at least two of: paleness, swollen ankles, hair changing colour	Fewer items available in the data specification
Jaundice	Acute jaundice (yellowing of the whites of the eyes during the disease that lead to death) with fever and/or itching but without history of alcohol abuse	Had yellowness/jaundice and fever, and no history of alcohol use	No record of itching
Herpes zoster	Ever suffered from zoster	Ever had shingles/herpes zoster	
Abscesses or sores	Had abscesses or sores	Had an ulcer, abcess or sore	
Oral candidiasis	Had two or three of the following: ulcers in the mouth, difficulty swallowing, white patches inside the mouth and tongue	Had mouth sores or white patches on the mouth or tongue, and difficulty drinking	
Acute respiratory tract illness	Trouble breathing, cough lasting 3–27 days with fever but not recent TB, weight loss, or wasting, as above	A cough, chest pain/difficulty breathing, fever, not TB, weight loss, or wasting	Cough was not prescribed in length, as options were shorter or longer than three weeks; timing of TB is not known
Vaginal tumours	Vaginal tumour for at least one month with or without bleeding	Female with genital swelling	Potentially inadequate

Table 34: Signs and symptoms predictive of HIV/AIDS-related deaths (from Lopman et al 2006[172])

III. Re-deriving the Lopman algorithm using the present data

I applied the Lopman method, based on the likelihood ratio and ROC curve, with the present data. Rather than use combinations of reported VA items to derive "symptoms", I used individual VA items from data specification 8.1 (described in the General Methods chapter), as there was no obviously valid way of deciding which combinations ought to be created; the variable `va_wasting` was already a combination variable in Spec 8.1, constructed from reports of individual VA questionnaire items. All symptom variables, beginning `va_` in the data specification, were constructed as binary variables. Where an item in the data specification was categorical, I created a single binary variable for each category. For example, "Cough" was categorised as cough <2 weeks in duration, cough ≥2 weeks in duration, and cough of unknown duration; I created three binary variables reflecting a positive report or not for each of these categories. I treated responses coded missing as negative responses to the question, equal to those coded zero.

The definition of AIDS in the reference standard was the same as that used above in section I. I calculated the specificity, sensitivity and proportion of deaths correctly classified, as well as the proportion of deaths assigned to HIV/AIDS, in both the training and testing datasets, and compared this to the proportion of "true positive" cases in the reference standard using the Z test of two proportions.

As the algorithm is data-derived and constructed maximising specificity, it is not instructive to investigate symptoms associated with false-positive classification of deaths of HIV-negative people: these symptoms are those that form the algorithm, and they are by definition included only if their inclusion brings a greater gain in sensitivity than the loss in specificity.

i. Sensitivity analysis of the length of the post-negative period

I investigated whether there was any effect on specificity of using a cut-off of three years rather than five for the length of time a person was assumed to be HIV-negative following a negative HIV test. This involved maintaining the allocation of records to the training and testing datasets, but deriving the algorithm again using the dataset containing fewer HIV-

negative people and excluding those with a negative HIV test result between three and five years prior to death. I compared the estimated specificity using cut-offs of one, three, five and seven years in Kisesa. In Manicaland, the assumed length of HIV-negative status following a negative test result is 3.75 years and this was compared to lengths of one, five and seven years.

ii. Investigating Lopman's assumption about the composition of the reference standard

As described above in the methods for section II, I investigated the algorithm as applied to HIV-positive people with reported symptoms of obstetric causes and injuries. I also investigated whether specificity was affected by the definition of the reference standard, comparing that originally used by Lopman and colleagues to the HIV-status-only standard used elsewhere in this thesis.

IV. Variability in the performance of the algorithm according to the composition of its training dataset

The eligible records are split at random into a training and a testing dataset. In the course of re-deriving the algorithm, it became clear that the difference in the records assigned to the respective training and testing datasets can result in major differences in the symptoms that form the algorithm. As this may in turn affect the performance of the algorithm in terms of specificity, sensitivity and proportion correctly classified, I added the third objective above: to investigate the variability of the composition and performance of the algorithm to the composition of the training dataset. To investigate this variation, I derived the algorithm 20 times (an arbitrary number chosen to be manageably small but large enough to illustrate potential variability), recording for each version the symptoms included, their order and the validation metrics. The range of symptoms and validation metrics in the variant algorithms were compared to those from the first derivation.

V. Reciprocal application

I applied the Lopman algorithm derived in the Kisesa data to the Manicaland data, and vice versa, reporting the same validation metrics as above.

4. Results: Kisesa

I. Constructing the reference standard in the present Kisesa data

Of 1246 deaths with VA records, just over half had unknown HIV status (648/1246, 52.0%) and were excluded from analysis; among those with known HIV status at death, 322 (25.8%) were HIV-negative within five years prior to death, and 276 (22.2%) were HIV-positive. Of these 276, ten (3.6%) had suffered injuries, major peri-partum bleeding or symptoms of eclampsia or died peri-partum. For five of these people, injuries were reported (three assaults, one suicide and one road transport collision), and five had major bleeding in late pregnancy, delivery or the post-partum. They were classified as HIV-negative in the reference standard, which consisted of deaths of 598 people: 332 non-HIV-associated (55.5%) and 266 HIV-associated (44.5%).

II. Applying the original Lopman algorithm to the Kisesa data

Applied to this dataset, the original Lopman algorithm achieved a specificity of 66.0% (95% confidence interval (CI) 60.7–70.9%) and a sensitivity of 65.0% (95% CI 59.1–70.5%); it correctly classified 65.6% of deaths, and classified 47.8% of deaths as HIV/AIDS-related, 3.3 percentage points higher than the proportion HIV/AIDS-related in the reference standard (p=0.246). The notable differences between the ROC curve derived in Figure 13 and the curves in the original application by Lopman and colleagues (Figure 12) are that in the present dataset, weight loss has higher sensitivity, and weight loss and jaundice have lower specificity.

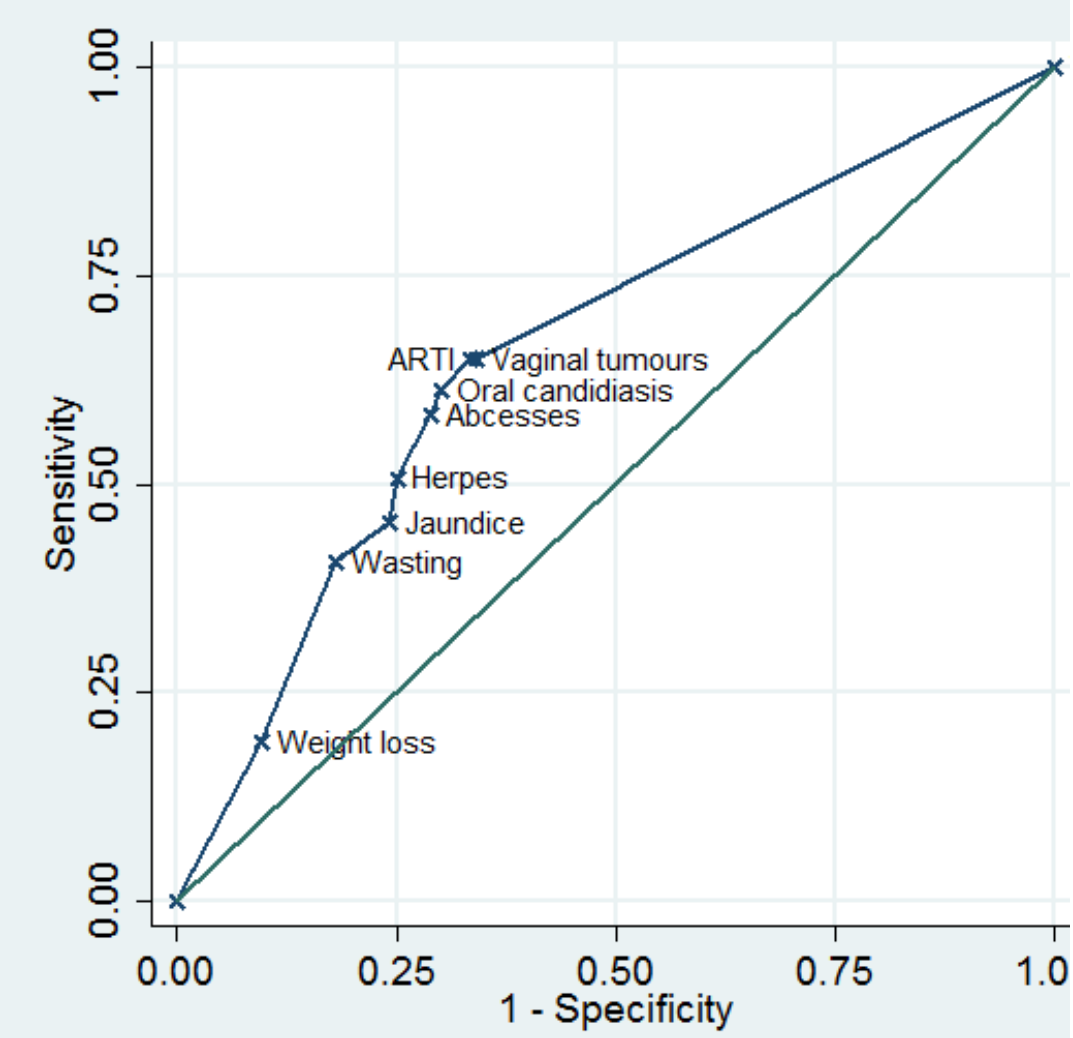

Figure 13: ROC curve applying the original Lopman algorithm to the present Kisesa dataset

i. Sensitivity analysis of the length of the post-negative period

There was no difference in the specificity of the original Lopman algorithm comparing the five-year assumed HIV-negative period used here with the three years assumed by Lopman and colleagues, or with periods of one or seven years (Table 35).

Assumed HIV-negative period following negative HIV test	Specificity (%)	Z-test *p*-value compared with an assumed HIV-negative period of five years following a negative HIV test result
One year	68.0	0.694
Three years	66.2	0.929
Five years	66.0	Reference category
Seven years	66.0	Identical to reference category

Table 35: Specificity of the original Lopman algorithm by HIV-negative period following a negative HIV test, Kisesa

ii. Investigating Lopman's assumption about the composition of the reference standard

Half of the ten HIV-positive people classified as having died of obstetric causes or injuries had at least one symptom included in the algorithm: the one person with reported suicide, and four women with major bleeding in pregnancy or the post-partum period. Three had one symptom (two jaundice, one wasting) and two had three symptoms (one wasting, herpes and abscesses; and one – the person with reported suicide – weight loss, herpes and abscesses).

Using a reference standard consisting only of HIV status, and thus not recategorising as "non-HIV-associated" those HIV-positive people with symptoms of obstetric causes or injuries, specificity was similar to that achieved using Lopman's reference standard: 64.5% compared to 66.0%.

III. Re-deriving the Lopman algorithm using the present data

For the purposes of re-deriving the algorithm, the 598 records with known HIV status were split into a training and a testing dataset, based on the allocation of random numbers described above: 461 (77.1%) in the training dataset and 137 (22.9%) in the testing dataset. The proportion of deaths in the training and testing datasets that were HIV/AIDS-related in the reference standard was 44.9% and 43.1%, respectively (chi-squared p=0.704).

The likelihood ratios for all symptoms are shown in Appendix 11. Fifteen symptoms had likelihood ratio ≥1.92 and were eligible for inclusion in the algorithm. I graphed all these eligible symptoms in a ROC curve included in the order of highest likelihood ratio (Figure 14), as described above. Productive cough was the first symptom the inclusion of which meant a

loss of specificity greater than the gain in sensitivity. The cut-off for the algorithm was therefore made prior to productive cough, at abnormal hair colour; abnormal hair colour was also marginally closer to the top-left of the plot than productive cough. The algorithm was defined as indicating an HIV/AIDS-related death when any of the following seven symptoms was present: herpes zoster, diagnosis of HIV, oral candidiasis, sunken eyes, ulcers (not on feet), diagnosis of tuberculosis or abnormal hair colour (Figure 15, Figure 16, Table 36).

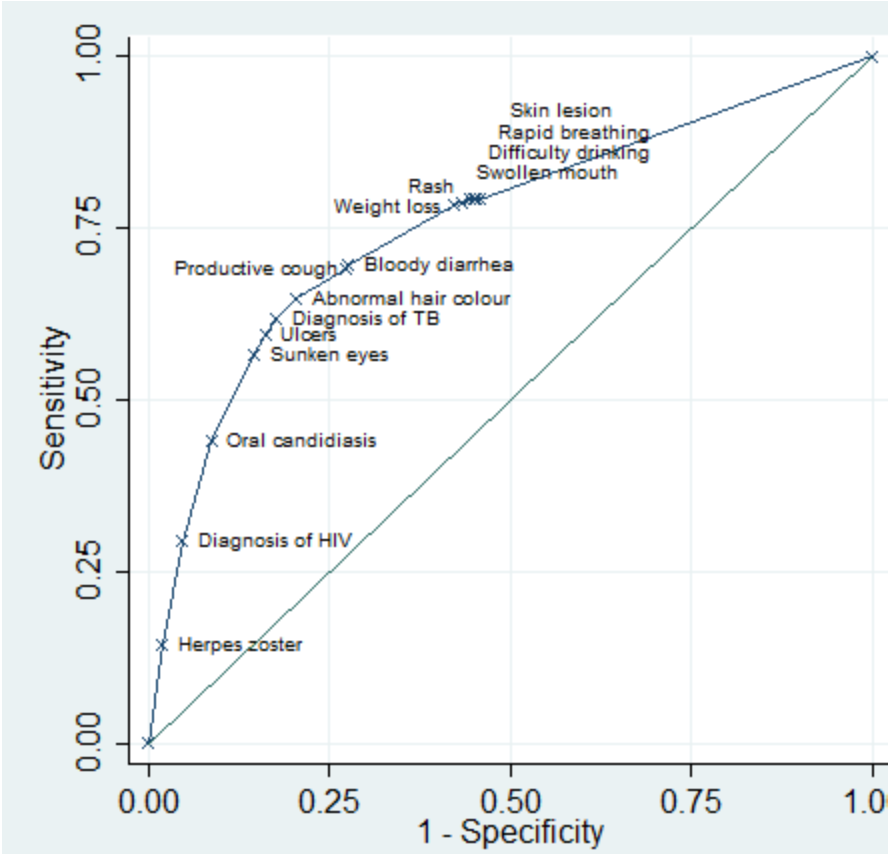

Figure 14: ROC curve with all eligible symptoms, in Kisesa

In the training dataset, this algorithm achieved specificity of 79.5% (95% CI 74.1–84.0%) and sensitivity of 64.7% (95% CI 58.0–70.9%) (Table 37). It correctly classified 72.9% of deaths and classified 40.3% of deaths as HIV/AIDS-related. This was 4.6 percentage points lower than the 44.9% in the reference standard, and not statistically different (p=0.162). Applied to the testing dataset, the algorithm achieved specificity of 74.4% (95% CI 63.7–82.7%) and sensitivity of 49.2% (95% CI 36.8–61.6%) (Table 37). It correctly classified 63.5% of deaths and classified 35.8% of deaths as HIV/AIDS-related – 7.3 percentage points lower than in the reference standard of the testing dataset, but not statistically significant (p=0.216). Sunken eyes was the

symptom with the most strikingly discrepant impact between the training and testing datasets: when included in the training dataset it had a specificity of 94% and sensitivity of 22%, while in the testing dataset it had an unusually low specificity (88%) and much lower sensitivity than in the training dataset (6%) (Table 36). Although there was a similar prevalence of positive reports of sunken eyes at the point of its inclusion in the training and testing datasets (15+26 of 348 remaining deaths in training, 11.8%; 9+2 of 108 remaining deaths in testing, 10.2%), in training 63% of reports of sunken eyes were for HIV-positive people, while in testing this was just 18%.

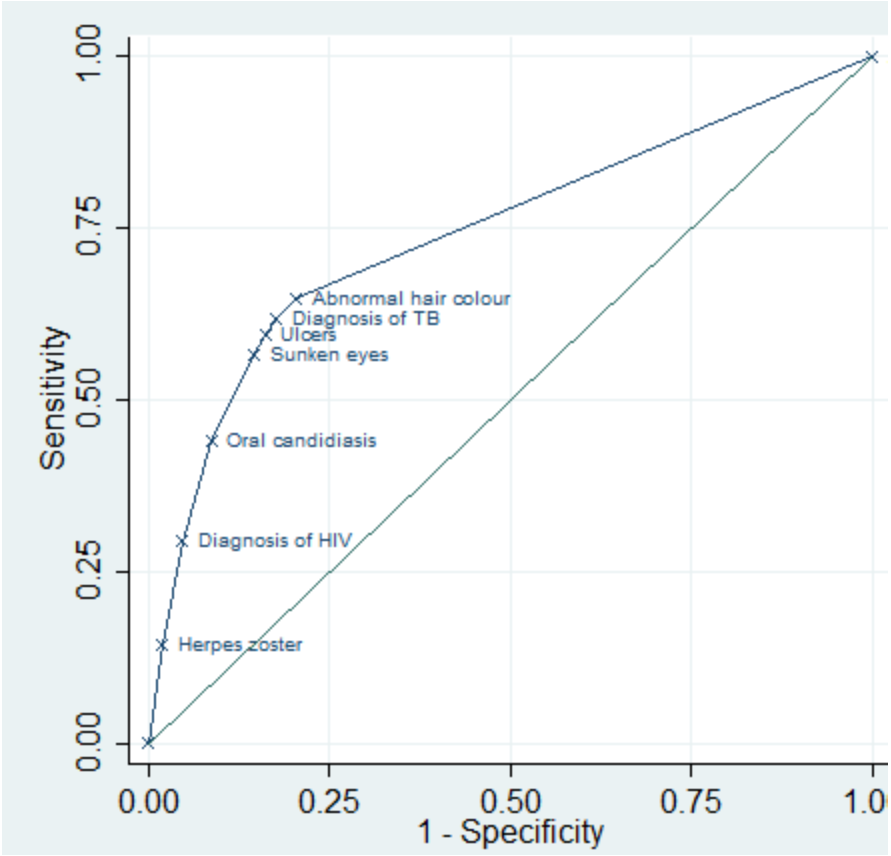

Figure 15: ROC curve of Lopman algorithm derived in the training dataset, with cut-off at "Abnormal hair colour"

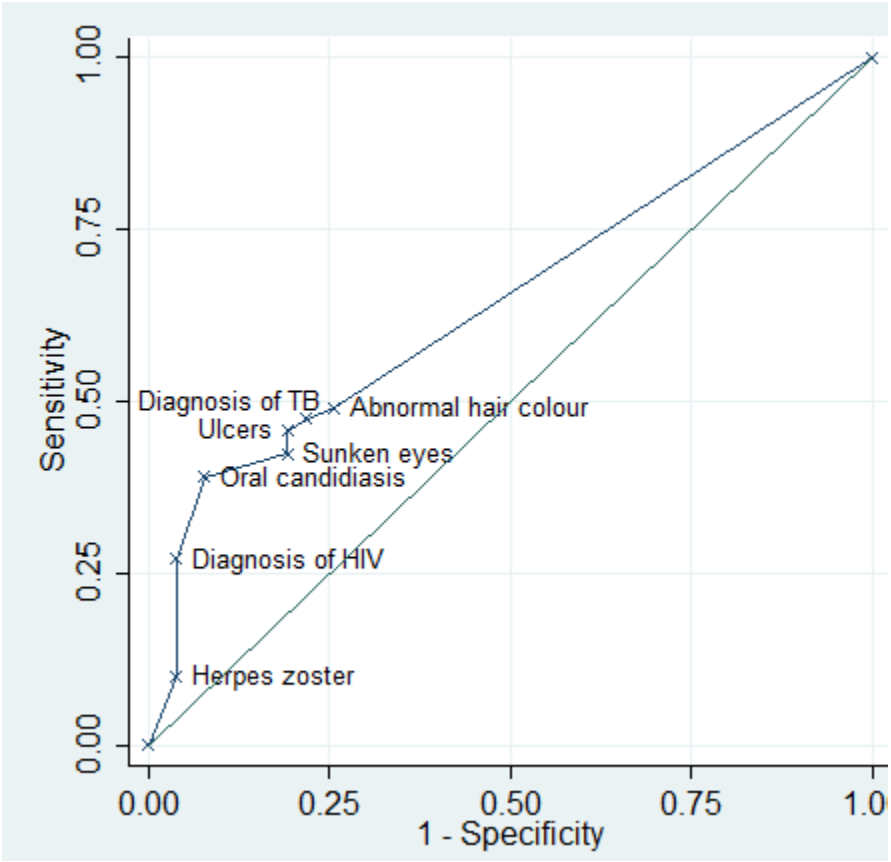

Figure 16: ROC curve of Lopman algorithm derived in the training dataset and applied to the testing dataset

Training dataset	N deaths remaining			N with symptom		Specificity and sensitivity when added	
Symptom	Total	non-HIV-associated	HIV-associated	non-HIV-associated	HIV-associated	Sp	Se
Herpes	461	254	207	5	30	98%	14%
Diagnosis of HIV	426	249	177	7	31	97%	18%
Oral candidiasis	388	242	146	10	30	96%	21%
Sunken eyes	348	232	116	15	26	94%	22%
Ulcers	307	217	90	4	6	98%	7%
Diagnosis of TB	297	213	84	4	5	98%	6%
Abnormal hair colour	288	209	79	7	6	97%	8%
Not classified as HIV/AIDS-related	275	202	73				

Testing dataset	N deaths remaining			N with symptom		Specificity and sensitivity when added	
Symptom	Total	non-HIV-associated	HIV-associated	non-HIV-associated	HIV-associated	Sp	Se
Herpes	137	78	59	3	6	96%	10%
Diagnosis of HIV	128	75	53	0	10	100%	19%
Oral candidiasis	118	75	43	3	7	96%	16%
Sunken eyes	108	72	36	9	2	88%	6%
Ulcers	97	63	34	0	2	100%	6%
Diagnosis of TB	95	63	32	2	1	97%	3%
Abnormal hair colour	92	61	31	3	1	95%	3%
Not classified as HIV/AIDS-related	88	58	30				

Table 36: The effects of individual symptoms on algorithm specificity and sensitivity in Kisesa, in the training and testing datasets

Dataset	Specificity %	Sensitivity %	% correctly classified	% assigned as HIV/AIDS-related	% HIV/AIDS-related in reference standard
Training	79.5	64.7	72.9	40.3	44.9
Testing	74.4	49.2	63.5	35.8	43.1

Table 37: Performance of Lopman algorithm in training and testing datasets in Kisesa

i. Sensitivity analysis of the length of the post-negative period

There was no difference in the specificity of the derived algorithm comparing the five-year assumed HIV-negative period used here (specificity = 79.5%) with the three years assumed by Lopman and colleagues (specificity = 78.1%, *p*=0.915) (Table 38), or with assumed periods of one or seven years.

Assumed HIV-negative period following negative HIV test	Specificity (%)	Z-test *p*-value compared with an assumed HIV-negative period of five years following a negative HIV test result
One year	73.4	0.253
Three years	78.1	0.915
Five years	79.5	Reference category
Seven years	79.5	(no difference)

Table 38: Specificity of the Lopman algorithm in Kisesa, by length of HIV-negative period following a negative HIV test

ii. Investigating Lopman's assumption about the composition of the reference standard

Four of the ten HIV-positive people re-classified as non-HIV-associated due to having reported obstetric causes or injuries had at least one symptom included in the algorithm: the one person with reported suicide, and three women with major bleeding in pregnancy or postpartum. Two had one symptom (one with sunken eyes, one with abnormal hair colour), one – the person with reported suicide – had three symptoms (herpes, oral candidiasis and abscesses) and one had six symptoms (herpes, diagnosis of HIV, oral candidiasis, sunken eyes, abscesses and abnormal hair colour).

Using a reference standard consisting only of HIV status, specificity and sensitivity in the training dataset were almost identical to those achieved with this algorithm using Lopman's reference standard: specificity was respectively 79.8% and 79.5% (*p*=0.931), and sensitivity was respectively 63.8% and 64.7% (*p*=0.851).

IV. Variability in the performance of the algorithm according to the composition of its training dataset

I derived 20 variants of the Lopman algorithm using the same method as used to derive the version used in the above analyses ("the version used"). Four symptoms occurred in all 20 variants (diagnosis of HIV, ulcers/abscesses, herpes zoster and sunken eyes) (Table 39). Oral candidiasis occurred in 17/20 variants. The number of symptoms in the variants ranged from seven to 11, compared to seven in the version used. The number of eligible symptoms (those with LR≥1.92) also varied between variants, from 14 to 18 (Table 40).

All four symptoms that occurred in all the variant algorithms also occurred in the version used. Diagnosis of HIV was always ranked first or second in the algorithm, while the other symptoms that occurred in all variants (ulcers/abscesses, herpes zoster, sunken eyes) had a range of ranks. Oral candidiasis featured both in the version used and frequently across the variants, but productive cough did not feature in the version used. The version used featured two symptoms that occurred less frequently across the variants: abnormal hair colouring (4/20) and premortem medical diagnosis of tuberculosis (5/20). Abnormal hair colouring ranked second in three of its four appearances in the variants, while diagnosis of TB was ranked medium-to-low.

Herpes zoster, ulcers/abscesses and oral candidiasis occurred frequently across the variants and also in the original Lopman algorithm – although by contrast, the most specific symptoms from the original algorithm played a minor part across the variants derived in the present dataset: weight loss and wasting, as well as jaundice, did not feature in any variant.

The specificities and sensitivities of the variant algorithms as applied to their respective training datasets differed widely, with specificity ranging from 72% to 84%, and sensitivity ranging from 56% to 73% (Figure 17). The proportion of deaths correctly classified ranged from 70% to 75%. The proportion assigned to HIV/AIDS ranged from 34% to 48%, and the absolute difference compared to the proportion of "true positives" in the reference standard ranged from zero to 11 percentage points; seven variant algorithms had an absolute difference of over five percentage points (Table 40).

As per the methods for deriving the version used in the analyses above, I made the cut-off for the variant algorithms at the point before the first symptom whose inclusion meant a loss of specificity greater than the gain in sensitivity (the "specificity cut-off"), rather than the point

closest to the upper left-hand corner of the plot. In ten of the 20 variants, these two conditions resulted in the same algorithm. However, in the other ten variants, using the upper leftmost point resulted in a greater number of symptoms in the algorithm than using the specificity cut-off: in four variants, the upper leftmost point cut-off resulted in one more symptom included, in another four variants it resulted in two more symptoms included, and in one variant each the upper leftmost point cut-off resulted in four and five more symptoms. Although the impact on the proportion of deaths correctly classified was minimal, versions using the upper leftmost point cut-off had lower specificity and higher sensitivity than versions using the specificity cut-off (Figure 18).

V. Summary

The version of the Lopman algorithm derived in the present data achieved specificity of 80% in the training dataset and 74% in the testing dataset, and sensitivity of 65% and 49% respectively. It correctly classified 73% and 64% of deaths in the respective datasets, and classified 40% and 36% as due to HIV/AIDS. There was a great range of specificity and sensitivity achieved by different variants of the algorithm derived on the same dataset.

Short name	Symptom	Frequency across 20 variants	Present in version used?
HIV	Medical diagnosis of HIV/AIDS	20	Yes
Ulce	Ulcers/ abscesses or sores on body, apart from feet	20	Yes
Herp	Herpes zoster	20	Yes
Eyes	Eyes sunken	20	Yes
OrCa	Oral candidiasis	17	Yes
NiSw	Excessive night sweats	14	
ProCo	Productive cough	10	
RapBr	Rapid breathing	8	
Armp	Lump or lesion in armpit	6	
TB	Diagnosis of TB	5	Yes
BlCof	Coughing with blood	4	
Hair	Abnormal hair colouring	4	Yes
Neck	Stiff or painful neck < 1 wk	4	
Breast	Any breast lump or lesion	3	
Rash	Rash	2	
BlVom	Any vomiting with blood	2	
Haem	Haematuria	2	
Rehyd	Oral rehydration required during final illness	1	
Convul	Convulsions	1	

Table 39: Symptoms and symptom frequency across 20 variant Lopman algorithms, Kisesa

	Twenty variant algorithms																				Version used
Variant no.	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	
Symptoms in algorithm	***HIV***	***HIV***	***HIV***	***HIV***	***HIV***	***HIV***	***HIV***	***HIV***	***HIV***	***HIV***	***HIV***	***HIV***	***HIV***	***HIV***	***HIV***	***HIV***	***HIV***	OrCa	***HIV***	Herp	Herp
	Herp	Ulce	Breast	Ulce	OrCa	Hair	Hair	OrCa	BlVom	OrCa	OrCa	OrCa	Herp	Hair	OrCa	Herp	OrCa	***HIV***	OrCa	***HIV***	***HIV***
	Ulce	OrCa	OrCa	OrCa	Eyes	NiSw	NiSw	Ulce	Ulce	Ulce	Ulce	Ulce	OrCa	Herp	Herp	Ulce	Ulce	BlVom	Ulce	OrCa	OrCa
	Breast	Armp	Herp	Herp	RapBr	Haem	Ulce	Eyes	Herp	Herp	Herp	Herp	Ulce	NiSw	Ulce	NiSw	Herp	Breast	Herp	Ulce	Eyes
	NiSw	Neck	Ulce	Armp	Herp	ProCo	Armp	ProCo	OrCa	TB	Eyes	Eyes	Eyes	Ulce	Eyes	Armp	Rash	Herp	Armp	Armp	Ulce
	Eyes	Herp	Eyes	Eyes	NiSw	Herp	Herp	Herp	Eyes	Eyes	TB	RapBr	NiSw	ProCo	TB	OrCa	Eyes	Ulce	Rash	Eyes	TB
	BlCof	ProCo	RapBr	Neck	BlCof	OrCa	Eyes	NiSw	ProCo	ProCo	ProCo	Neck	ProCo	Eyes	RapBr	Eyes	RapBr	Eyes	Eyes	RapBr	Hair
	RapBr	Eyes	ProCo	BlCof	ProCo	Ulce	RapBr	Neck	NiSw	NiSw	NiSw	TB	BlCof								
	Hair	NiSw	NiSw	NiSw	Ulce	Eyes															
	Haem	TB	Rehyd																		
	Convul																				
# symptoms	11	10	10	9	9	9	8	8	8	8	8	8	8	7	7	7	7	7	7	7	7
# eligible symptoms	18	16	16	17	14	14	16	15	15	14	16	16	16	15	15	14	16	18	14	17	15
% specificity	78	73	74	83	73	73	80	74	73	72	74	80	76	73	82	79	80	84	81	83	80
% sensitivity	69	69	70	65	68	68	66	69	68	70	73	65	67	69	59	60	59	56	61	64	65
% correctly classified	74	71	72	75	71	71	74	72	71	71	74	73	72	71	72	70	71	72	72	75	73
% assigned to HIV/AIDS	44	45	45	39	45	45	41	45	45	47	48	40	45	46	36	39	37	34	38	37	40
% "true" HIV/AIDS	46	44	43	46	45	45	45	44	45	45	46	44	48	46	44	46	43	45	45	43	45
Difference, pp	–2.2	1.3	2.2	–6.7	0.2	0.8	–4.8	1.1	0.4	1.8	1.6	–4.1	–3.4	0.0	–7.9	–7.1	–6.4	–11.0	–7.3	–5.6	–4.6
p for difference	0.503	0.687	0.501	0.039	0.946	0.786	0.141	0.733	0.895	0.59	0.637	0.204	0.309	1.000	0.013	0.031	0.049	0.001	0.026	0.088	0.162

Table 40: Composition and performance of 20 random variant Lopman algorithms, and the version used in the present analyses; specificity, sensitivity, % correctly classified, % assigned HIV/AIDS and comparison with % HIV/AIDS in reference standard, Kisesa. pp=percentage points

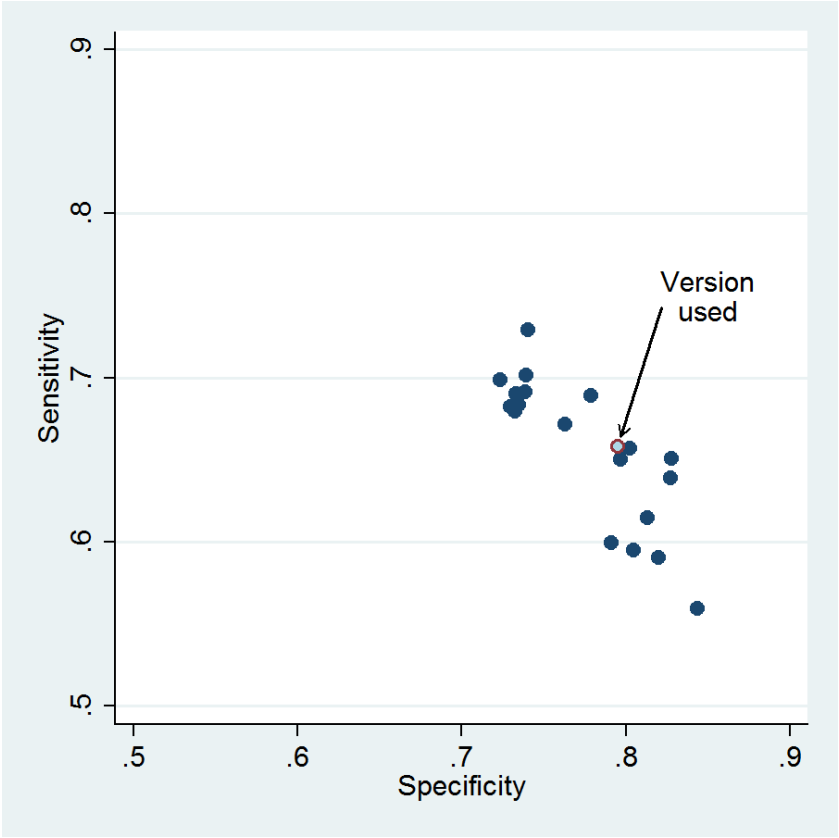

Figure 17: Specificities and sensitivities of 20 variant Lopman algorithms and the version used, in their respective training datasets

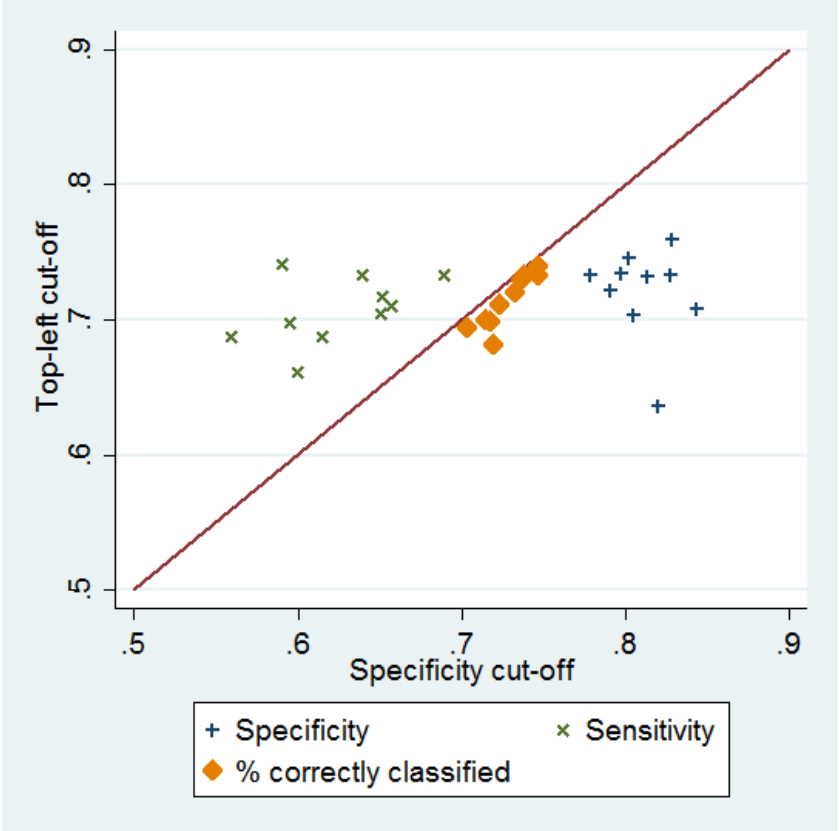

Figure 18: Specificity, sensitivity and % correctly classified in training datasets for variant Lopman algorithms according to whether the specificity cut-off or the upper-leftmost point cut-off was used, Kisesa

5. Results: Manicaland

I. Constructing the reference standard in the present Manicaland data

Of 1021 deaths with VA records, 5.5% had unknown HIV status (56/1021) and were excluded from analysis; 187 (18.3%) were HIV-negative within 3.75 years of death, and 778 (76.2%) were HIV-positive. Of these 778, 21 (2.7%) had suffered injuries or major peri-partum bleeding, symptoms of eclampsia or died peri-partum. For 17 of these people, injuries were reported (four homicides, three poisonings, two suicides, two transport collisions, one burn and five unspecified injuries), and four had major bleeding in late pregnancy, delivery or the post-partum. They were classified as HIV-negative in the reference standard, which consisted of deaths of 965 people: 208 non-HIV-associated (21.6%) and 757 HIV-associated (78.4%).

II. Applying the original Lopman algorithm to the present Manicaland data

Applied to this dataset, the original Lopman algorithm achieved a specificity of 54.8% (95% CI 48.0–61.4%) and a sensitivity of 81.5% (95% CI 78.6–84.1%); it correctly classified 75.8% of deaths, and classified 73.7% of deaths as HIV/AIDS-related, 3.7% lower than the 78.4% of deaths HIV/AIDS-related in the reference standard (p=0.014). The most notable difference between the ROC curve derived in the present data (Figure 19) and the original application by Lopman and colleagues (Figure 12) is the much more prominent role of wasting, both in lower specificity and in much higher sensitivity.

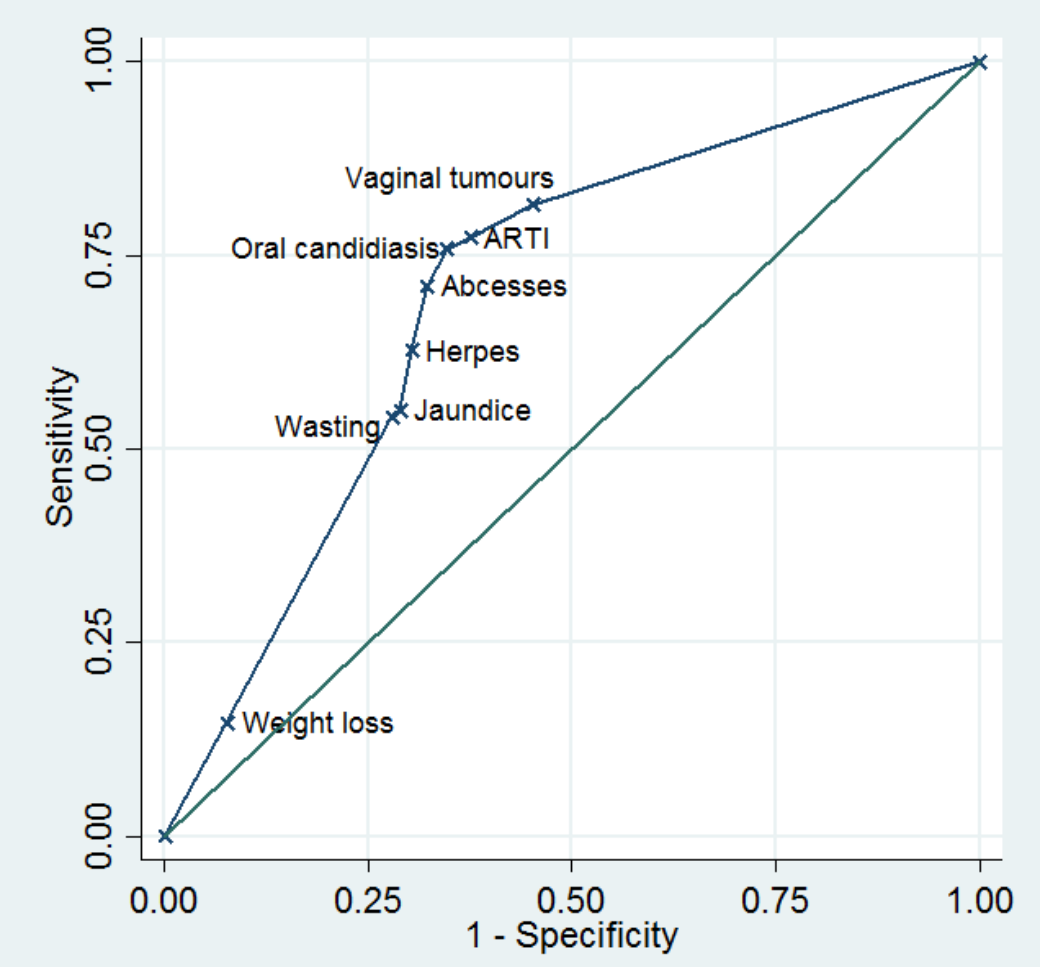

Figure 19: ROC curve applying the original Lopman algorithm to the present Manicaland dataset

i. Sensitivity analysis of the length of the post-negative period

There was no difference in the specificity of the original Lopman algorithm comparing the 3.75-year assumed HIV-negative period used here with periods of one, five or seven years (Table 41).

Assumed HIV-negative period following negative HIV test	Specificity (%)	Z-test *p*-value compared with an assumed HIV-negative period of 3.75 years following a negative HIV test result
One year	63.6	0.239
3.75 years	54.8	Reference standard
Five years	55.0	0.964
Seven years	55.0	0.964

Table 41: Specificity of the original Lopman algorithm by HIV-negative period following a negative HIV test, Manicaland

ii. Investigating Lopman's assumption about the composition of the reference standard

Five of the 21 HIV-positive people classified as having died of obstetric causes or injuries had at least one symptom included in the algorithm: one person with reported suicide, one road traffic collision, one homicide, one unspecified injury and one woman with major bleeding in pregnancy or postpartum. Three had one symptom (one ARTI, one vaginal tumours, one wasting) and two had two symptoms (wasting and herpes; wasting and vaginal tumours).

Using a reference standard consisting only of HIV status, and thus not recategorising as "HIV-negative" those HIV-positive people with symptoms of obstetric causes or injuries, specificity and sensitivity were identical to those achieved using Lopman's reference standard: 52.4% and 80.0% respectively.

III. Re-deriving the Lopman algorithm using the present data

The 965 VA records with known HIV status were split into a training and a testing dataset: 717 (74.3%) in the training dataset and 248 (25.7%) in the testing dataset. The proportion of deaths in the training and testing datasets that were HIV/AIDS-related in the reference standard was respectively 78.9% and 77.0% (chi-squared p=0.525).

The likelihood ratios for all symptoms are shown in Appendix 12. Twelve symptoms had a likelihood ratio ≥1.92 and were eligible for inclusion in the algorithm. Figure 20 shows that both rash lasting at least one week, and any rash, were eligible. The likelihood ratio for "rash" was calculated after all deaths with reported rash lasting at least one week had already been classified as HIV/AIDS-related and removed from the dataset; therefore, "rash" means any reported rash either shorter than one week or of unknown duration. Abnormal hair colour was the first symptom the inclusion of which meant a loss of specificity greater than the gain in sensitivity. The cut-off for the algorithm was therefore made prior to abnormal hair colour, at rash, although fever lasting at least two weeks was slightly closer to the top-left of the plot. The algorithm was defined as indicating an HIV/AIDS-related death when any of the following seven symptoms was present: herpes zoster, wasting, rash lasting at least one week, productive cough, fever lasting at least two weeks, ulcers (not on feet) or rash (Figure 21, Figure 22, Table 42).

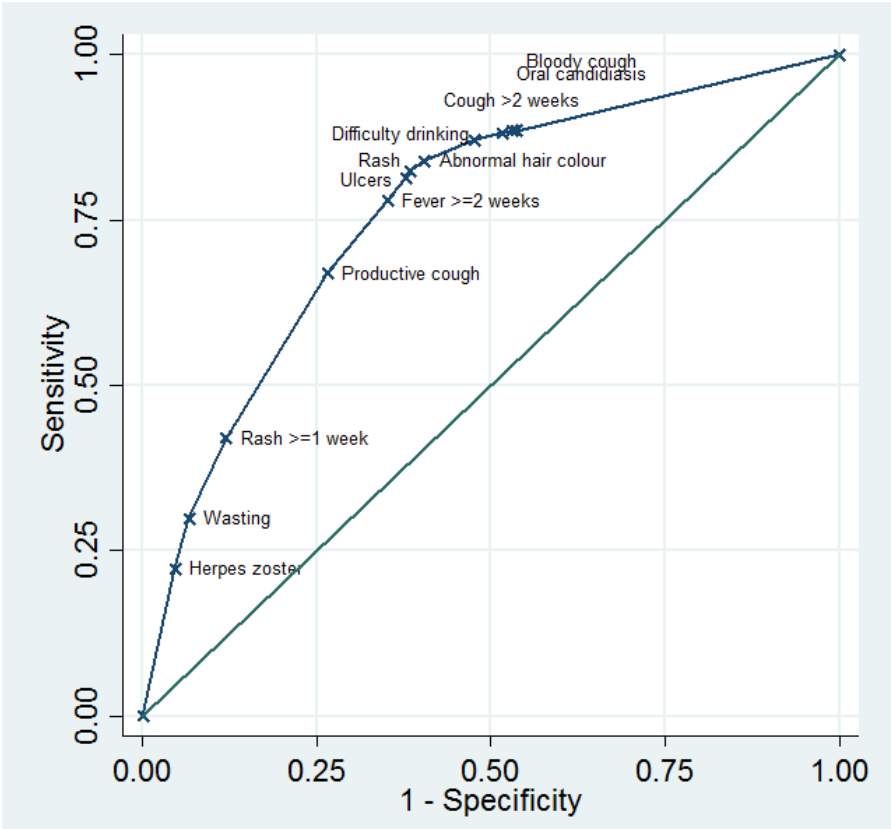

Figure 20: ROC curve with all eligible symptoms, Manicaland

In the training dataset, this algorithm achieved specificity of 61.6% (95% CI 53.6–69.0%) and sensitivity of 82.5% (95% CI 79.2–85.4%) (Table 43). It correctly classified 78.1% of deaths and classified 73.2% of deaths as HIV/AIDS-related. This was 5.7 percentage points lower than the 78.9% in the reference standard of the training dataset, a difference that was statistically significant (p=0.011). Applied to the testing dataset, the algorithm achieved specificity of 64.9% (95% CI 51.9–76.0%) and sensitivity of 77.0% (95% CI 70.49–82.4%) (Table 43). It correctly classified 74.2% of deaths and classified 67.3% of deaths as HIV/AIDS-related. This was 9.7 percentage points lower than the 77.0% in the reference standard of the testing dataset, a significant difference (p=0.016).

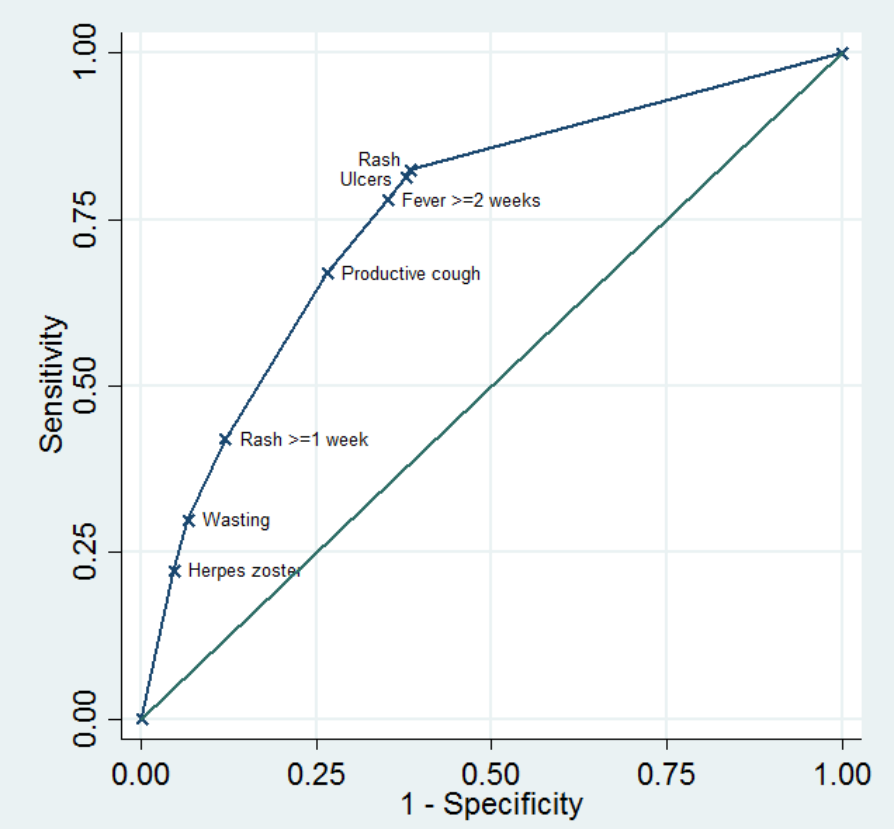

Figure 21: ROC curve of Lopman algorithm derived in the training dataset, with cut-off at "Rash"

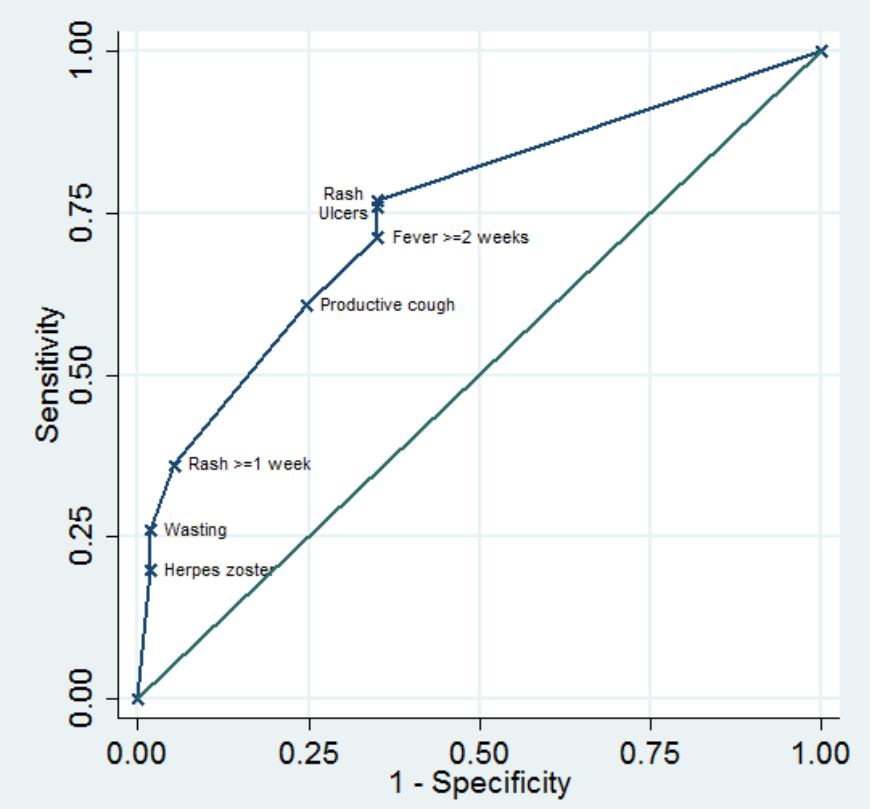

Figure 22: ROC curve of Lopman algorithm derived in the training dataset and applied to the testing dataset

Training dataset	N deaths remaining			N with symptom		Specificity and sensitivity when added	
Symptom	Total	HIV–	HIV+	HIV–	HIV+	Specificity (%)	Sensitivity (%)
Herpes	717	151	566	7	126	95	22
Wasting	584	144	440	3	43	98	10
Rash ≥1 week	538	141	397	8	69	94	17
Productive cough	461	133	328	22	141	83	43
Fever ≥2 weeks	298	111	187	13	62	88	33
Ulcers	223	98	125	4	19	96	15
Rash	200	94	106	1	7	99	7
Not classified as HIV/AIDS-related	192	93	99				

Testing dataset	N deaths remaining			N with symptom		Specificity and sensitivity when added	
Symptom	Total	HIV–	HIV+	HIV–	HIV+	Specificity (%)	Sensitivity (%)
Herpes	248	57	191	1	38	98	20
Wasting	209	56	153	0	12	100	8
Rash ≥1 week	197	56	141	2	19	96	13
Productive cough	176	54	122	11	47	80	39
Fever ≥2 weeks	118	43	75	6	20	86	27
Ulcers	92	37	55	0	9	100	16
Rash	83	37	46	0	2	100	4
Not classified as HIV/AIDS-related	81	37	44				

Table 42: The effects of individual symptoms on algorithm specificity and sensitivity, Manicaland, in the training and testing datasets

Dataset	Specificity %	Sensitivity %	% correctly classified	% assigned as HIV/AIDS-related	% HIV/AIDS-related in reference standard
Training	61.6	82.5	78.1	73.2	78.9
Testing	64.9	77.0	74.2	67.3	77.0

Table 43: Performance of Lopman algorithm in training and testing datasets in Manicaland

i. Sensitivity analysis of the length of the post-negative period

There was no significant difference in the specificity of the derived algorithm comparing the 3.75-year assumed HIV-negative period used here (specificity = 61.6%) with one year (specificity = 70.0%, p=0.327) (Table 44); there was no difference in the records classified as

HIV-negative using an assumed period of five or seven years, as nobody had an interval between HIV test and death of five years or more.

Assumed HIV-negative period following negative HIV test	Specificity (%)	Z-test *p*-value compared with an assumed HIV-negative period of 3.75 years following a negative HIV test result
One year	70.0	0.327
3.75 years	61.6	Reference category
Five years	61.6	No difference
Seven years	61.6	No difference

Table 44: Specificity of the Lopman algorithm by HIV-negative period following a negative HIV test, Manicaland

ii. Investigating Lopman's assumption about the composition of the reference standard

Three of the 21 HIV-positive people re-classified as HIV-negative due to having reported obstetric causes or injuries had one symptom included in the algorithm: one person with reported homicide also had had a fever for at least two weeks, and two people with unspecified injuries had respectively had a productive cough and herpes zoster.

Using a reference standard consisting only of HIV status, specificity and sensitivity in the training dataset were the same as those achieved with this algorithm using Lopman's reference standard: specificity was respectively 58.2% and 61.6% (p=0.561), and sensitivity was respectively 80.5% and 82.5% (p=0.369).

IV. Variability in the performance of the algorithm according to the composition of its training dataset

I derived 20 variants of the Lopman algorithm using the same method as used to derive the version used in the above analyses ("the version used"). Herpes zoster occurred in all 20 variants, while ulcers and wasting occurred in 19 and rash in 17 (Table 45). The number of symptoms in the variants ranged from six to 10, compared to seven in the version used. The number of eligible symptoms (those with LR≥1.92) also varied between variants, from 10 to 17 (Table 46).

The most common symptoms in the variant algorithms – herpes zoster, ulcers, wasting and rash – also occurred in the version used. Herpes zoster was always ranked first or second in the algorithm and wasting was almost always in the top three, while other symptoms had a range of ranks. The version used featured two symptoms that occurred less frequently across the variants: fever lasting at least two weeks (5/20) and productive cough (5/20). The version used and 7/20 variants included two rash symptoms, and only one variant contained no rash symptom. Herpes zoster, ulcers/abscesses and wasting also occurred in the original Lopman algorithm – although given that the original Lopman algorithm was trained on VA data from Manicaland, it is perhaps unsurprising that there is similarity between the present version used/variants and the original.

The specificities and sensitivities of the variant algorithms as applied to their respective training datasets differed widely, with specificity ranging from 62% to 73%, and sensitivity ranging from 73% to 84% (Figure 23). The proportion of deaths correctly classified ranged from 73% to 80%. The proportion assigned to HIV/AIDS ranged from 54% to 71% across the variants, and the absolute difference compared to the proportion of "true positives" in the reference standard ranged from seven to 23 percentage points; in all variant algorithms the proportion of deaths assigned to HIV/AIDS was highly significantly lower than the proportion due to HIV/AIDS in the reference standard (Table 46).

As per the methods for deriving the version used in the analyses above, I made the cut-off for the variant algorithms at the specificity cut-off, rather than the point closest to the upper left-hand corner of the plot. In 15/20 variants, these two conditions resulted in the same algorithm. In the other five variants, using the upper leftmost point resulted in fewer symptoms in the algorithm than using the specificity cut-off: either one fewer (four cases) or two fewer (one case). Using the specificity cut-off rather than the upper-leftmost point cut-off meant lower specificity, higher sensitivity and a slightly higher proportion of deaths correctly classified when applied to the training datasets across the variant algorithms where the different cut-offs resulted in different algorithms (Figure 24).

V. Summary

The version of the Lopman algorithm derived in the present data achieved specificity of 62% in the training dataset and 65% in the testing dataset, and sensitivity of 83% and 77%

respectively. It correctly classified 78% and 74% of deaths in the respective datasets, and classified 73% and 67% as due to HIV/AIDS. Across variants of the algorithm derived on the same dataset there was great variation in specificity, sensitivity, the proportion of deaths assigned to HIV/AIDS and the difference between this proportion and the proportion due to HIV/AIDS in the reference standard; across all variants the estimated proportion was significantly lower than the proportion in the reference standard, by a median of –14 percentage points.

Short name	Symptom	Frequency across 20 variants	Present in version used?
Herp	Herpes zoster	20	Yes
Ulce	Ulcers/ abscesses or sores on body, apart from feet	19	Yes
Wast	Wasting	19	Yes
Rash	Rash	17	Yes
OrCa	Oral candidiasis	14	
HIV	Medical diagnosis of HIV/AIDS	11	
DiaL	Diarrhea lasting at least four weeks	9	
Hair	Abnormal hair colouring	9	
RashL	Rash lasting at least one week	6	Yes
FevL	Fever lasting at least two weeks	5	Yes
Drink	Difficulty drinking	5	
ProCo	Productive cough	5	Yes
DiaU	Diarrhea of unknown duration	4	
RashU	Rash of unknown duration	3	
DiaM	Diarrhea lasting 2-4 weeks	2	
UriEx	Excessive urination	2	
UriRet	Urinary retention	1	

Table 45: Symptoms and symptom frequency across 20 variant Lopman algorithms, Manicaland

Variant no.	Twenty variant algorithms																				Version used
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	
Symptoms in algorithm	Herp	Herp	Herp	Herp	Herp	Herp	Herp	Herp	Herp	Herp	Rash	Herp	Herp	Herp	Herp	Wast	Herp	Wast	Wast	Herp	Herp
	Wast	Wast	Wast	Wast	Wast	Ulce	Wast	Wast	Wast	Wast	Herp	UriEx	Wast	Ulce	Wast	Herp	Wast	Herp	Herp	Wast	Wast
	Ulce	Ulce	RashL	Hair	Hair	Wast	RashL	UriRet	Ulce	Ulce	Ulce	Wast	RashL	Rash	RashL	Ulce	Ulce	Ulce	UriEx	Rash	RashL
	RashU	Rash	OrCa	Ulce	DiaU	DiaL	Ulce	Hair	RashU	RashU	OrCa	RashL	Ulce	***HIV***	OrCa	Rash	DiaL	Rash	RashL	Ulce	ProCo
	DiaL	***HIV***	Ulce	Rash	Ulce	***HIV***	Rash	DiaU	Rash	Rash	DiaL	ProCo	Rash	OrCa	Ulce	DiaL	***HIV***	OrCa	FevL	OrCa	FevL
	Rash	Hair	Rash	DiaL	Rash	Rash	OrCa	***HIV***	***HIV***	Hair	Hair	Ulce	OrCa	FevL	Rash	***HIV***	Rash	Drink	OrCa	Drink	Ulce
	HIV	DiaU	DiaL	FevL	DiaL	Hair	Drink	ProCo	DiaL	ProCo	DiaU	OrCa	ProCo	Wast	Drink	OrCa					Rash
	Hair	OrCa	***HIV***	DiaM	OrCa	FevL	ProCo	Ulce	OrCa												
	FevL	Drink	Hair	***HIV***	***HIV***																
	DiaM																				
# symptoms	10	9	9	9	9	8	8	8	8	7	7	7	7	7	7	7	6	6	6	6	7
# eligible symptoms	15	12	16	15	16	16	17	15	13	12	15	14	12	11	11	13	14	12	17	10	12
% specificity	65	66	70	67	71	64	63	68	73	69	72	68	66	69	73	69	72	72	69	71	62
% sensitivity	84	78	79	83	78	82	83	77	76	77	78	77	78	79	74	76	76	73	75	74	83
% correctly classified	80	75	77	80	76	78	78	75	76	75	77	75	75	77	74	75	75	73	74	73	78
% assigned to HIV/AIDS	71	65	65	71	64	70	70	63	60	63	63	64	64	66	60	62	54	59	65	60	73
% “true” HIV/AIDS	78	79	78	79	78	79	78	79	76	78	78	78	77	79	79	78	78	78	77	78	79
Difference, pp	–7	–14	–13	–8	–14	–9	–8	–16	–16	–15	–15	–14	–13	–14	–18	–16	–23	–19	–12	–18	–6
p for difference	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01	0.011

Table 46: Composition and performance of 20 random variant Lopman algorithms, and the version used in the present analyses; specificity, sensitivity and % correctly classified, % assigned HIV/AIDS and comparison with % HIV/AIDS in reference standard, Manicaland. pp=percentage points

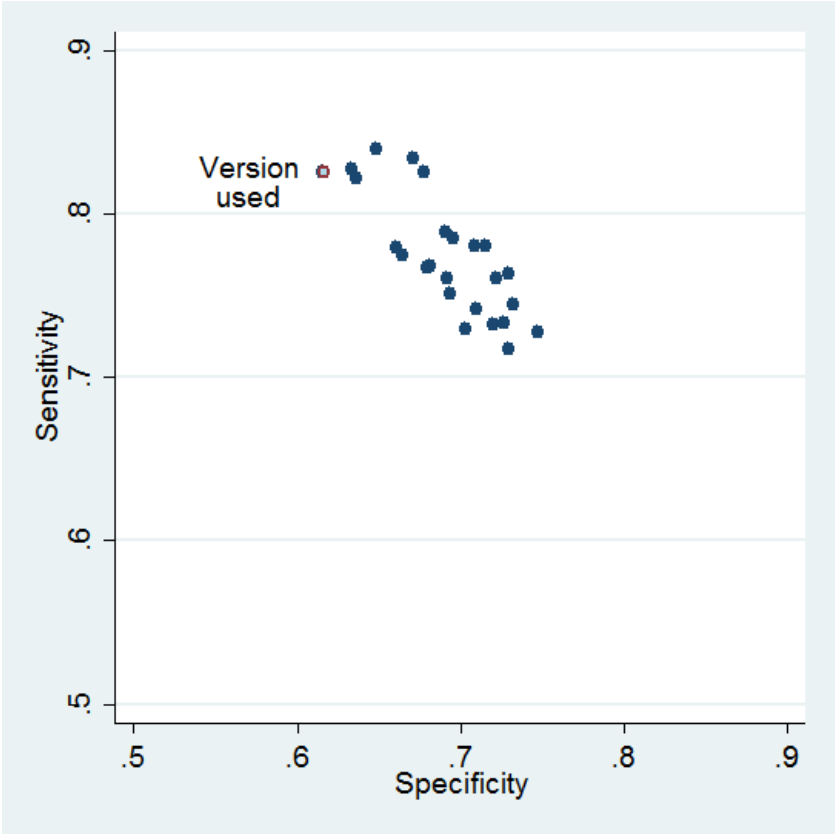

Figure 23: Specificities and sensitivities of 20 variant Lopman algorithms and the version used, in their respective training datasets, Manicaland

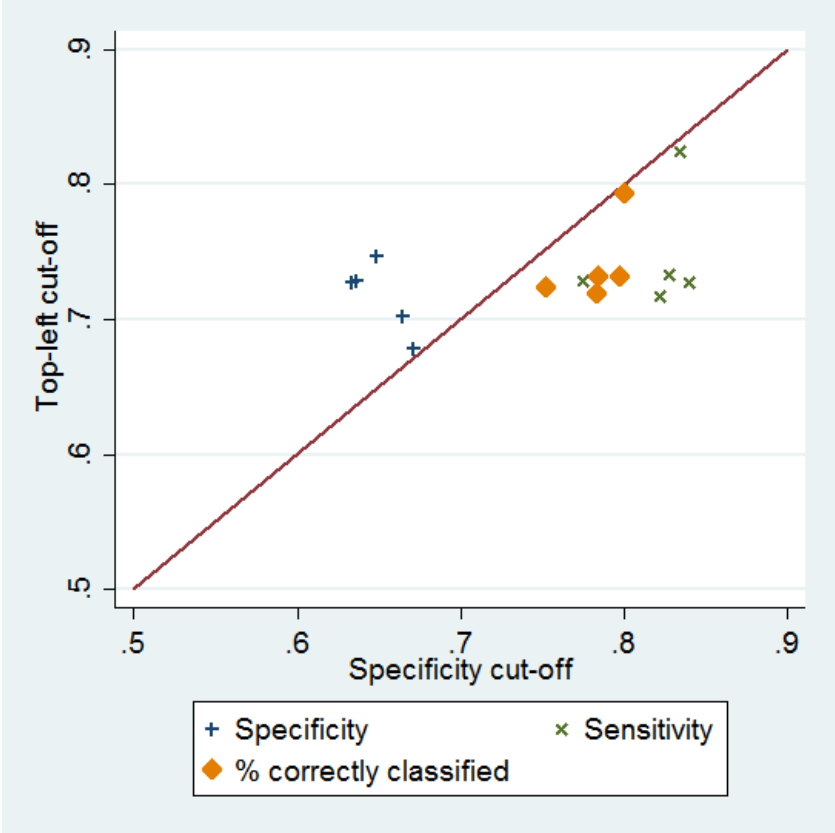

Figure 24: Specificity, sensitivity and % correctly classified in training datasets for variant Lopman algorithms according to whether the specificity cut-off or the upper-leftmost point cut-off was used, Manicaland

6. Applying the Kisesa-derived and Manicaland-derived algorithms in Manicaland and Kisesa, respectively

Applying the Kisesa-derived algorithm to 965 deaths with known HIV status from Manicaland (assuming a post-negative period of 3.75 years) achieved specificity of 67.8% and sensitivity of 70.1%. The proportion correctly classified was 69.6%, and 62.0% of deaths were assigned to HIV/AIDS, which was 16.5 percentage points lower than the reference standard proportion of 78.4%.

Applying the Manicaland-derived algorithm to 598 deaths with known HIV status from Kisesa (assuming a post-negative period of five years) achieved specificity of 61.4% and sensitivity of 71.4%. The proportion correctly classified was 65.9%, and 53.2% of deaths were assigned to HIV/AIDS, which was 8.7 percentage points higher than the reference standard proportion of 44.5%.

7. Discussion

This chapter investigated the validity of the Lopman algorithm, both of its previously developed version and of versions derived in the present data, and allows some conclusions about the quality of this method for ascertaining HIV/AIDS-related mortality from VA data. Specificity of the Lopman algorithm derived in the present dataset was fairly low in Kisesa and very low in Manicaland. The original Lopman algorithm also achieved very poor specificity in the Kisesa and Manicaland datasets. The large variation in the performance of the algorithm across variants derived using identical methods illustrates an inherent problem of data-derived methods as an automated alternative to physician review.

I. Findings of other studies

The Lopman algorithm has not appeared in the published literature since its original development and later validation by Lopman and colleagues. Those papers therefore provide the only direct comparison for the present performance of the method. The specificity of the present version of the algorithm in Kisesa (80% in training, 74% in testing) was virtually identical to that in the original derivation by Lopman and colleagues (78% in training, 76% in testing); in the present version of the algorithm in Manicaland specificity was lower (62% in

training, 65% in testing). Sensitivity in the original application was 71% in training and 66% in testing, somewhat higher than in the data from Kisesa (65% in training, 49% in testing) but lower than in Manicaland (83% in training, 77% in testing).

II. Variation in estimates

The similarity in validity between the original application of the Lopman algorithm and its first derivation in the Kisesa dataset was due to chance: variants of the algorithm, derived by exactly the same method, had a wide range of specificities and other measures in both settings, determined by the randomly generated composition of their respective training and testing datasets. One of the stated advantages of automated methods over physician review is their reliability[65]. The central importance of random-number allocation in the Lopman algorithm introduces unreliability. This unreliability differs from that encountered in the changeable outcomes of physician review, in that it is not biased. Nonetheless, its arbitrary selection of symptoms and great variation in specificity and sensitivity raise doubt over the advisability of using this method as currently conceived as a means of ascertaining HIV/AIDS-related mortality across time and location, and even undermine confidence in its results in a single time and place.

Lopman and colleagues stated "A classification system with known sensitivity and specificity has practical applications" (2006: 1277[172]). The present analysis suggests that it is misleading to describe the Lopman algorithm as having "known sensitivity and specificity", due both to the difference between the performance of the original algorithm applied in their original data and in the present data, and to the demonstrated variability in composition, specificity and sensitivity of algorithms derived from the present dataset.

There were substantial differences in symptom composition between the original Lopman algorithm, the version used in the present data and the variants derived; there were also great differences in the performance of the original algorithm in the original and the present data, and of the Kisesa-derived algorithm in the Manicaland data and vice versa. The original Lopman algorithm applied to the present data had poor specificity (66% in Kisesa, 55% in Manicaland), values much lower than those it achieved in its original application (78% in training and 76% in testing). The performances of the Manicaland version in the Kisesa data and vice versa were more or less consistent with the findings across variants internal to those datasets. This suggests that deriving an algorithm consisting of a set of symptoms in one

context and applying that algorithm to ascertain HIV/AIDS-related mortality in another context may not necessarily give worse results than deriving the algorithm anew, with reference-standard data, in every application. It is nonetheless not a ringing endorsement, given that the similarity includes overall low levels of specificity and sensitivity and large absolute errors.

III. Selection of the cut-off for the Lopman algorithm

Lopman and colleagues present two conditions for selecting the cut-off point of the algorithm: the symptom closest to the upper left-hand corner of the ROC plot, and the symptom that occurs prior to the one whose inclusion means a loss of specificity greater than the gain in sensitivity. These conditions are presented as interchangeable in the original publications, but are not in fact the same. In the variants presented for analysis, I used the specificity cut-off, as Lopman and colleagues prioritised specificity in their calculation of likelihood ratios. In half of the 20 variants in Kisesa, the specificity cut-off gave a shorter algorithm than would have been given by the upper left-hand corner method, while the reverse was true in a quarter of the variants in Manicaland, with the specificity cut-off resulting in a longer algorithm. In Kisesa and Manicaland, the different cut-offs led to notable differences in sensitivity and specificity.

IV. Validation and the choice of reference standard

The analyses in this chapter differ from the analyses in previous chapters, due to the definition of "true positive" and "false negative" deaths as well as "true negative" and "false positive" deaths. This allowed calculation of sensitivity, the percentage correctly classified and the absolute classification error. The Lopman algorithm requires sensitivity in order to calculate likelihood ratios for symptoms – it cannot be constructed using specificity alone, as the most specific symptoms are those occurring in no or few HIV-negative people, and as many of these symptoms also occur in no or few HIV-positive people, the resulting algorithm would both be highly inefficient and contain many symptoms that occur too infrequently for statements about their association with HIV status to be valid. While widespread access to HIV-status data may be easier to achieve than widespread medical certification of causes of death, the requirement for these data does limit the potential application of the Lopman algorithm.

It is notable that the version of the Lopman algorithm used in Manicaland that came closest to accurately estimating the proportion of HIV/AIDS-related deaths in the reference standard was the version with the lowest specificity. That value, 62%, is far below any of the thresholds

suggested for minimum acceptable specificity[66, 98, 126]. This finding encourages caution toward specificity as a useful measure of validation.

It is an inherent limitation of the HIV-status reference standard that HIV-positive people can die of causes other than HIV/AIDS. Lopman and colleagues attempted to address this with their classification of HIV-positive people for whom obstetric causes or injuries are reported as having suffered non-HIV-associated deaths. The present results show several HIV/AIDS-related symptoms for people with reported injuries or obstetric symptoms. HIV/AIDS could be the underlying cause of a death of which the immediate cause was suicide, but this cannot be told from the VA data available (especially without the narrative section of the VA interview, in which the respondent might elaborate on the mental state of the deceased, or potentially on the reasons for their suicide). It is also possible that the reports of "excessive" bleeding do not necessarily indicate bleeding of a scale sufficient to constitute a fatal obstetric haemorrhage, but rather bleeding that seemed "excessive" to the respondent, alongside fatal HIV/AIDS-related diseases – it is known that where HIV/AIDS is a major cause of disease and death, HIV-positive pregnant women die of AIDS and of non-AIDS-defining infections at a much higher rate than HIV-negative women, and a substantial minority of these suffer symptoms of obstetric complications as well[33]. Definitions of criteria for HIV/AIDS-related disease as a cause of pregnancy-related mortality are rarely reported in cause-specific pregnancy-related mortality studies[51].

As outlined in the introduction to this thesis, this is a question of whether we are attempting to ascertain deaths **from** HIV/AIDS, or to ascertain which deaths were of people **with** HIV/AIDS. Beyond injuries and obvious direct obstetric causes of death, the validity of including HIV-positive people in the denominator for specificity calculations becomes much more uncertain: the fundamental distinction is that HIV-negative people are not at risk of death from HIV/AIDS, while HIV-positive people are. This demands that even when the mortality profile of HIV-positive people is substantially changed by ART, caution is exercised in treating HIV-positive ART users as similar to HIV-negative people in reporting mortality, at least until the mortality profile of HIV-positive ART users is longer established and better understood. Of course, it is to be hoped that the primary effect of ART is to extend the life of HIV-positive people, and that the latter would consequently become less relevant to research with a focus on causes of death among younger adults (such as the 15–59 year-olds in the present work).

Using a relatively crude reference standard such as that used by Lopman and colleagues serves to ascertain deaths with HIV/AIDS, removing only those HIV-positive people who have very likely other causes of death. This might suggest that the findings of the proportion of population deaths ascertained would be greater than the proportion truly from HIV/AIDS, which makes it notable that all Manicaland variants and the version used, and more than a quarter of variants in Kisesa, significantly underestimated the proportion of true positive deaths, by up to 23 percentage points.

V. Limitations

Lopman and colleagues calculated estimated misclassification of deaths of HIV-positive people as HIV/AIDS-related based on the population underlying hazard of death – I was unable to do this with the present data, which is a limitation of this analysis.

I used less stringent definitions of weight loss and wasting, which would decrease the likelihood ratio if it increased the proportion of deaths of HIV-negative people in which the symptom occurred. The present dataset from Manicaland overlaps with that data on which the algorithm was trained in the original publication. However, this might be expected to over-estimate the similarity in the performance of the method between the two applications. In practice the algorithm originally derived and the version used in Manicaland had a minority of symptoms in common, suggesting this limitation did not cause important bias.

VI. Conclusion

Given that the procedure for deriving the algorithm is the same each time and that variation is only due to random differences in the composition of the training dataset, there is no valid way of deciding between variants. This is true also of the "original Lopman algorithm" used in this chapter, which can only be said to comprise one instance of the performance of the Lopman algorithm in that original dataset. Even disregarding the low specificity of the algorithm derived in the present dataset, the variation in performance suggests that as it stands, the Lopman algorithm should not be used to draw conclusions about the proportion of mortality due to HIV/AIDS. The variation, and the findings with regard to the cut-off, suggest that at the very least, the method would need to be more rigorously conceptualised before being used to measure HIV/AIDS-related mortality.

6. Assessment of potential bias

1. Introduction

One advantage of data from demographic surveillance systems is that these sites are designed to represent or constitute a population and have both internal validity and enough external validity to allow them to be the basis for useful epidemiological studies[197, 198]. However, data collection is imperfect and the data available for the present analyses are subject to probable biases, the risks of only some of which can be estimated.

Not every death is followed by a VA interview, and not every VA record receives a physician review, or contains the necessary information to allow InterVA to assign a cause of death. This chapter investigates potential selection biases for deaths of people more or less likely to have truly died of HIV/AIDS, that might lead to an over- or under-estimation of the proportion of deaths due to HIV/AIDS. To do this, I assessed potential selection bias in which deaths received a VA interview, which affects the composition of the datasets overall and therefore the results given by all methods of interpreting VA data. I also assessed potential selection bias in which VA records received a physician review and which records had a cause of death assigned by the InterVA model.

2. Assessment of potential selection bias

In order to identify the scale and direction of possible selection bias, I investigated whether deaths of people with particular characteristics were more likely to be included in the analyses. The characteristics investigated were:

- age, because HIV incidence and attributable mortality can vary across age groups[24, 199, 200];
- sex, because substantially higher attributable mortality to HIV/AIDS has been reported in women than in men[17, 24], although this may not be true in Manicaland[168];
- urban versus rural residence, because HIV prevalence and incidence have been seen to be higher in urban than rural areas[160, 201, 202];
- year of death, reflecting whether anti-retroviral therapy was available, because ART availability affects the level of mortality due to HIV/AIDS[29, 155]; and
- HIV status, because the proportion of deaths due to HIV/AIDS will be affected if a greater proportion of deaths of HIV-positive or HIV-negative people are included.

Chi-squared tests were used to assess the association between records being available for analyses and these characteristics. I summarised the overall risk of selection bias as high, medium or low risk of upward or downward bias for each characteristic, taking account of both the association and the magnitude of the differences between groups: a difference may be statistically significant but be too small to warrant concern that it introduces much bias in the proportion of deaths in the dataset that were truly due to HIV/AIDS.

3. Assessment of potential selection bias in Kisesa

I. Potential selection bias in Kisesa in which deaths received verbal autopsy

Table 47 shows the distribution of 1837 deaths eligible for verbal autopsy interview by categories of other variables; 1246 of these deaths (68%) received a verbal autopsy interview. The proportion of deaths receiving a VA interview varied by age, from 58.8% in 40-44 year-olds to 77.8% among 50-54 year-olds. There was no pattern by age. There was no difference by

sex. Rural residents had a greater probability of receiving a VA after death (71.0%), than peri-urban residents (67.4%) or urban residents (62.2%) (p=0.003). Few deaths with no year of death recorded received a VA interview (10.2%), while coverage among those who died from 1994 to 2004 (73.6%) was lower than among those who died in 2005–2011 (82.1%) (p<0.001). There was no difference in the proportion of deaths of HIV-negative (71.9%) and HIV-positive people (72.1%) that received a VA interview (p=0.952). Deaths of people with known HIV status were significantly more likely to receive a VA interview than those without (72.0% vs 64.4%, p=0.001).

Variable		**Received VA interview**				**X^2 p-value**
		Yes		**No**		
Age group	15–19	59.0%	(85/144)	41.0%	(59/144)	
	20–24	66.0%	(126/191)	34.0%	(65/191)	
	25–29	65.3%	(169/259)	34.7%	(90/259)	
	30–34	69.4%	(202/291)	30.6%	(89/291)	
	35–39	70.9%	(175/247)	29.1%	(72/247)	0.001
	40–44	58.6%	(126/215)	41.4%	(89/215)	
	45–49	74.2%	(132/178)	25.8%	(46/178)	
	50–54	77.8%	(130/167)	22.2%	(37/167)	
	55–59	69.7%	(101/145)	30.3%	(44/145)	
Sex	Men	67.5%	(647/959)	32.5%	(312/959)	0.729
	Women	68.2%	(599/878)	31.8%	(279/878)	
Residence	Rural	71.0%	(669/942)	29.0%	(273/942)	
	Peri-urban	67.3%	(264/392)	32.7%	(128/392)	0.003
	Urban	62.2%	(313/503)	37.8%	(190/503)	
Year of death	1994–2004	73.6%	(746/1014)	26.4%	(268/1014)	
	2005–2011	82.1%	(476/580)	23.8%	(104/580)	<0.001
	Unknown year	9.9%	(24/243)	90.1%	(219/243)	
HIV status at death	HIV-negative	71.9%	(322/448)	28.1%	(126/448)	0.952
	HIV-positive	72.1%	(276/383)	27.9%	(107/383)	
	Known HIV+/–	72.0%	(598/831)	28.0%	(233/831)	0.001
	Unknown	64.4%	(648/1006)	35.6%	(358/1006)	
Total		**67.8%**	**(1246/1837)**	**32.2%**	**(591/1837)**	**–**

Table 47: Distribution of deaths of people resident in the Kisesa DSS area, by whether they received a VA interview, and other variables

II. Potential selection bias in Kisesa in which VA records had cause of death assigned by physician review

Of 1246 VA records from Kisesa, 462 (37%) had cause of death assigned by physician review (Table 48). The proportion of deaths with a physician-assigned cause was broadly greater in older compared to younger people, but the proportion in most five-year age groups was similar and there was no discernible pattern. Excluding 50–54 year-olds (10.4% of the records), among whom coverage was substantially higher, the difference between age groups was insignificant (p=0.632). The differences in men compared to women and rural compared to urban residents were small and not significant (p=0.285 and 0.094, respectively).

The VA records for which there are physician-assigned causes of death were largely from the period 2005–2011 when ART was available: VA records from this period were far more likely to have had a physician review, compared to those from the ART-naïve period until 2004; records from 2005–2011 comprised 87.7% of the total VA records with physician review (405/462). The proportion of VA records with physician-assigned cause of death was higher among HIV-negative (47.8%) than HIV-positive people (38.0%, p=0.016). Deaths of people with known HIV status were more likely than those with unknown HIV status to receive a physician review (43.3% vs 31.3%, p<0.001).

Variable		VA had cause of death assigned by physician				X^2 *p*-value
		Yes		No		
Age group	15–19	35.3%	(30/85)	64.7%	(55/85)	
	20–24	35.7%	(45/126)	64.3%	(81/126)	
	25–29	33.1%	(56/169)	66.9%	(113/169)	
	30–34	32.2%	(65/202)	67.8%	(137/202)	
	35–39	35.4%	(62/175)	64.6%	(113/175)	0.032
	40–44	40.5%	(51/126)	59.5%	(75/126)	
	45–49	33.3%	(44/132)	66.7%	(88/132)	
	50–54	50.8%	(66/130)	49.2%	(64/130)	
	55–59	42.6%	(43/101)	57.4%	(58/101)	
Sex	Men	38.5%	(249/647)	61.5%	(398/647)	0.285
	Women	35.6%	(213/599)	64.4%	(386/599)	
Residence	Rural	38.9%	(260/669)	61.1%	(409/669)	
	Peri-urban	38.5%	(102/265)	61.5%	(163/265)	0.094
	Urban	31.9%	(100/313)	68.1%	(213/313)	
Year of death	1994–2004	5.1%	(38/746)	94.9%	(708/746)	
	2005–2011	85.1%	(405/476)	14.9%	(71/476)	<0.001
	Unknown year	79.2%	(19/24)	20.8%	(5/24)	
HIV status at death	HIV-negative	47.8%	(154/322)	52.2%	(168/322)	0.016
	HIV-positive	38.0%	(105/276)	62.0%	(171/276)	
	Known HIV+/–	43.3%	(259/598)	56.7%	(339/598)	<0.001
	Unknown	31.3%	(203/648)	68.7%	(445/648)	
Total		**37.1%**	**(462/1246)**	**62.9%**	**(784/1246)**	**–**

Table 48: Distribution of VA records with physician-assigned cause of death, by categories of other variables

III. Potential selection bias in Kisesa in which VA records had cause of death assigned by InterVA

Of 1246 VA records from Kisesa, 1107 (89%) had cause of death assigned by InterVA (Table 49). There was no difference in the proportion of records receiving a cause of death by InterVA across categories of age (*p*=0.997), or year of death (p=0.403). Women were slightly more likely than men to receive a cause of death by InterVA (90.8% vs 86.7%, *p*=0.026), and rural residents (93.0%) were more likely than either peri-urban (83.4%) or urban residents (84.3%, *p*<0.001). There was no difference in the proportion of deaths assigned cause by InterVA comparing HIV-negative with HIV-positive people (90.4% vs 90.6%, *p*=0.932) or comparing those with known HIV status to those without (90.5% vs 87.3%, *p*=0.080).

Variable		**VA had cause of death assigned by InterVA**				**X^2 *p*-value**
		Yes		**No**		
Age group	15–19	89.4%	(76/85)	10.6%	(9/85)	
	20–24	90.5%	(114/126)	9.5%	(12/126)	
	25–29	88.2%	(149/169)	11.8%	(20/169)	
	30–34	88.1%	(178/202)	11.9%	(24/202)	
	35–39	89.7%	(157/175)	10.3%	(18/175)	0.996
	40–44	89.7%	(113/126)	10.3%	(13/126)	
	45–49	87.1%	(115/132)	12.3%	(17/132)	
	50–54	89.2%	(116/130)	10.8%	(14/130)	
	55–59	88.1%	(89/101)	11.9%	(12/101)	
Sex	Men	86.9%	(562/647)	13.1%	(85/647)	0.021
	Women	91.0%	(545/599)	9.0%	(54/599)	
Residence	Rural	93.0%	(622/669)	7.0%	(47/669)	
	Peri-urban	83.7%	(221/264)	16.3%	(43/264)	<0.001
	Urban	84.3%	(264/313)	15.7%	(49/313)	
Year of death	1994–2004	88.1%	(657/746)	11.9%	(89/746)	
	2005–2011	90.1%	(429/476)	9.9%	(47/476)	0.526
	Unknown year	87.5%	(21/24)	12.5%	(3/24)	
HIV status at death	HIV-negative	90.4%	(291/322)	9.6%	(31/322)	0.932
	HIV-positive	90.6%	(250/276)	9.4%	(26/276)	
	Known HIV+/–	90.5%	(541/598)	9.5%	(57/598)	0.080
	Unknown	87.3%	(566/648)	12.7%	(82/648)	
Total		**88.8%**	**(1107/1246)**	**11.2%**	**(139/1246)**	**–**

Table 49: Distribution of VA records by whether InterVA assigned a cause of death, and by categories of other variables

4. Assessment of potential selection bias in Manicaland

I. Potential selection bias in Manicaland in which deaths received verbal autopsy

Table 50 shows the distribution of 3155 deaths in the Manicaland DSS area by other variables; 1021 deaths (32%) received a VA interview. Coverage of VA interviews by age varied from below 20% among 15–19 year-olds and 55–59 year-olds, up to 43.1% among 35–39 year-olds (p<0.001). Women were more likely to have a VA than men (34.7% vs 29.2%, p=0.001); there was no difference by residence. People who died in the period 1998–2005 before ART was available were more likely to receive a VA interview than those who died in 2006–2011 when ART was available (44.7% vs 18.7%, p<0.001). HIV-positive people were substantially more likely to receive a VA interview than HIV-negative people (64.7% vs 38.2%, p<0.001). Deaths of people with known HIV status were much more likely to receive an interview than those of people with unknown HIV status (57.0% vs 3.8%, p<0.001).

Variable		**Received VA interview**		**X^2 p-value**
		Yes	**No**	
Age group	15–19	17.4% (32/184)	82.6% (152/184)	<0.001
	20–24	24.3% (79/325)	75.7% (246/325)	
	25–29	29.2% (133/456)	70.8% (323/456)	
	30–34	37.2% (184/495)	62.8% (311/495)	
	35–39	43.1% (209/485)	56.9% (276/485)	
	40–44	37.6% (150/399)	62.4% (249/399)	
	45–49	35.5% (123/346)	64.5% (223/346)	
	50–54	33.2% (81/244)	66.8% (163/244)	
	55–59	13.6% (30/221)	86.4% (191/221)	
Sex	Men	29.2% (399/1365)	70.8% (966/1365)	0.001
	Women	34.7% (622/1390)	65.3% (1168/1790)	
Residence	Subsistence farming	31.2% (375/1203)	68.8% (828/1203)	0.564
	Roadside trading	32.4% (213/658)	67.6% (445/658)	
	Agricultural estates	32.7% (264/807)	67.3% (543/807)	
	Commercial centres	34.7% (169/487)	65.3% (318/487)	
Year of death	1998–2005	44.7% (740/1654)	55.3% (914/1654)	<0.001
	2006–2011	18.7% (281/1501)	81.3% (1220/1501)	
HIV status at death	HIV-negative	38.2% (187/489)	61.8% (302/489)	<0.001
	HIV-positive	64.7% (778/1203)	35.3% (425/727)	
	Known HIV+/–	57.0% (965/1692)	43.0% (727/1692)	<0.001
	Unknown	3.8% (56/1465/)	96.2% (1407/1463)	
Total		**32.4% (1021/3155)**	**67.6% (2134/3155)**	**–**

Table 50: Distribution of deaths in the Manicaland DSS area, by whether they received a VA interview

II. Potential selection bias in Manicaland in which VA records had cause of death assigned by physician review

As stated under "Data sources and study population" in the General Methods section above, the physician review data from Manicaland were in a format inconsistent with that needed for the physician review analyses, meaning this analysis was not carried out.

III. Potential selection bias in Manicaland in which VA records had a cause assigned by InterVA

Of 1021 deaths for which a VA was conducted, 1016 (99.5%) were assigned a cause by InterVA. I did not consider it meaningful to investigate bias in the remaining five records.

5. Summary of potential selection bias by characteristics of the deceased

Almost all potential selection biases investigated in Kisesa were too small to present any concern over bias affecting the results, even where differences were statistically significant. For example, the proportion of deaths in Kisesa that received a VA was 87% for men and 91% for women, which is statistically significant (p=0.021) but unlikely to be a large enough difference to bias any estimates (Table 51). In Kisesa overall, there may be a slight underestimation of the contribution of HIV to mortality due to bias in the area of residence and year of death. There is also a small risk of downward bias in the proportion of deaths due to HIV in the InterVA analysis due to area of residence.

In Manicaland overall, there is a substantial risk of overestimation of the proportion of deaths due to HIV, due primarily to there being more VA records available for HIV-positive people, but also because of greater VA coverage in the ART-naïve period, among women and among people in the age groups with highest HIV-related mortality. The greater availability of VA records for HIV-positive people also means that the whole-population estimates of mortality due to HIV/AIDS assigned by the Lopman algorithm may be inflated.

Area	Potential source of selection bias	Whether death received VA	Whether VA received cause of death by physician review	Whether VA received cause of death by InterVA
Kisesa	Age	↕	–	–
	Sex	–	–	–
	Residence	↓	–	↓
	Year of death	↓	↓↓↓	–
	HIV status	–	↓	–
Manicaland	Age	↑	NI	NI
	Sex	↑	NI	NI
	Residence	–	NI	NI
	Year of death	↑↑	NI	NI
	HIV status	↑↑↑	NI	NI

–	No risk of bias
↑ / ↑↑ / ↑↑↑	Low/moderate/high risk of upward bias
↓ / ↓↓ / ↓↓↓	Low/moderate/high risk of downward bias
↕ / ↕↕ / ↕↕↕	Low/moderate/high risk of bias, direction unclear
NI	Not investigated

Table 51: Assessing the risk of selection bias in estimating the proportion of deaths due to HIV

6. Biases the risk of which could not be assessed

Any selection bias in the coverage of verbal autopsy potentially affects the accuracy of the estimates of cause-specific mortality distributions and the proportion of deaths assigned to HIV/AIDS. Specificity, by contrast, is unlikely to be biased by selection in the population receiving VA: such a bias would require differential likelihood of HIV/AIDS being assigned as the cause of death of an HIV-negative person according to whether their HIV status was known; this seems unlikely but, not knowing the distribution of HIV-negative and HIV-positive people in the group of people with unknown HIV status, I cannot assess the risk of this bias. In the distribution of causes of death by HIV status assigned by Physician Review and InterVA, the group with unknown status was intermediate between the HIV-negative and HIV-positive groups, implying a similar mix of HIV status among those with unknown status as among those with known status. In any case, with regard to potential selection bias in the estimates of specificity, it is worth noting that the results in the Lopman algorithm chapter suggest that the external validity of specificity estimates is highly limited.

There are factors that could have contributed to selection bias of which no measure is available. For example, we do not have any measures of formal educational level or socioeconomic status; both of these factors have shown associations with incidence or prevalence of HIV, although findings have been mixed in the direction and magnitude of their effects[203-207]. It has been seen that ART uptake is affected by distance to ART treatment centres[208], of which there is no measure in the present study.

The inherent imprecision of VA is a source of potential information biases, such as those due to imperfect understanding of questions or inaccurate recording of answers. In discussing the use of VA to assess trends in cause-specific mortality, Herbst and colleagues note that "Data quality […] could not be monitored on an ongoing basis" (Herbst et al 2011: 9[32]) and that changes in data quality could affect the results. Information biases due to poor quality of VA data cannot be investigated, as there is no valid method of assessing the quality of VA data.

I. Representativity of the study settings

As is shown in Table 4 and Table 5 in the General Methods chapter, there are some differences between the study settings and the countries in which they are located. Compared to Tanzania as a whole, Kisesa has higher HIV prevalence, higher fertility and a lower death rate, but similar rural/urban composition. Compared to Zimbabwe as a whole, Manicaland has higher fertility but a similar death rate and HIV prevalence, and it is much more rural (83% vs 67%). There are not national COD estimates to compare with the DSS estimates. Validation metrics would almost certainly differ nationally compared with DSS estimates, given that validation metrics vary by changes in the underlying mortality distribution.

Both Kisesa and Manicaland are open, geographically defined cohorts with high participation levels, which ought to remain representative of the areas where they are located. DSS estimates do not represent the country in which they are derived, or even all notionally similar parts of the country (e.g. rural areas)[66, 133]; nonetheless, the results derived from DSS research are informative for wider populations and should not be treated as only having local relevance[3].

7. Conclusion

Among the potential sources of selection bias presented in this chapter, there is overall little risk of bias in the Kisesa dataset. By contrast, the Manicaland data have substantial risk of upward bias in the overall proportion of deaths assigned to HIV/AIDS. Reporting causes of death by HIV status therefore not only provides useful information on the mortality differences between HIV-negative and HIV-positive people, but also avoids the potential biasing of estimates through differential inclusion of people by HIV status.

7. Discussion

The three interpretation methods estimated the proportion of mortality due to HIV/AIDS among 15–59 year-olds to be 30–53% in Kisesa, and 58–73% in Manicaland. Specificities for estimating HIV/AIDS-related mortality ranged from 61–88% in Kisesa and 62–68% in Manicaland. There was no clear relationship between specificity and the estimated proportion of HIV/AIDS-related mortality, and it was impossible to judge which estimate was closest to the true proportion of HIV/AIDS-related mortality. Reasons for the variation, and the potential of validation for assessing the performance of verbal autopsy methods, are discussed below.

1. Results with regard to existing literature

Figure 25 plots the point estimates and confidence intervals of the proportion of deaths assigned to HIV/AIDS in Kisesa and Manicaland by the interpretation methods presented in the results chapters. Across all interpretation methods, the HIV/AIDS-related CSMF in Manicaland was clearly higher than it was by any method in Kisesa, which is consistent with the magnitude of the epidemic in the respective settings. The overall HIV/AIDS-related CSMF in Kisesa ranged from 30% to 53% across the methods, with relatively little risk of selection bias in the data. This is higher than the CSMFs in Rufiji in Tanzania (1999–2002) and in Ifakara in Tanzania

(2000–02), both rural DSSes, where the proportions of deaths of people aged 15+ assigned to HIV by physician review of VA were 13.5% and 11.6% respectively[141]. The inclusion of people aged 60+ in those two Tanzanian studies would have reduced the HIV/AIDS-related CSMF somewhat, and the settings are different, meaning there is no reason to expect close correspondence of estimates. The present estimates for Kisesa are very similar to the 35% estimated by physician review of VA for deaths in Kisesa between 1994 and 1996[18]. In 2010 Lopman and colleagues[87] reported that in the Kisesa testing dataset of 15–44 year-olds, the true proportion of deaths due to HIV/AIDS was 51%; this is at the high end of estimates for 15–59 year-olds in the present data, as would be expected given the ages and the definition of "true positive" HIV/AIDS-related deaths as including all HIV-positive people without symptoms of injury or obstetric complications. It is also higher than the 43% estimated among 15–44 year-olds in the Kisesa testing dataset from the present data (data not shown).

In Manicaland the HIV/AIDS-related CSMF range was 58% to 67%, but there may have been upward bias due to the composition of the dataset, particularly considering the low proportion of HIV-negative compared with HIV-positive people whose deaths received a VA interview. The original publication on the Lopman algorithm contains the only other population-level estimate of HIV/AIDS-related mortality in Manicaland, but did not report the proportion of deaths assigned to HIV/AIDS by the algorithm. It did report that for Manicaland, 74–75% of deaths were HIV/AIDS-related in the reference standard[172]; from reported specificity and sensitivity figures the proportion assigned by the algorithm to HIV/AIDS in the original publication was 55–58%, lower than in the bulk of the variant Lopman algorithms for Manicaland.

The HIV/AIDS-related CSMF of HIV-positive people in Kisesa ranged from 49% (Lopman in testing) to 71% (Manicaland-derived Lopman), although the estimates by other methods varied less, from 56% to 65% with strongly overlapping confidence intervals. This is consistent with the 60% estimate for people in Kisesa aged 15+ found by InterVA-4[102]. In the Manicaland data the HIV/AIDS-related CSMF of HIV-positive people ranged from 63% (InterVA-4) to 77% (Lopman in testing dataset). These were higher than the 58% found for HIV-positive people in Manicaland by Byass and colleagues' InterVA-4 study[102]. It is unfortunate that more estimates of cause-specific mortality disaggregated by HIV status are not available for comparison.

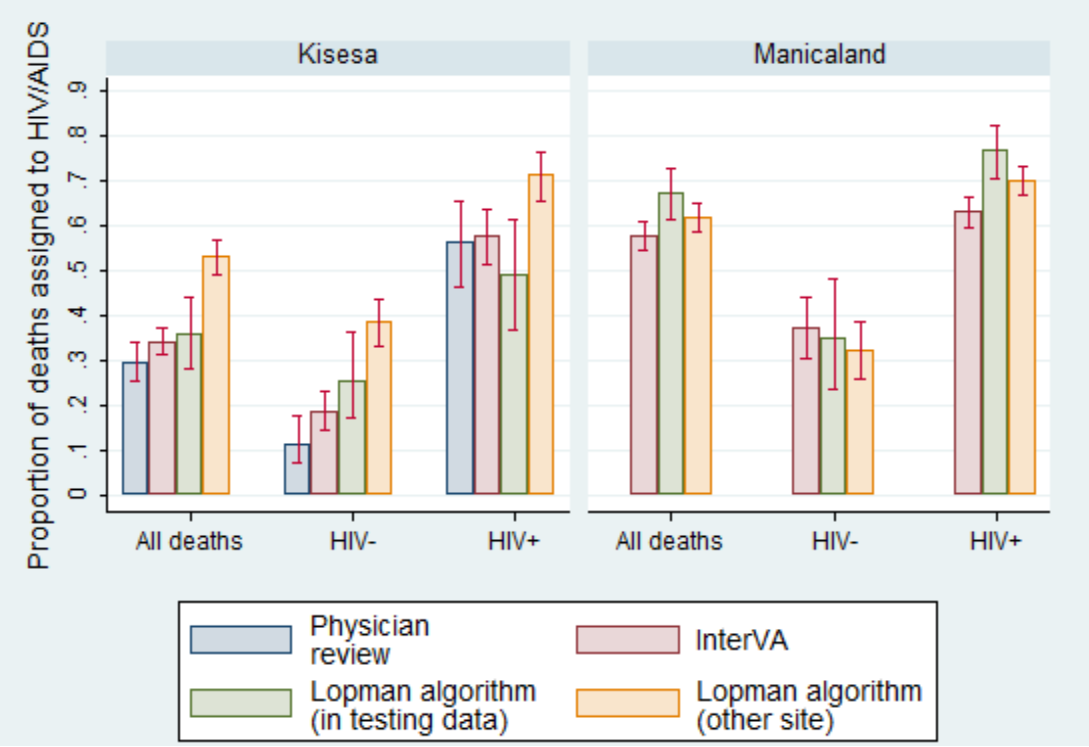

Figure 25: Proportion of deaths assigned to HIV/AIDS by interpretation method, HIV status and study location, with 95% confidence intervals

Method/variant	Estimated CSMF for HIV/AIDS (%)	Specificity (%)
Kisesa		
Physician review (final causes assigned to 462 deaths)	30	88
InterVA-4	34	81
Lopman algorithm, derived in present data, applied in testing dataset	36	74
Lopman algorithm, derived in present data from Manicaland	53	61
Manicaland		
InterVA-4	58	63
Lopman algorithm, derived in present data, applied in testing dataset	67	65
Lopman algorithm, derived in present data from Kisesa	62	68

Table 52: Summary of estimates of the HIV/AIDS-related mortality fraction and specificity assigned by methods investigated

The individual methods performed comparably to the reports for each in the literature: the specificities of all three methods (and the reliability of physician review) were within the ranges that have been reported in other studies[19, 64, 87, 96, 97, 100-102, 116, 120, 172]. It is important to note that direct comparisons – even of identical interpretation methods – are limited by the use of heterogeneous methods in different settings, including different reference standards and different degrees of precision in the causes of death assigned. Few estimates can robustly be used to assess agreement with the thesis results.

Population-based reports of adult deaths in sub-Saharan Africa disaggregated by both cause-of-death and HIV status are very unusual[102]. Such disaggregated estimates exist for selected populations, comparison with which is limited but may be illustrative. In a cohort of mine-workers in South Africa, HIV-negative workers compared with HIV-positive workers had significantly higher proportions of deaths due to unnatural causes (e.g. industrial negligence) or non-communicable diseases; deaths from infections were significantly lower for HIV-negative than HIV-positive workers[58]. Among deaths in pregnant/postpartum women reported to the South African Confidential Enquiries into Maternal Deaths (CEMD), the proportion of deaths due to infections among HIV-negative women was less than one quarter of the proportion among HIV-positive women – even among those who did not have AIDS and were not eligible for ART[33].

Both these populations are probably subject to selection bias due respectively to the "healthy worker effect"[209, 210] and the "healthy pregnant woman effect" [55, 211]. This means the eligible population (mine-workers or pregnant women) is healthier overall than the general population, and is different to the general population in both level and cause-structure of disease. Such selection effects might explain why studies of selected populations report lower proportions of deaths due to non-HIV/AIDS-related infectious diseases among HIV-negative compared to HIV-positive people, whereas in the present data there was no difference in this proportion between HIV-negative and HIV-positive people. This is nonetheless consistent with the suggestion from the lack of elevated mortality from non-HIV/AIDS-related infections among HIV-negative people in the present data that HIV infection is associated with deaths from infectious causes other than or not identified as HIV/AIDS.

2. The importance of definitions

The comparability of estimates from different studies is limited by differences in the (implicit or explicit) definitions of what constitutes an HIV/AIDS-related death. Discussing why estimates of the proportion of pregnancy-related deaths that is due to HIV/AIDS differ so widely, Grollman and Ronsmans state:

> The distinction between deaths "with HIV" and deaths "from HIV" may account for some of the observed difference. Hogan and colleagues [in their model] considered all deaths in HIV-positive pregnant/postpartum women as maternal; population attributable fractions assume that all mortality in HIV-positive pregnant/postpartum women beyond the level observed in HIV-negative pregnant/postpartum women is due to HIV. Neither the Hogan model nor the PAF method consider whether individual deaths are "from HIV" or merely "with HIV". By contrast, in cause-of-death studies physicians or other methods of assigning causes to deaths would only consider a death as HIV-related if the verbal autopsy or clinical history suggests a fatal role of HIV disease or AIDS, and their estimates are therefore of deaths "from HIV". (Grollman and Ronsmans 2014: 91[51])

The latter statement should be broken down somewhat: what it means for verbal autopsy methods to consider a death as HIV/AIDS-related varies according to the interpretation method, particularly whether the method is expert-opinion-based or data-derived. The definitions used by simple expert algorithms such as the 1994 WHO algorithm[38] are clear. For more complex expert-opinion-based methods such as physician review or InterVA, the "definition" is reductive – HIV/AIDS-related deaths are deaths the method assigned to HIV/AIDS given how its expert opinion acted on the data. However, expert-opinion-based methods will usually attempt to identify the deaths in which HIV/AIDS was the underlying cause, akin to identifying deaths "from HIV". Sometimes reviewing physicians may be asked to identify which deaths were of HIV-positive people – deaths "with HIV"[19]. By contrast, whether data-derived methods ascertain deaths "with HIV" or "from HIV" depends on the reference standard used to train them: a reference standard of clinical diagnoses of AIDS such as that used in the PHMRC validation dataset[61] will give different results to a reference standard of HIV status alone, or HIV status plus some VA indicators, as used by Lopman and colleagues[172]. In the present work, the differences between the validity estimates derived using Lopman's reference standard and using the HIV-status-only reference standard (that is, not

recategorising as non-HIV/AIDS-associated those deaths of HIV-positive people with obstetric or injury-related indicators) were small and not significant.

No single definition of HIV/AIDS-related death is correct, but it is important to realise the implications of using given definitions. For example, there is evidence that ART rollout is associated with declines in all-cause mortality as well as AIDS mortality[29], and that HIV-positive people have an elevated risk of mortality from diverse individual causes of death[102]. This suggests that the strictest definition of HIV/AIDS-related mortality – such as the requirement for AIDS used in the PHRMC validation dataset – may miss some deaths that could be prevented by HIV/AIDS-related interventions. This is supported by the finding in Kisesa of significantly higher proportions of deaths being assigned to non-infectious-disease categories among HIV-negative people compared to HIV-positive people, but finding no difference in the proportion assigned to non-HIV/AIDS-related infections.

3. Interpreting measures of validity

The definition of HIV/AIDS-related deaths is not only important for determining what it is that methods diagnose, but is also central to assessing the validity of any method, either data-derived or expert-opinion-based. The degree to which the reference standard comprises death "with HIV" or "from HIV" should therefore affect the interpretation of reported validity.

An HIV-status-only reference standard assesses the validity of interpretation methods for diagnosing deaths of HIV-negative people as not due to HIV/AIDS; such a standard also diagnoses deaths of HIV-positive people as due to HIV/AIDS, which is to validate findings of deaths "with HIV". The Lopman reference standard, which says nothing positive about the symptoms of HIV/AIDS-related deaths, but excludes those people with apparent injuries or obstetric causes of death, is slightly closer to a "from HIV" standard.

4. Validity of the interpretation methods

Among the interpretation methods investigated in this thesis, physician review had the highest specificity (88%), followed by InterVA-4 (81%) and then the Lopman algorithm (74–80%), all in

Kisesa. In Manicaland all methods had lower specificity, with none above 70% and the highest being the version of the Lopman algorithm derived in the Kisesa training dataset (68%). Among the variant Lopman algorithms the range of specificity was 63–73%.

No further validation was possible for the findings of physician review and InterVA-4. As the Lopman algorithm required definition of "true" cases, it was possible to calculate sensitivity and absolute percentage error. Where a single cause is responsible for a very large proportion of deaths, as seems to be the case with HIV/AIDS in Manicaland, any shortfall in sensitivity requires specificity to be substantially lower in order to produce an accurate CSMF. The findings regarding absolute error showed that the "best" validation results in terms of specificity do not imply least error, which is consistent with the findings in both hypothetical datasets[124, 129] and empirical data[125]. This will also be true for the findings of physician review or InterVA, and suggests that attempts to find a "best" interpretation method[129, 130] are unlikely to succeed.

No VA interpretation method has an intrinsic validity. Validity is an attribute not of a method but of a method applied in a given dataset, and datasets vary by symptom composition, symptom–cause associations and cause distributions[123, 124, 129]. Therefore a method can only have a range of performance; the "robust metrics" proposed by Murray and colleagues[129] adjust for the overestimation of performance due to chance, but there is no way to assign a definitive measure of performance to any method.

5. Limitations

I. Inherent in VA

In practice there are common limitations to the quality of VA input data – that is, the accuracy of the reported lists of symptoms – such as recall and forgetting, knowledge of the respondent, memorability of symptoms, local symptom recognition[212] and more. The effect of these factors on the validity of VA is poorly understood[65], and there are ongoing disagreements about the importance of basic questions such as whether symptom duration should form part of the VA data collected[84, 138]. But even were all these problems solved, that is with "perfect" input information to an interpretation method, VA is still vulnerable to the

problems of clinical diagnosis, namely that different causes of death sometimes present similarly[60].

Moreover, HIV/AIDS-related disease itself has a wide variety of presentation. As an example of variation in reported symptoms (whether due to true variation in presence or to another source of variation), among 178 VA records for community-based deaths in Kisesa in 1991–1992, only three reported diarrhea, of whom two were HIV-positive[89]; among 78 deaths in a representative sub-cohort in Masaka in 1990–1996, around two fifths of the 63 HIV-positive deaths were preceded by diarrhea[132]. Differences in presentation led early researchers in Uganda to suggest that the "slim disease" observed there, strongly associated with HIV (then called HTLV–III), was different to the "AIDS" and "AIDS-related complex" seen in nearby countries[36]. In my analyses, there were differences in symptom patterns associated with false-positive assignment of HIV/AIDS by physician review and InterVA-4 in Kisesa and Manicaland. In each case the symptoms associated with false-positive assignment were the same as those associated with assigning HIV/AIDS to deaths of HIV-positive people, and combinations of these symptoms were common among false-positive cases. This underlines the difficulty of achieving high specificity.

II. Methods

The methods of my analyses had several limitations. Several investigations (including the physician-review analyses and many of the HIV-status-specific analyses, particularly for Kisesa) were done on small numbers of deaths, resulting in wide confidence intervals (see e.g. Figure 25). Not all deaths had linked VA data available, and incomplete coverage of VA interviews in Manicaland may have introduced bias to the point estimates of HIV/AIDS-related mortality (though it should not have affected the analyses by HIV status).

Not all those deaths with VA data had the necessary information to use all the interpretation methods. I used all available and eligible records for each analysis, rather than using only those records for which a cause could be assigned by all methods. This was primarily due to the concern over reducing the number of available records even further, but given the small number of records used in each case it somewhat detracts from the within-setting comparability of the results.

Although there is still elevated mortality among HIV-positive people on ART compared with HIV-negative people[29], the applicability of the HIV-status-only reference standard may be limited in the era of ART. Unfortunately, I could not stratify my results by ART status as I did not have individual-level records of ART use.

The reference standard I used to calculate specificity, with deaths in people recently HIV-negative as the denominator, has the advantage for use in resource-limited settings of not requiring clinical diagnoses or confirmatory tests. It is a limitation of my estimate that HIV-positive people who truly died of non-HIV-related causes are not included in the denominator for calculating specificity, but this is a limitation imposed by having data without a "gold standard" assessment of cause-of-death, as will be the case in settings where VA is necessary for monitoring mortality. In such settings, it may be more feasible in the short term to establish widespread knowledge of HIV status than to have widespread valid diagnosis of AIDS-defining conditions, although there are challenges to uptake of voluntary counselling and testing services[213].

III. Data quality

There were also limitations to the quality of the data that were available. The length of recall between death and VA interview, especially in Manicaland, frequently exceeded one year, which is the recommended upper limit in the WHO 2012 VA standards[171], although there is little high quality evidence on effects of timing or many other aspects of the VA procedure[65]. Intervals of a year or more between death and VA interview will probably be fairly common in non-research national sample vital registration systems, which will lack the resources for the frequent rounds of enumeration conducted in typical demographic surveillance systems.

The proportion of VA records assigned unknown cause by the interpretation method has been seen as a sign of poor-quality data[4, 189]. The proportions with unknown cause assigned by InterVA-4 were 8–10%, which is similar to other applications of both InterVA[102, 117] and physician review[78, 97, 187]. The 26% of records with unknown cause by physician review in Kisesa was partly an artefact of the lack of reconciliation of discrepant cause-assignment by individual physicians, although such high proportions of deaths with unknown cause have been seen, both by physician review[184, 214, 215] and InterVA[88, 118].

In Manicaland, 11% (14/124) of people HIV-negative prior to death were reported to have had a diagnosis of HIV/AIDS. In Kisesa the proportion was 3% (8/275). It is impossible to know whether this reflects knowledge of a positive valid HIV test result received outside the research programme subsequent to the last negative result within the programme, or an error in the recording of HIV status within the research programme, or whether it raises doubt about the meaning of VA reports of positive HIV test results. In neither setting, particularly not in Manicaland, was the seroconversion and progression to death of this proportion of recently HIV-negative people plausible. It would be desirable if VA interview instruments could allow assessment of the source and quality of diagnoses when asking about those diagnoses.

The data used in this thesis were collected by many different people over different time periods, using varying methods and with limited prospective supervision. They were therefore probably below the quality that could be expected of rigorous data collection established with clear standard operating procedures for collecting and storing data and training staff. This should be borne in mind when considering the findings of this thesis. Nonetheless, the population-based nature of the data used in my analyses makes them an important source of cause-of-death information. The link with known HIV status enabled a useful and unusual analysis, describing cause-specific mortality distributions by HIV status, and allowed valuable investigations of validity.

6. Using VA data

There are different uses of VA data. For global-level rankings of leading diseases and comparisons of countries, a simple and consistent tool that allows for clear comparison of country-level results globally may be most desirable. Such uses may have little direct impact on resourcing decisions, so the accuracy and precision of their estimates may not be a primary concern.

By contrast, there may be greater concern over accuracy and precision of estimates where these more directly influence healthcare resource allocation. Exercises in prioritising healthcare resources can be opaque – one exploration of criteria for priority-setting in Uganda, for example, refers to “benefit of intervention” and “severity of condition”, but does not say whether these are to be understood at the level of the patient or of the population[216]. While

the importance of "reliable mortality statistics"[217], "mortality data"[4] and so on is universally agreed[2, 3, 5, 10], it is not always obvious how diseases come to be regarded as important or what degree of precision is needed in estimates of cause-specific mortality in order to avoid catastrophic misallocation of resources. Tools such as LiST (the Lives Saved Tool), can be applied at national or subnational level (although to date this tool has been used for child and maternal mortality rather than general adult mortality). LiST uses CSMDs as an input and is therefore affected by this question. Variability in cause-of-death data is acknowledged as a limitation of the method, though Bryce and colleagues recommend that "where feasible, [accurate information on the causes of child deaths] should be collected routinely through verbal autopsies conducted as a part of nationally representative household surveys" (2010: i46[218]).

There are not published sensitivity analyses of the implications for tools such as LiST of variation in the accuracy of CSMDs. Such variation may be very important – as noted, discussing lives not saved due to error in estimating CSMFs, Murray and colleagues suggest that "the negative consequence scales to the absolute error in cause estimation" (2011: 7[129]). On the other hand, Kahn and colleagues present the dual purposes of the Agincourt demographic surveillance system as being "to establish and evaluate innovative subdistrict health centre programmes; and to provide valid population-based data to inform this process" (1999: 434[219]). They go on to say that "the verbal autopsy approach has proved vital in determining the ranking of causes of death in Agincourt, rural South Africa – information vital to district level priority-setting and planning" (1999: 439). A mere ranking of causes of death need not involve precise CSMF estimates and may better recognise the inherent limitations of working with VA data.

It behoves researchers to remember that technical solutions to relatively peripheral aspects of vital registration cannot ultimately do much in the face of political and economic failings with regard to both vital registration and the health system more broadly[10, 220-222]. Nonetheless, if estimating cause-specific mortality as accurately as possible is important – for uses with more tangible implications for health resourcing – estimating cause-specific mortality using more than a single interpretation of VA data may be desirable.

I. Triangulation or data synthesis

The findings of VA analyses ought not to be treated as precise estimates as they are inherently of low accuracy. Comparison of the findings of multiple interpretation methods applied to the same data may help in assessing the likely true answer, and it is even possible that attempts at adjustment using a range of plausible values for sensitivity and specificity would also be helpful[135]. Where the aim is accurate CSMFs, VA interpretation methods should be given as much information as possible, including HIV status of the deceased if known – or last test result and test date. This includes using as much information from the health care experience as is available – some validation studies have looked at the performance of methods both with and without such information[121]. Sometimes the VA reviewer is given access to the known HIV status collected in DSS research studies[30]. Encouraging knowledge of HIV status through voluntary counselling and testing may improve VA-based estimation of HIV/AIDS-related mortality.

However, it may be valuable to include non-VA sources of data in assessing cause-specific mortality, where possible. Compared to other causes of death, HIV/AIDS may be unusually well placed to allow the use of other data sources. These could include estimated HIV/AIDS-related mortality levels based on HIV prevalence and incidence data and data on access to ART (as go into the Spectrum model[215]); records from health facilities; and data on tuberculosis-related mortality, as this is a common comorbidity with HIV/AIDS and is often confused with HIV/AIDS in cause-of-death studies[29, 34, 141].

Such data sources might be combined following a methodology of public health triangulation[4, 223]. An advantage of triangulation is that it allows formal, structured synthesis of data with different strengths and weaknesses, such as the fact that physician review and InterVA may tend to assign a smaller proportion of deaths to HIV/AIDS than data-derived methods trained on HIV status. Routinely recording data on potentially important aspects of the VA process, including on data-collection procedures (questionnaires, continuous or periodic enumeration of deaths) and estimates of time from death to VA interview and of non-response, may help in assessing the quality of VA data and its weighting in any triangulation process.

However, synthesis of data from a large array of sources is a substantial use of person-time, and does not necessarily result in a clear picture. Commenting on a previous extensive review

of cause-specific mortality recorded in epidemiological studies in sub-Saharan Africa prior to 1996 (Adetunji et al 1996, cited in Rao et al 2006[189]), Rao and colleagues stated that:

> Compiling information from various sources, despite the enormous effort involved, still results in substantial uncertainty about the cause structure of mortality owing to biases in the way the data were collected and the high proportions of unspecified causes in the reports. (2006, p45)

A study comparing InterVA-4 with the Spectrum mortality model to estimate HIV/AIDS-related deaths in slums in Nairobi showed that insufficient input data means tools cannot always be used as intended, which undermines the clarity available in interpreting and comparing their output[215]. A triangulation study investigating the effects of national anti-HIV/AIDS programmes in Swaziland found the approach useful, despite limitations to data quality and difficulty in bringing together all relevant partners[224]. A further consideration in using multiple sources of data to investigate mortality in SVR systems is that adjustments to, for example, the HIV/AIDS-related fraction in a VA-based cause-specific mortality distribution would necessitate changes to the fractions due to other causes, and it is not clear how those changes would be made.

An alternative approach to achieve greater accuracy and precision in using VA data would be to aim to achieve less detail in cause-specific mortality estimates. Instead of assigning deaths to one of fifty or more groups, Joshi and colleagues used 16 chapter headings from ICD-10, with results for injuries, cardiovascular causes and infectious diseases further subdivided, giving 25 classifications in total[173]. Being able to monitor the relative importance of infectious and non-communicable diseases, injuries and maternal conditions, possibly even at a level of resolution similar to the broad cause categories in the present thesis, might provide useful overall guidance for public health resource allocation; smaller scale studies using more resource-intensive methods could supplement these to achieve more detailed understandings of prevention and treatment needs.

II. Sample vital registration with verbal autopsy

Sample vital registration systems have routinely collected VA data but by definition do not have valid causes of death known and available to use as reference-standard data. This means that general validation is not possible, and also means the data are not available on which to train data-derived VA interpretation methods. The findings applying the original Lopman

algorithm to the present datasets suggest that caution must be exercised before applying data-derived algorithms derived in one setting to another setting. The potential to use data-derived methods to assign causes to deaths in SVR systems is therefore severely limited.

Data-derived methods have limited potential for other reasons:

- Data-derived methods only assess the relationship between symptoms and causes in the setting from which they are gathered. They suffer from an inherent bias in being trained in the same dataset in which they are tested, meaning they return answers that are internally valid in the dataset in which they are trained, but may not have external validity[86, 138], particularly for causes for which the training dataset has relatively few cases[120]. This does not mean they cannot be used, but does mean that this bias must be acknowledged when interpreting their output;

- Even where reference-standard data are available, data-derived methods are particularly vulnerable to bias or errors in those data: while biased or inaccurate recording of symptoms also affects the performance of expert-opinion-based methods, the effect is more critical with data-derived methods because they incorporate the bias and error into their reference standard as well as in their estimation.

The lack of valid COD data limits both HIV/AIDS-specific and all-cause validation.

Accepting that validation is not possible in many settings where SVR is useful, some authors have suggested that plausibility, consistency or other measures may be useful for assessing data accuracy[64, 65, 137]. Some studies that do not seek to validate their findings nonetheless use comparison with physician review in order to assess consistency between the new method and physician review, as physician review is the default method of interpreting VA data[32, 88, 101, 115, 118, 121]. Several studies have reported reliability or agreement rather than validity – the metrics reported have been percentage agreement[32, 118, 190], and kappa[32, 101, 120, 136]. One study assessed cause-specific agreement as the lowest of the CSMFs assigned by respective interpretation methods[88]. Desai and colleagues did not present their analysis as a validation study, as it lacked the true cause-of-death needed for formal validation[111]. Nonetheless they calculated positive predictive value, partial chance-corrected concordance and CSMF accuracy, which are validation metrics. All these metrics have the same vulnerability to the true CSMD as those discussed above for specificity[129].

Despite having performed relatively well in terms of specificity in the present analysis, in practice physician review remains costly and slow compared to automated methods. Although there has been some equivocation more recently regarding abandoning the method altogether, it seems unlikely that large-scale roll-out of VA-based routine cause-of-death surveillance in sub-Saharan Africa will use physician review if any other method is available. The alignment of the recently revised WHO VA instrument with InterVA-4[84] is a vote of confidence in that method or methods that improve upon it[225]. Validation of routine cause-specific mortality estimates will not be possible, but continuing to assess the performance of interpretation methods in datasets where validation is possible[62] is desirable: it does not inspire confidence when, among the generally poor Lopman algorithm variants, the "best performing" – after false positives and false negatives have cancelled out – is that which diagnoses almost two fifths of HIV-negative people with HIV/AIDS. As an aside, a further reason it may be desirable to have some confidence in the accuracy of individual diagnoses by VA is that one future use of VA could be to tell people – with appropriate caveats about uncertainty – what their loved ones died of.

III. Antiretroviral therapy and HIV/AIDS as a cause of death

Probably the biggest contemporary challenge in VA analysis for estimating HIV/AIDS-related mortality is the change in HIV/AIDS-related morbidity and mortality caused by antiretroviral therapy. ART has already reduced HIV/AIDS-related mortality massively in settings where it is available[29, 50, 70, 142, 226]. Population-based data from Masaka, Uganda, with estimated ART coverage of 70% of people needing treatment, suggest that even with such high ART coverage, death rates among HIV-positive people remain over ten times greater than among HIV-negative people[29]. Reductions in HIV/AIDS-related mortality may occur without any change in the rate of non-HIV/AIDS-related infections[30, 227].

There are no studies of the symptoms reported in VA for deceased HIV-positive people on ART and not on ART, nor proposals for how VA interpretation methods ought to ascertain HIV/AIDS-related deaths among people on ART. Although absolute levels of mortality will fall dramatically, data from wealthy settings suggest some diseases may increase among people using ART (including liver disease[228] and pre-eclampsia[229]). Many people who do receive ART nonetheless go on to develop AIDS and die of classic HIV/AIDS-related diseases that occur in

ART-naïve people[230], although the AIDS-defining conditions that HIV-positive people suffer have been seen to differ between pre-ART periods and periods of ART availability[230, 231]. Experience in low-income countries may differ from that in middle- and high-income countries, not least due to differences in background infectious-disease epidemiology[139]. VA methods need to respond to this change. For example, some researchers are restructuring VA reporting to allow primary and contributing causes[30]. Such subtleties will be a particular challenge for automated methods of VA interpretation.

There is a wider conceptual question of whether and how HIV ought to be understood as a cause of death. For the purposes of public health officials, it is useful to consider HIV to be a cause of death distinct from the immediate causes that are ultimately fatal: HIV causes morbidity and mortality that would not otherwise occur; and there are HIV-specific prevention and treatment interventions that can reduce that morbidity and mortality.

There may be a reluctance to define HIV as a cause of death derived from VA, which would mirror the removal of unspecified HIV disease from the draft 11th revision of the International Statistical Classification of Diseases (ICD-11)[§§]. One alternative might be to assign a cause of death from a list that does not include HIV/AIDS, and give a separate indication of whether the deceased person was believed to have been HIV-positive. As noted, such an approach has been followed in physician coding in Manicaland. Some (non-VA) investigations of pregnancy-related mortality have also used the approach of reporting deaths by cause and HIV status[33, 53].

If validation remains desirable, the widespread use of ART presents a conceptual challenge. Lopman and colleagues worked from the starting point that "we knew a priori that most adult deaths were HIV-associated" (2006: 1275[172]). This is reasonable in an ART-naïve population with high HIV prevalence, but becomes problematic as an increasing proportion of HIV-positive adults die of causes other than classic AIDS-defining conditions[139]. Arguably, the simple reference standard of HIV status is inadequate when a large proportion of mortality among HIV-positive adults is no longer attributable to HIV/AIDS – authors studying mortality in an ART cohort in Cameroon wrote that "to attribute [all deaths in the cohort] to HIV overestimates the AIDS-related mortality rate" (Sieleunou et al 2009: 41[232]).

[§§] ICD-11 is in Beta Draft at apps.who.int/classifications/icd11. It is frequently updated and subject to change. Last accessed 1st April 2014. For comparison, ICD-10 is at apps.who.int/classifications/icd10/.

7. Conclusion

The findings of the different VA interpretation methods within each setting were fairly similar, but not similar enough to provide even a crude point estimate. Moreover, validity measures cannot be used to assess which estimate might be most correct. It is possible that the true proportion of deaths that is HIV/AIDS-related is higher than the highest estimate or lower than the lowest but, in the absence of a correct answer, having several estimates allows one to consider their different approaches and infer more than would be possible with a single estimate. For example, I suggest that since the Lopman algorithm was trained to detect deaths of HIV-positive people (albeit those without injuries or obstetric complications) and both physician review and InterVA-4 work by considering HIV/AIDS as a potential cause of death alongside all other causes, it is reasonable to treat physician review and InterVA-4 as giving lower estimates and the Lopman algorithm as giving an upper estimate.

The uses of VA data vary[4, 65], but one requirement of SVR is for estimates of cause-specific mortality that can be used for health planning. It is attractive to seek a single CSMD for a population from simple application of a VA interpretation method, but the accuracy of such an approach cannot be known. Users may have more confidence in estimates derived from interpretation of more than a single use of VA; on the basis of the present findings I would not recommend any single method for ascertaining HIV/AIDS-related mortality from VA data. Research into VA for estimating HIV/AIDS-related mortality in populations with substantial use of ART should be a priority.

References

1. Lozano, R., et al., *Global and regional mortality from 235 causes of death for 20 age groups in 1990 and 2010: a systematic analysis for the Global Burden of Disease Study 2010.* Lancet, 2012. **380**(9859): p. 2095-128.
2. Mathers, C.D., et al., *Counting the dead and what they died from: an assessment of the global status of cause of death data.* Bull World Health Organ, 2005. **83**(3): p. 171-7.
3. Ye, Y., et al., *Health and demographic surveillance systems: a step towards full civil registration and vital statistics system in sub-Sahara Africa?* BMC Public Health, 2012. **12**: p. 741.
4. Joubert, J., et al., *Characteristics, availability and uses of vital registration and other mortality data sources in post-democracy South Africa.* Glob Health Action, 2012. **5**: p. 1-19.
5. Mudenda, V., et al., *Tuberculosis and tuberculosis/HIV/AIDS-associated mortality in Africa: the urgent need to expand and invest in routine and research autopsies.* J Infect Dis, 2012. **205 Suppl 2**: p. S340-6.
6. Mathers, C.D., T. Boerma, and D. Ma Fat, *Global and regional causes of death.* Br Med Bull, 2009. **92**: p. 7-32.
7. Chandramohan, D., et al., *Verbal autopsies for adult deaths: issues in their development and validation.* Int J Epidemiol, 1994. **23**(2): p. 213-22.
8. Jha, P., *Reliable direct measurement of causes of death in low- and middle-income countries.* BMC Med, 2014. **12**: p. 19.
9. Rao, C., A.D. Lopez, and Y. Hemed, *Causes of Death*, in *Disease and Mortality in Sub-Saharan Africa*, D.T. Jamison, et al., Editors. 2006, World Bank
The International Bank for Reconstruction and Development/The World Bank.: Washington (DC).
10. Setel, P.W., et al., *Sample registration of vital events with verbal autopsy: a renewed commitment to measuring and monitoring vital statistics.* Bull World Health Organ, 2005. **83**(8): p. 611-7.
11. Gajalakshmi, V. and R. Peto, *Verbal autopsy of 80,000 adult deaths in Tamilnadu, South India.* BMC Public Health, 2004. **4**: p. 47.
12. Setel, P.W., et al., *Core verbal autopsy procedures with comparative validation results from two countries.* PLoS Med, 2006. **3**(8): p. e268.
13. Mudenda, S.S., et al., *Feasibility of using a World Health Organization-standard methodology for Sample Vital Registration with Verbal Autopsy (SAVVY) to report leading causes of death in Zambia: results of a pilot in four provinces, 2010.* Popul Health Metr, 2011. **9**: p. 40.
14. WHO, *WHO case definitions of HIV for surveillance and revised clinical staging and immunological classification of HIV-related disease in adults and children*. 2007, Geneva: World Health Organization.
15. UNAIDS, *Global report: UNAIDS report on the global AIDS epidemic 2013*. 2013.
16. Sewankambo, N.K., et al., *Demographic impact of HIV infection in rural Rakai district, Uganda: results of a population-based cohort study.* Aids, 1994. **8**(12): p. 1707-13.
17. Sewankambo, N.K., et al., *Mortality associated with HIV infection in rural Rakai District, Uganda.* Aids, 2000. **14**(15): p. 2391-400.
18. Boerma, J., et al., *Levels and causes of adult mortality in rural Tanzania with special reference to HIV/AIDS.* Health Transition Review 1997. **7**(supp.2): p. 63-74.
19. Mayanja, B.N., et al., *Using verbal autopsy to assess the prevalence of HIV infection among deaths in the ART period in rural Uganda: a prospective cohort study, 2006-2008.* Popul Health Metr, 2011. **9**: p. 36.
20. Groenewald, P., et al., *Identifying deaths from AIDS in South Africa.* Aids, 2005. **19**(2): p. 193-201.

21. Sartorius, B., et al., *Dying in their prime: determinants and space-time risk of adult mortality in rural South Africa.* Geospat Health, 2013. **7**(2): p. 237-49.
22. Lopman, B.A., et al., *Assessing adult mortality in HIV-1-afflicted Zimbabwe (1998 - 2003).* Bull World Health Organ, 2006. **84**(3): p. 189-97.
23. Nunn, A.J., et al., *Mortality associated with HIV-1 infection over five years in a rural Ugandan population: cohort study.* Bmj, 1997. **315**(7111): p. 767-71.
24. Urassa, M., et al., *The impact of HIV/AIDS on mortality and household mobility in rural Tanzania.* Aids, 2001. **15**(15): p. 2017-23.
25. Gregson, S., et al., *Critique of early models of the demographic impact of HIV/AIDS in sub-Saharan Africa based on contemporary empirical data from Zimbabwe.* Proc Natl Acad Sci U S A, 2007. **104**(37): p. 14586-91.
26. Porter, K. and B. Zaba, *The empirical evidence for the impact of HIV on adult mortality in the developing world: data from serological studies.* Aids, 2004. **18 Suppl 2**: p. S9-s17.
27. Crampin, A.C., et al., *Long-term follow-up of HIV-positive and HIV-negative individuals in rural Malawi.* Aids, 2002. **16**(11): p. 1545-50.
28. Ford, N., A. Calmy, and E.J. Mills, *The first decade of antiretroviral therapy in Africa.* Global Health, 2011. **7**: p. 33.
29. Floyd, S., et al., *The effect of antiretroviral therapy provision on all-cause, AIDS and non-AIDS mortality at the population level--a comparative analysis of data from four settings in Southern and East Africa.* Trop Med Int Health, 2012. **17**(8): p. e84-93.
30. Chihana, M., et al., *Adult mortality and probable cause of death in rural northern Malawi in the era of HIV treatment.* Trop Med Int Health, 2012. **17**(8): p. e74-83.
31. Herbst, A.J., et al., *Adult mortality and antiretroviral treatment roll-out in rural KwaZulu-Natal, South Africa.* Bull World Health Organ, 2009. **87**(10): p. 754-62.
32. Herbst, A.J., T. Mafojane, and M.L. Newell, *Verbal autopsy-based cause-specific mortality trends in rural KwaZulu-Natal, South Africa, 2000-2009.* Popul Health Metr, 2011. **9**: p. 47.
33. NCCEMD. *Saving Mothers 2008-2010: Fifth report on the confidential enquiries into maternal deaths in South Africa*. 2012 [cited; Available from: http://www.doh.gov.za/docs/reports/2012/Report_on_Confidential_Enquiries_into_Maternal_Deaths_in_South_Africa.pdf.
34. Hosegood, V., A.M. Vanneste, and I.M. Timaeus, *Levels and causes of adult mortality in rural South Africa: the impact of AIDS.* AIDS, 2004. **18**(4): p. 663-71.
35. Groenewald, P., et al., *Identifying deaths from AIDS in South Africa: an update.* Aids, 2005. **19**(7): p. 744-5.
36. Serwadda, D., et al., *Slim disease: a new disease in Uganda and its association with HTLV-III infection.* Lancet, 1985. **2**(8460): p. 849-52.
37. WHO. *Bangui definition of AIDS*. 1985 [cited; Available from: http://www.who.int/hiv/strategic/en/bangui1985report.pdf.
38. WHO, *WHO case definitions for AIDS surveillance in adults and adolescents.* Wkly Epidemiol Rec, 1994. **69**(37): p. 273-5.
39. Quinn, T.C., et al., *AIDS in Africa: an epidemiologic paradigm. 1986.* Bull World Health Organ, 2001. **79**(12): p. 1159-67.
40. WHO, *Interim WHO clinical staging of HIV/AIDS and HIV/AIDS case definitions for surveillance: African region*. 2005, Geneva: World Health Organization.
41. WHO, *International Statistical Classification of Diseases and Related Health Problems 10th Revision Volume 2 Instruction manual*. 2011, Geneva: World Health Organization.
42. Birnbaum, J.K., C.J. Murray, and R. Lozano, *Exposing misclassified HIV/AIDS deaths in South Africa.* Bull World Health Organ, 2011. **89**(4): p. 278-85.
43. Diaz, T., et al., *Surveillance methods to monitor the impact of HIV therapy programmes in resource-constrained countries.* Aids, 2005. **19 Suppl 2**: p. S31-7.

44. Campaign, T.A. *Understanding the Confidential Enquiry into Maternal Deaths: A TAC briefing*. 2006 [cited; Available from: http://www.tac.org.za/documents/AnalysisOfMaternalDeaths.pdf.
45. Le Coeur, S., et al., *HIV and the magnitude of pregnancy-related mortality in Pointe Noire, Congo.* Aids, 2005. **19**(1): p. 69-75.
46. Zaba, B., et al., *Effect of HIV infection on pregnancy-related mortality in sub-Saharan Africa: secondary analyses of pooled community-based data from the network for Analysing Longitudinal Population-based HIV/AIDS data on Africa (ALPHA).* Lancet, 2013. **381**(9879): p. 1763-71.
47. NCCEMD. *Saving Mothers 2005-2007: Fourth report on the confidential enquiries into maternal deaths in South Africa*. 2008 [cited; Available from: http://www0.sun.ac.za/ruralhealth/ukwandahome/rudasaresources2009/DOH/savingmothers%2005-07%5B1%5D.pdf.
48. Calvert, C. and C. Ronsmans, *HIV and risk of direct obstetric complications: a systematic review and meta-analysis.* PLoS ONE, 2013. **8**(10): p. e74848.
49. Garenne, M., et al., *Protective Effect of Pregnancy in Rural South Africa: Questioning the Concept of "Indirect Cause" of Maternal Death.* PLoS ONE, 2013. **8**: p. 5.
50. Garenne, M., et al., *Maternal mortality in rural South Africa: the impact of case definition on levels and trends.* Int J Womens Health, 2013. **5**: p. 457-63.
51. Grollman, C. and C. Ronsmans, *Systematic review of the proportion of pregnancy-related deaths attributed to HIV in population-based studies in sub-Saharan Africa.* Tropical Medicine & International Health, 2014. **19**(1): p. 83-97.
52. Kanyighe, C., et al., *Determinants of post-partum maternal mortality at Queen Elizabeth Central Hospital, Blantyre, Malawi: a case-control study 2001-2002.* Afr J Reprod Health, 2008. **12**(3): p. 35-48.
53. Menendez, C., et al., *An autopsy study of maternal mortality in Mozambique: the contribution of infectious diseases.* PLoS Med, 2008. **5**(2): p. e44.
54. Gabrysch, S., et al., *Tracking progress towards safe motherhood: meeting the benchmark yet missing the goal? An appeal for better use of health-system output indicators with evidence from Zambia and Sri Lanka.* Trop Med Int Health, 2011. **16**(5): p. 627-39.
55. Ronsmans, C., et al., *Evidence for a 'healthy pregnant woman effect' in Niakhar, Senegal?* Int J Epidemiol, 2001. **30**(3): p. 467-73; discussion 474-5.
56. Institute for Health Metrics and Evaluation, *The Global Burden of Disease: Generating Evidence, Guiding Policy*. 2013, Seattle, WA: IHME.
57. Morris, S.S., R.E. Black, and L. Tomaskovic, *Predicting the distribution of under-five deaths by cause in countries without adequate vital registration systems.* Int J Epidemiol, 2003. **32**(6): p. 1041-51.
58. Murray, J., et al., *Cause of death and presence of respiratory disease at autopsy in an HIV-1 seroconversion cohort of southern African gold miners.* AIDS, 2007. **21 Suppl 6**: p. S97-S104.
59. Cox, J.A., et al., *Autopsy causes of death in HIV-positive individuals in sub-Saharan Africa and correlation with clinical diagnoses.* AIDS Rev, 2010. **12**(4): p. 183-94.
60. Fligner, C.L., J. Murray, and D.J. Roberts, *Synergism of verbal autopsy and diagnostic pathology autopsy for improved accuracy of mortality data.* Popul Health Metr, 2011. **9**: p. 25.
61. Murray, C.J., et al., *Population Health Metrics Research Consortium gold standard verbal autopsy validation study: design, implementation, and development of analysis datasets.* Popul Health Metr, 2011. **9**: p. 27.
62. Laserson, K., *[Presentation] Cause of death determination: report on pilot development. Comparison of post mortem (PM) to verbal autopsy (VA) to improve HDSS mortality data*, in *11th INDEPTH scientific conference*. 2011: Maputo, Mozambique.

63. Cox, J.A., et al., *An autopsy study describing causes of death and comparing clinico-pathological findings among hospitalized patients in Kampala, Uganda.* PLoS One, 2012. **7**(3): p. e33685.
64. Fantahun, M., et al., *Assessing a new approach to verbal autopsy interpretation in a rural Ethiopian community: the InterVA model.* Bull World Health Organ, 2006. **84**(3): p. 204-10.
65. Fottrell, E. and P. Byass, *Verbal autopsy: methods in transition.* Epidemiol Rev, 2010. **32**(1): p. 38-55.
66. Soleman, N., D. Chandramohan, and K. Shibuya, *Verbal autopsy: current practices and challenges.* Bull World Health Organ, 2006. **84**(3): p. 239-45.
67. Campbell, O. and C. Ronsmans, *Verbal autopsies for maternal deaths: report of a WHO workshop, London, 10-13 January 1994*. 1995, Geneva: World Health Organization.
68. Zimicki, S., *Approaches to assessment of the cause structure of mortality: a case study from Bangladesh*, in *Measurement and Analysis of Mortality*, J. Vallin, S. D'Souza, and A. Palloni, Editors. 1990, OUP: New York.
69. King, G. and Y. Lu, *Verbal autopsy methods with multiple causes of death.* Statistical Science, 2008. **23**(1): p. 78-91.
70. Reniers, G., et al., *Steep declines in population-level AIDS mortality following the introduction of antiretroviral therapy in Addis Ababa, Ethiopia.* Aids, 2009. **23**(4): p. 511-8.
71. Floyd, S., et al., *The long-term social and economic impact of HIV on the spouses of infected individuals in northern Malawi.* Trop Med Int Health, 2008. **13**(4): p. 520-31.
72. Rankin, W.W., et al., *The stigma of being HIV-positive in Africa.* PLoS Med, 2005. **2**(8): p. e247.
73. Burger, E.H., et al., *Validation study of cause of death statistics in Cape Town, South Africa, found poor agreement.* J Clin Epidemiol, 2012. **65**(3): p. 309-16.
74. Chandramohan, D., et al., *Ethical issues in the application of verbal autopsies in mortality surveillance systems.* Trop Med Int Health, 2005. **10**(11): p. 1087-9.
75. Doctor, H.V. and A.A. Weinreb, *Estimation of AIDS adult mortality by verbal autopsy in rural Malawi.* Aids, 2003. **17**(17): p. 2509-13.
76. Yang, G., et al., *Mortality registration and surveillance in China: History, current situation and challenges.* Popul Health Metr, 2005. **3**(1): p. 3.
77. Doctor, H.V., *Variations in under-five mortality estimates in Nigeria: explanations and implications for program monitoring and evaluation.* Matern Child Health J, 2013. **17**(8): p. 1355-8.
78. Araya, T., et al., *Burial surveillance detected significant reduction in HIV--related deaths in Addis Ababa, Ethiopia.* Trop Med Int Health, 2011. **16**(12): p. 1483-9.
79. Mathers, C.D. and D. Loncar, *Projections of global mortality and burden of disease from 2002 to 2030.* PLoS Med, 2006. **3**(11): p. e442.
80. Byass, P., et al., *Strengthening standardised interpretation of verbal autopsy data: the new InterVA-4 tool.* Glob Health Action, 2012. **5**: p. 1-8.
81. InterVA. *InterVA-4 User Guide, version 4.RC1 2012-08-14*. 2012 [cited; Available from: www.interva.net
82. Hurt, L., et al., *Effect of vitamin A supplementation on cause-specific mortality in women of reproductive age in Ghana: a secondary analysis from the ObaapaVitA trial.* Bull World Health Organ, 2013. **91**(1): p. 19-27.
83. WHO, *WHO technical consultation verbal autopsy: Final report: Review of the literature and currently used verbal autopsy tools*. 2005, World Health Organization: Geneva.
84. Leitao, J., et al., *Revising the WHO verbal autopsy instrument to facilitate routine cause-of-death monitoring.* Glob Health Action, 2013. **6**: p. 21518.
85. WHO, *Verbal autopsy standards: Ascertaining and attributing cause of death*. 2007, Geneva: WHO.

86. Byass, P., *Usefulness of the Population Health Metrics Research Consortium gold standard verbal autopsy data for general verbal autopsy methods.* BMC Medicine, 2014. **12**: p. 23.
87. Lopman, B., et al., *Verbal autopsy can consistently measure AIDS mortality: a validation study in Tanzania and Zimbabwe.* J Epidemiol Community Health, 2010. **64**(4): p. 330-4.
88. Ramroth, H., et al., *Cause of death distribution with InterVA and physician coding in a rural area of Burkina Faso.* Trop Med Int Health, 2012. **17**(7): p. 904-13.
89. Todd, J., et al., *HIV-associated adult mortality in a rural Tanzanian population.* Aids, 1997. **11**(6): p. 801-7.
90. Murray, C.J., et al., *Simplified Symptom Pattern Method for verbal autopsy analysis: multisite validation study using clinical diagnostic gold standards.* Popul Health Metr, 2011. **9**: p. 30.
91. Flaxman, A.D., et al., *Direct estimation of cause-specific mortality fractions from verbal autopsies: multisite validation study using clinical diagnostic gold standards.* Popul Health Metr, 2011. **9**: p. 35.
92. Flaxman, A.D., et al., *Random forests for verbal autopsy analysis: multisite validation study using clinical diagnostic gold standards.* Popul Health Metr, 2011. **9**: p. 29.
93. Lozano, R., et al., *Performance of InterVA for assigning causes of death to verbal autopsies: multisite validation study using clinical diagnostic gold standards.* Popul Health Metr, 2011. **9**: p. 50.
94. Lozano, R., et al., *Performance of physician-certified verbal autopsies: multisite validation study using clinical diagnostic gold standards.* Popul Health Metr, 2011. **9**: p. 32.
95. James, S.L., A.D. Flaxman, and C.J. Murray, *Performance of the Tariff Method: validation of a simple additive algorithm for analysis of verbal autopsies.* Popul Health Metr, 2011. **9**: p. 31.
96. Kamali, A., et al., *Verbal autopsy as a tool for diagnosing HIV-related adult deaths in rural Uganda.* Int J Epidemiol, 1996. **25**(3): p. 679-84.
97. Chandramohan, D., et al., *Verbal autopsies for adult deaths: their development and validation in a multicentre study.* Trop Med Int Health, 1998. **3**(6): p. 436-46.
98. Quigley, M.A., D. Chandramohan, and L.C. Rodrigues, *Diagnostic accuracy of physician review, expert algorithms and data-derived algorithms in adult verbal autopsies.* Int J Epidemiol, 1999. **28**(6): p. 1081-7.
99. Lulu, K. and Y. Berhane, *The use of simplified verbal autopsy in identifying causes of adult death in a predominantly rural population in Ethiopia.* BMC Public Health, 2005. **5**: p. 58.
100. Setel, P.W., et al., *Validity of verbal autopsy procedures for determining cause of death in Tanzania.* Trop Med Int Health, 2006. **11**(5): p. 681-96.
101. Bauni, E., et al., *Validating physician-certified verbal autopsy and probabilistic modeling (InterVA) approaches to verbal autopsy interpretation using hospital causes of adult deaths.* Popul Health Metr, 2011. **9**: p. 49.
102. Byass, P., et al., *InterVA-4 as a public health tool for measuring HIV/AIDS mortality: a validation study from five African countries.* Glob Health Action, 2013. **6**: p. 22448.
103. Quigley, M.A., et al., *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): p. 33-9.
104. Munjanja, S., et al. *Maternal and perinatal mortality study*. 2007 [cited; Available from: http://www.unicef.org/zimbabwe/ZMPMS_report.pdf.
105. Negin, J., et al., *High rates of AIDS-related mortality among older adults in rural Kenya.* J Acquir Immune Defic Syndr, 2010. **55**(2): p. 239-44.
106. Eddy, D.M., *Variations in physician practice: the role of uncertainty.* Health Aff (Millwood), 1984. **3**(2): p. 74-89.

107. Todd, J.E., et al., *The limitations of verbal autopsy in a malaria-endemic region.* Ann Trop Paediatr, 1994. **14**(1): p. 31-6.
108. Mills, S., et al., *Maternal mortality decline in the Kassena-Nankana district of northern Ghana.* Maternal and Child Health Journal, 2008. **12**(5): p. 577-85.
109. Fottrell, E., *Advances in verbal autopsy: pragmatic optimism or optimistic theory?* Popul Health Metr, 2011. **9**: p. 24.
110. Leitao, J., et al., *Comparison of physician-certified verbal autopsy with computer-coded verbal autopsy for cause of death assignment in hospitalized patients in low- and middle-income countries: systematic review.* BMC Med, 2014. **12**(1): p. 22.
111. Desai, N., et al., *Performance of four computer-coded verbal autopsy methods for cause of death assignment compared with physician coding on 24,000 deaths in low- and middle-income countries.* BMC Med, 2014. **12**(1): p. 20.
112. Byass, P., *Who needs cause-of-death data?* PLoS Med, 2007. **4**(11): p. e333.
113. Kalter, H.D., et al., *Validation of postmortem interviews to ascertain selected causes of death in children.* Int J Epidemiol, 1990. **19**(2): p. 380-6.
114. Byass, P., D.L. Huong, and H.V. Minh, *A probabilistic approach to interpreting verbal autopsies: methodology and preliminary validation in Vietnam.* Scand J Public Health Suppl, 2003. **62**: p. 32-7.
115. Byass, P., et al., *Refining a probabilistic model for interpreting verbal autopsy data.* Scand J Public Health, 2006. **34**(1): p. 26-31.
116. Tensou, B., et al., *Evaluating the InterVA model for determining AIDS mortality from verbal autopsies in the adult population of Addis Ababa.* Trop Med Int Health, 2010. **15**(5): p. 547-53.
117. Fottrell, E., et al., *Mortality measurement in transition: proof of principle for standardised multi-country comparisons.* Trop Med Int Health, 2010. **15**(10): p. 1256-65.
118. Oti, S.O. and C. Kyobutungi, *Verbal autopsy interpretation: a comparative analysis of the InterVA model versus physician review in determining causes of death in the Nairobi DSS.* Popul Health Metr, 2010. **8**: p. 21.
119. Tadesse, S. and T. Tadesse, *Evaluating the performance of interpreting Verbal Autopsy 3.2 model for establishing pulmonary tuberculosis as a cause of death in Ethiopia: a population-based cross-sectional study.* BMC Public Health, 2012. **12**: p. 1039.
120. Boulle, A., D. Chandramohan, and P. Weller, *A case study of using artificial neural networks for classifying cause of death from verbal autopsy.* Int J Epidemiol, 2001. **30**(3): p. 515-20.
121. Murray, C.J., et al., *Using verbal autopsy to measure causes of death: the comparative performance of existing methods.* BMC Med, 2014. **12**(1): p. 5.
122. Aleksandrowicz, L., et al., *Performance criteria for verbal autopsy-based systems to estimate national causes of death: development and application to the Indian Million Death Study.* BMC Med, 2014. **12**(1): p. 21.
123. Chandramohan, D., P. Setel, and M. Quigley, *Effect of misclassification of causes of death in verbal autopsy: can it be adjusted?* Int J Epidemiol, 2001. **30**(3): p. 509-14.
124. Anker, M., *The effect of misclassification error on reported cause-specific mortality fractions from verbal autopsy.* Int J Epidemiol, 1997. **26**(5): p. 1090-6.
125. Quigley, M.A., J.R. Armstrong Schellenberg, and R.W. Snow, *Algorithms for verbal autopsies: a validation study in Kenyan children.* Bull World Health Organ, 1996. **74**(2): p. 147-54.
126. Araya, T., et al., *Accuracy of Physicians in Diagnosing HIV and AIDS-Related Death in the Adult Population of Addis Ababa, Ethiopia.* World Journal of AIDS, 2012. **2**: p. 89-96.
127. Maude, G.H. and D.A. Ross, *The effect of different sensitivity, specificity and cause-specific mortality fractions on the estimation of differences in cause-specific mortality*

rates in children from studies using verbal autopsies. Int J Epidemiol, 1997. **26**(5): p. 1097-106.
128. Kalter, H., *The validation of interviews for estimating morbidity.* Health Policy Plan, 1992. **7**(1): p. 30-9.
129. Murray, C.J., et al., *Robust metrics for assessing the performance of different verbal autopsy cause assignment methods in validation studies.* Popul Health Metr, 2011. **9**: p. 28.
130. Chandramohan, D., *Validation and validity of verbal autopsy procedures.* Popul Health Metr, 2011. **9**: p. 22.
131. Reeves, B.C. and M. Quigley, *A review of data-derived methods for assigning causes of death from verbal autopsy data.* Int J Epidemiol, 1997. **26**(5): p. 1080-9.
132. Okongo, M., et al., *Causes of death in a rural, population-based human immunodeficiency virus type 1 (HIV-1) natural history cohort in Uganda.* Int J Epidemiol, 1998. **27**(4): p. 698-702.
133. Masiira, B., et al., *Mortality and its predictors among antiretroviral therapy naive HIV-infected individuals with CD4 cell count >/=350 cells/mm(3) compared to the general population: data from a population-based prospective HIV cohort in Uganda.* Glob Health Action, 2014. **7**: p. 21843.
134. Chandramohan, D., et al., *The validity of verbal autopsies for assessing the causes of institutional maternal death.* Stud Fam Plann, 1998. **29**(4): p. 414-22.
135. Korenromp, E.L., et al., *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): p. 349-58.
136. Tadesse, S., *Validating the InterVA model to estimate the burden of mortality from verbal autopsy data: a population-based cross-sectional study.* PLoS One, 2013. **8**(9): p. e73463.
137. Byass, P., et al., *Using verbal autopsy to track epidemic dynamics: the case of HIV-related mortality in South Africa.* Popul Health Metr, 2011. **9**: p. 46.
138. Garenne, M., *Prospects for automated diagnosis of verbal autopsies.* BMC Med, 2014. **12**(1): p. 18.
139. Lucas, S., *Causes of death in the HAART era.* Curr Opin Infect Dis, 2012. **25**(1): p. 36-41.
140. Timaeus, I.M. and M. Jasseh, *Adult mortality in sub-Saharan Africa: evidence from Demographic and Health Surveys.* Demography, 2004. **41**(4): p. 757-72.
141. Adjuik, M., et al., *Cause-specific mortality rates in sub-Saharan Africa and Bangladesh.* Bull World Health Organ, 2006. **84**(3): p. 181-8.
142. Jahn, A., et al., *Population-level effect of HIV on adult mortality and early evidence of reversal after introduction of antiretroviral therapy in Malawi.* Lancet, 2008. **371**(9624): p. 1603-11.
143. Narh-Bana, S.A., et al., *Adult deaths and the future: a cause-specific analysis of adult deaths from a longitudinal study in rural Tanzania 2003-2007.* Trop Med Int Health, 2012. **17**(11): p. 1396-404.
144. van Eijk, A.M., et al., *Causes of deaths using verbal autopsy among adolescents and adults in rural western Kenya.* Trop Med Int Health, 2008. **13**(10): p. 1314-24.
145. Agan, T.U., et al., *Trends in maternal mortality at the University of Calabar Teaching Hospital, Nigeria, 1999-2009.* Int J Womens Health, 2010. **2**: p. 249-54.
146. Traore, B., et al., *[Maternal mortality at the Gynecology-Obstetrics Service of the Segou Regional Hospital Center of Mali. Retrospective study of 138 cases].* Mali Med, 2010. **25**(2): p. 42-7.
147. Ziraba, A.K., et al., *Maternal mortality in the informal settlements of Nairobi city: what do we know?* Reprod Health, 2009. **6**: p. 6.
148. National Bureau of Statistics Tanzania. *2012 Populations and Housing Census*. 2013 [cited; Available from:

http://www.nbs.go.tz/sensa/PDF/Census%20General%20Report%20-%2029%20March%202013_Combined_Final%20for%20Printing.pdf/.
149. National Bureau of Statistics Tanzania. *Tanzania in Figures 2010*. 2011 [cited; Available from: http://www.nbs.go.tz/takwimu/references/Tanzania_in_Figures2010.pdf.
150. UNDESA (Population Division). *World Population Prospects: The 2012 Revision, DVD edition*. 2013 [cited; Available from: http://esa.un.org/wpp/Excel-Data/mortality.htm.
151. National Bureau of Statistics Tanzania and ICF Macro, *Tanzania Demographic and Health Survey 2010*. 2011: Dar es Salaam, Tanzania.
152. Tanzania Commission for AIDS (TACAIDS), et al., *Tanzania HIV/AIDS and Malaria Indicator Survey 2011-12*. 2013: Dar es Salaam, Tanzania.
153. Boerma, J.T., et al., *Spread of HIV infection in a rural area of Tanzania.* AIDS, 1999. **13**(10): p. 1233-40.
154. Isingo, R., et al., *Trends in the uptake of voluntary counselling and testing for HIV in rural Tanzania in the context of the scale up of antiretroviral therapy.* Trop Med Int Health, 2012. **17**(8): p. e15-25.
155. Marston, M., et al., *The impact of antiretroviral therapy on adult mortality in rural Tanzania.* Trop Med Int Health, 2012. **17**(8): p. e58-65.
156. Anon, *The impact of HIV on fertility trends in Kisesa, Tanzania*. 2009, London School of Hygiene & Tropical Medicine.
157. Cawley, C., et al., *Low rates of repeat HIV testing despite increased availability of antiretroviral therapy in rural Tanzania: findings from 2003-2010.* PLoS One, 2013. **8**(4): p. e62212.
158. Zaba, B., et al., *Using age-specific mortality of HIV infected persons to predict anti-retroviral treatment need: a comparative analysis of data from five African population-based cohort studies.* Trop Med Int Health, 2012. **17**(8): p. e3-14.
159. Urassa M, et al. *Recent declines in HIV prevalence and incidence in Magu DSS, 1994-2007*. Presentation to the INDEPTH conference, Dar-es-Salaam, September 22nd to 26th 2008 2008 [cited; Available from: http://www.indepth-network.org/AGM2008/2.Tuesday/Morning/Para%202/4.PrevInciTrendsINDEPTH%20mark%20urassa.pdf.
160. Wambura, M., et al., *HIV prevalence and incidence in rural Tanzania: results from 10 years of follow-up in an open-cohort study.* J Acquir Immune Defic Syndr, 2007. **46**(5): p. 616-23.
161. Boerma, J., et al., *Levels and causes of adult mortality in rural Tanzania with special reference to HIV/AIDS.* Health Transition Review, 1997. **7**(supp. 2): p. 63-74.
162. Indepth Network. *Magu DSS Site Description*. [cited 18th June 2013]; Available from: http://www.indepth-network.org/dss_site_profiles/maguprofile.pdf.
163. Zimstat. *Zimbabwe National Population Census 2012*. 2013 [cited 24 Feb 2014]; Available from: http://www.zimstat.co.zw/dmdocuments/Census/CensusResults2012/National_Report.pdf.
164. Zimbabwe National Statistics Agency and ICF International, *Zimbabwe Demographic and Health Survey 2010-11*. 2012, Calverton, Maryland: ZIMSTAT and ICF International Inc.
165. Lopman, B., et al., *HIV incidence and poverty in Manicaland, Zimbabwe: is HIV becoming a disease of the poor?* Aids, 2007. **21 Suppl 7**: p. S57-66.
166. Gregson, S., et al., *Recent upturn in mortality in rural Zimbabwe: evidence for an early demographic impact of HIV-1 infection?* Aids, 1997. **11**(10): p. 1269-80.
167. Smith, J., et al., *Changing patterns of adult mortality as the HIV epidemic matures in Manicaland, eastern Zimbabwe.* Aids, 2007. **21 Suppl 6**: p. S81-6.

168. Gregson, S., et al., *Transmission Dynamics Underlying a Decade of HIV Prevalence Decline in Manicaland, Zimbabwe, 1998-2008*, in *AIDS 2010*. 2012, International AIDS Society: Washington, DC.
169. Campbell, C., et al., *Building adherence-competent communities: factors promoting children's adherence to anti-retroviral HIV/AIDS treatment in rural Zimbabwe.* Health Place, 2012. **18**(2): p. 123-31.
170. Gregson, S., et al., *Sexual mixing patterns and sex-differentials in teenage exposure to HIV infection in rural Zimbabwe.* Lancet, 2002. **359**(9321): p. 1896-903.
171. WHO, *Verbal Autopsy Standards: 2012 WHO Verbal Autopsy Instrument*. 2012, Geneva: WHO.
172. Lopman, B.A., et al., *Creating and validating an algorithm to measure AIDS mortality in the adult population using verbal autopsy.* PLoS Med, 2006. **3**(8): p. e312.
173. Joshi, R., et al., *Verbal autopsy coding: are multiple coders better than one?* Bull World Health Organ, 2009. **87**(1): p. 51-7.
174. Araya T, et al., *Accuracy of Physicians in Diagnosing HIV and AIDS-Related Death in the Adult Population of Addis Ababa, Ethiopia.* World Journal of AIDS, 2012. **2**(2): p. 89-96.
175. WHO, *International Statistical Classification of Diseases and Health Related Problems, tenth revision*. 2004, Geneva: World Health Organization
176. Landis, J.R. and G.G. Koch, *The measurement of observer agreement for categorical data.* Biometrics, 1977. **33**(1): p. 159-74.
177. Kirkwood, B. and J. Sterne, *Essential Medical Statistics, second edition*. 2003, Oxford: Blackwell Science.
178. Newcombe, R.G., *Two-sided confidence intervals for the single proportion: comparison of seven methods.* Stat Med, 1998. **17**(8): p. 857-72.
179. Mills, S., et al., *Maternal mortality decline in the Kassena-Nankana district of northern Ghana.* Matern Child Health J, 2008. **12**(5): p. 577-85.
180. Warrell, D., T. Cox, and J. Firth, *Oxford Textbook of Medicine* 5th ed. 2010, Oxford: Oxford University Press
181. UNAIDS, *Global report: UNAIDS report on the global AIDS epidemic 2010*. 2010.
182. DFID and GURT. *Policy implications of adult morbidity and mortality: End of phase one report*. 1997 [cited; Available from: http://research.ncl.ac.uk/ammp/site_files/public_html/ammp_rep/ammp_rpt.pdf.
183. Swartz, M., *Textbook of Physical Diagnosis: History and examination*. 5th ed. 2006: Saunders Elsevier.
184. Narh-Bana, S.A., et al., *Adult deaths and the future: a cause-specific analysis of adult deaths from a longitudinal study in rural Tanzania 2003-2007.* Trop Med Int Health, 2012.
185. Etard, J.F., B. Kodio, and C. Ronsmans, *Seasonal variation in direct obstetric mortality in rural Senegal: role of malaria?* American Journal of Tropical Medicine and Hygiene, 2003. **68**(4): p. 503-504.
186. Ba, M.G., B. Kodio, and J.F. Etard, *Verbal autopsy to measure maternal mortality in rural Senegal. [French].* Journal de Gynecologie Obstetrique et Biologie de la Reproduction, 2003. **32**(8 I): p. 728-735.
187. Becher, H., et al., *Patterns of malaria: cause-specific and all-cause mortality in a malaria-endemic area of west Africa.* Am J Trop Med Hyg, 2008. **78**(1): p. 106-13.
188. Hynes, M., et al., *A study of refugee maternal mortality in 10 countries, 2008-2010.* Int Perspect Sex Reprod Health, 2012. **38**(4): p. 205-13.
189. Rao, C., A. Lopez, and Y. Hemed, *Causes of death*, in *Disease and Mortality in Sub-Saharan Africa*. 2006, World Bank: Washington DC.
190. Fottrell, E., et al., *Revealing the burden of maternal mortality: a probabilistic model for determining pregnancy-related causes of death from verbal autopsies.* Popul Health Metr, 2007. **5**: p. 1.

191. Bell, J.S., et al., *The epidemiology of pregnancy outcomes in rural Burkina Faso.* Tropical Medicine and International Health, 2008. **13 Suppl 1**: p. 31-43.
192. Vergnano, S., et al., *Adaptation of a probabilistic method (InterVA) of verbal autopsy to improve the interpretation of cause of stillbirth and neonatal death in Malawi, Nepal, and Zimbabwe.* Popul Health Metr, 2011. **9**: p. 48.
193. Ferri, C.P., et al., *Socioeconomic factors and all cause and cause-specific mortality among older people in Latin America, India, and China: a population-based cohort study.* PLoS Med, 2012. **9**(2): p. e1001179.
194. Rankin, J.C., et al., *Exploring the role narrative free-text plays in discrepancies between physician coding and the InterVA regarding determination of malaria as cause of death, in a malaria holo-endemic region.* Malar J, 2012. **11**: p. 51.
195. Fottrell, E., et al., *Probabilistic methods for verbal autopsy interpretation: InterVA robustness in relation to variations in a priori probabilities.* PLoS One, 2011. **6**(11): p. e27200.
196. Todd, J., et al., *Time from HIV seroconversion to death: a collaborative analysis of eight studies in six low and middle-income countries before highly active antiretroviral therapy.* AIDS, 2007. **21 Suppl 6**: p. S55-63.
197. Kahn, K., et al., *Research into health, population and social transitions in rural South Africa: data and methods of the Agincourt Health and Demographic Surveillance System.* Scand J Public Health Suppl, 2007. **69**: p. 8-20.
198. Chuc, N.T. and V. Diwan, *FilaBavi, a demographic surveillance site, an epidemiological field laboratory in Vietnam.* Scand J Public Health Suppl, 2003. **62**: p. 3-7.
199. Mwaluko, G., et al., *Trends in HIV and sexual behaviour in a longitudinal study in a rural population in Tanzania, 1994-2000.* Aids, 2003. **17**(18): p. 2645-51.
200. Gregson, S., et al., *HIV decline associated with behavior change in eastern Zimbabwe.* Science, 2006. **311**(5761): p. 664-6.
201. Kumogola, Y., et al., *Trends in HIV & syphilis prevalence and correlates of HIV infection: results from cross-sectional surveys among women attending ante-natal clinics in Northern Tanzania.* BMC Public Health, 2010. **10**: p. 553.
202. Santelli, J.S., et al., *Behavioral, biological, and demographic risk and protective factors for new HIV infections among youth in Rakai, Uganda.* J Acquir Immune Defic Syndr, 2013. **63**(3): p. 393-400.
203. Barnighausen, T., et al., *The socioeconomic determinants of HIV incidence: evidence from a longitudinal, population-based study in rural South Africa.* AIDS, 2007. **21 Suppl 7**: p. S29-38.
204. Hargreaves, J.R., et al., *Explaining continued high HIV prevalence in South Africa: socioeconomic factors, HIV incidence and sexual behaviour change among a rural cohort, 2001-2004.* Aids, 2007. **21 Suppl 7**: p. S39-48.
205. Hargreaves, J.R., et al., *Systematic review exploring time trends in the association between educational attainment and risk of HIV infection in sub-Saharan Africa.* Aids, 2008. **22**(3): p. 403-14.
206. Wojcicki, J.M., *Socioeconomic status as a risk factor for HIV infection in women in East, Central and Southern Africa: a systematic review.* J Biosoc Sci, 2005. **37**(1): p. 1-36.
207. Msisha, W.M., et al., *Socioeconomic status and HIV seroprevalence in Tanzania: a counterintuitive relationship.* Int J Epidemiol, 2008. **37**(6): p. 1297-303.
208. Kunihira, N.R., et al., *Barriers to use of antiretroviral drugs in Rakai district of Uganda.* Afr Health Sci, 2010. **10**(2): p. 120-9.
209. Kleinschmidt, I. and G. Churchyard, *Variation in incidences of tuberculosis in subgroups of South African gold miners.* Occup Environ Med, 1997. **54**(9): p. 636-41.
210. Naidoo, R.N., et al., *Differential respirable dust related lung function effects between current and former South African coal miners.* Int Arch Occup Environ Health, 2005. **78**(4): p. 293-302.

211. Garenne, M., et al., *Protective effect of pregnancy in rural South Africa: questioning the concept of "indirect cause" of maternal death.* PLoS One, 2013. **8**(5): p. e64414.
212. Baqui, A.H., et al., *Causes of childhood deaths in Bangladesh: results of a nationwide verbal autopsy study.* Bull World Health Organ, 1998. **76**(2): p. 161-71.
213. South, A., et al., *Do accurate HIV and antiretroviral therapy knowledge, and previous testing experiences increase the uptake of HIV voluntary counselling and testing? Results from a cohort study in rural Tanzania.* BMC Public Health, 2013. **13**: p. 802.
214. Kahn, K., et al., *Validation and application of verbal autopsies in a rural area of South Africa.* Trop Med Int Health, 2000. **5**(11): p. 824-31.
215. Oti, S.O., et al., *InterVA versus Spectrum: how comparable are they in estimating AIDS mortality patterns in Nairobi's informal settlements?* Glob Health Action, 2013. **6**: p. 21638.
216. Kapiriri, L. and O.F. Norheim, *Criteria for priority-setting in health care in Uganda: exploration of stakeholders' values.* Bull World Health Organ, 2004. **82**(3): p. 172-9.
217. Bradshaw, D., et al., *Provincial mortality in South Africa, 2000--priority-setting for now and a benchmark for the future.* S Afr Med J, 2005. **95**(7): p. 496-503.
218. Bryce, J., et al., *LiST as a catalyst in program planning: experiences from Burkina Faso, Ghana and Malawi.* Int J Epidemiol, 2010. **39 Suppl 1**: p. i40-7.
219. Kahn, K., et al., *Who dies from what? Determining cause of death in South Africa's rural north-east.* Trop Med Int Health, 1999. **4**(6): p. 433-41.
220. Loewenson, R., *Structural adjustment and health policy in Africa.* Int J Health Serv, 1993. **23**(4): p. 717-30.
221. Pender, J., *From'structural adjustment'to'comprehensive development framework': conditionality transformed?* Third World Quarterly, 2001. **22**(3): p. 397-411.
222. O'Keefe, E. and A. Scott-Samuel, *Health impact assessment, human rights and global public policy: a critical appraisal.* Bulletin of the World Health Organization, 2007. **85**(3): p. 212-217.
223. Rutherford, G.W., et al., *Public health triangulation: approach and application to synthesizing data to understand national and local HIV epidemics.* BMC Public Health, 2010. **10**: p. 447.
224. van Schalkwyk, C., et al., *Outcomes and impact of HIV prevention, ART and TB programs in Swaziland--early evidence from public health triangulation.* PLoS One, 2013. **8**(7): p. e69437.
225. Clark, S., et al., *InSilicoVA: A Method to Automate Cause of Death Assignment for Verbal Autopsy*, C.f.S.a.t.S. Sciences, Editor. 2013, University of Washington.
226. Nyirenda, M., et al., *Mortality levels and trends by HIV serostatus in rural South Africa.* Aids, 2007. **21 Suppl 6**: p. S73-9.
227. Mee, P., et al., *Evidence for localised HIV related micro–epidemics associated with the decentralised provision of antiretroviral treatment in rural South Africa: a spatio–temporal analysis of changing mortality patterns (2007–2010).* Journal of global health, 2014. **4**(1).
228. Mocroft, A., et al., *Is there evidence for an increase in the death rate from liver-related disease in patients with HIV?* Aids, 2005. **19**(18): p. 2117-25.
229. Suy, A., et al., *Increased risk of pre-eclampsia and fetal death in HIV-infected pregnant women receiving highly active antiretroviral therapy.* Aids, 2006. **20**(1): p. 59-66.
230. Mocroft, A., et al., *Variable impact on mortality of AIDS-defining events diagnosed during combination antiretroviral therapy: not all AIDS-defining conditions are created equal.* Clin Infect Dis, 2009. **48**(8): p. 1138-51.
231. Grinsztejn, B., et al., *Changing mortality profile among HIV-infected patients in Rio de Janeiro, Brazil: shifting from AIDS to non-AIDS related conditions in the HAART era.* PLoS One, 2013. **8**(4): p. e59768.

232. Sieleunou, I., et al., *Determinants of survival in AIDS patients on antiretroviral therapy in a rural centre in the Far-North Province, Cameroon.* Trop Med Int Health, 2009. **14**(1): p. 36-43.

Appendices

Appendix 1: Verbal autopsy questionnaires used in the Kisesa and Manicaland DSSes

Attached as the final documents:

Kisesa VA questionnaire 1: Tanesa (p243)

Kisesa VA questionnaire 2: Indepth (p265)

Kisesa VA questionnaire 3: Tazama (p276)

Manicaland VA questionnaire (p294)

Appendix 2: Symptoms considered in calculating number of reported symptoms and investigating symptom profiles

Abdominal mass
Abdominal pain
Abnormal hair colouring
Abnormality of urine
Adequate vaccination
Alcohol use
Anemia
Animal bite/sting
Antibiotic injection
Assault
Attempted termination of pregnancy
Baby's delivery position abnormal
Bleeding between menstrual periods
Bleeding during/after pregnancy (unspecified)
Bleeding from mouth, nose and anus
Blood in urine
Bloody cough
Bloody diarrhea
Bloody vomit
Blurred vision
Blurred vision during pregnancy
Breastfeeding at death
Breathlessness (unspecified)
Breathlessness lying down
Breathlessness on exertion
Burn
Chest indrawing
Chest pain
Coma
Convulsions
Cough
Death in the dry season
Death in the wet season
Death within 24 hours of pregnancy ending
Delivered live baby within 6 weeks
Diagnosis of asthma
Diagnosis of cancer
Diagnosis of chronic obstructive pulmonary disease
Diagnosis of confusion
Diagnosis of dementia
Diagnosis of depression
Diagnosis of diabetes
Diagnosis of epilepsy
Diagnosis of heart disease
Diagnosis of hemoglobinopathy
Diagnosis of HIV/AIDS
Diagnosis of hypertension
Diagnosis of kidney disease
Diagnosis of liver disease
Diagnosis of malaria
Diagnosis of measles
Diagnosis of stroke
Diagnosis of tuberculosis
Diarrhea
Died in labour undelivered
Difficulty breathing
Difficulty drinking
Discharged from hospital ill
Drowning
Excessive thirst
Excessive urination
Fall
Fever
Final illness lasted >3 weeks
Final illness lasted ≤3 weeks
First pregnancy
Fits during pregnancy
Foul smelling vaginal discharge
Had professional assistance at delivery
Headache
Herpes zoster
High blood pressure during pregnancy
Homicide
Hysterectomy recently
Injury (unspecified)
Killed by force of nature
Labour longer than 24 hours
Major bleeding after delivery
Major bleeding during labour, before delivery
Major bleeding in first 6 months of pregnancy
Married at time of death
Measles rash
Menstruation stopped naturally
Method of delivery
More than 4 previous births
Multiple pregnancy
Night sweats
Oral candidiasis
Paralysis
Place of delivery
Poisoning
Pregnant at or prior to death
Previous Caesarian section
Productive cough
Professional assistance at delivery
Rapid breathing
Rash (non-measles)
Given blood transfusion
Given treatment by nose
Given intravenous drip
Given rehydration
Recent abortion
Recent early pregnancy ending
Recent operation
Retained placenta
Rigidity/lockjaw
Skin lesions/ulcers
Smoking habit
Stiff neck
Sudden death
Suicide
Sunken eyes
Surgery before death
Swelling (unspecified)
Swollen abdomen
Swollen ankles
Swollen armpit
Swollen breast
Swollen face
Swollen genitals
Swollen glands
Swollen legs
Swollen mouth
Swollen neck
Transport collision
Ulcers not on feet
Ulcers on feet
Urinary retention
Vaginal bleeding after menstruation stopped
Vomiting
Wasting
Weight loss
Wheezing
Whether received treatment
Whooping cough
Yellowness/jaundice

Appendix 3*:* Verbal autopsy data specification 8.1, for reporting cause-specific mortality in the Alpha network, and information availability in VA questionnaires

Variable name	Description	Coding	Data unavailable in questionnaire			
			Tanesa	Indepth	Tazama	Rakai
Idno	Person ID number	site specific				
study_name	Name of your study field site	site specific				
va_interview_date	Date of VA interview	in Stata format				
va_date_of_death	Reported date of death	in Stata format				
va_age_at_death	Age at death in years	12-89 as reported 90 = 90+ 99 not stated				
va_sex	Male or female	1 Male 2 Female				
va_final_ill	Did final illness last at least 3 weeks?	0 no,1 yes				
va_sudden	Was death very sudden or unexpected	0 no,1 yes				
va_vis_bl	Any blurred vision	0 no,1 yes		X		
va_drowsy	Any drowsiness	0 no,1 yes	X	X	X	X
va_bed_day	Was bed-bound for more than 1wk before death	0 no,1 yes	X	X	X	
va_coma	Was there a coma > 24hrs	0 no,1 yes				X
va_collapse	Did death follow sudden collapse	0 no,1 yes	X	X	X	
va_season	Season of death	0 dry,1 wet				
va_injury	Any obvious recent injury	0 no,1 yes				
va_transport	Was s/he in a transport accident	0 no,1 yes				
va_drowning	Did s/he drown	0 no,1 yes				
va_fall	Had s/he fallen recently	0 no,1 yes				
va_poison	Any poisoning, bite, sting	0 no,1 yes				
va_homicide	Any suggestion of homicide	0 no,1 yes		X		
va_suicide	Any suggestion of suicide	0 no,1 yes				
va_smoking	Was s/he a known smoker	0 no,1 yes	X			X
va_alcohol	Was s/he known to drink alcohol	0 no,1 yes				X
va_convul	Any convulsions or fits	0 no,1 yes	X			
va_headache	Any headache	0 no,1 yes				
va_paralysis	Was there paralysis	0 no paralysis 1 one side 2 both sides				
va_stiff_neck	Any stiff neck	0 no,1 yes				
va_or_cand	Any oral candidiasis	0 no,1 yes		X		X
va_rigidity	Any rigidity/lockjaw	0 no,1 yes				X
va_hair	Any abnormal hair colouring	0 no,1 yes		X	X	X
va_ch_pain	Any chest pain	0 no,1 yes				
va_cough_long	How long did cough last	0 no cough 1 ≤ 3 weeks 2 > 3 weeks 3 had cough, duration not known				
va_cough_pr	Any productive cough	0 no,1 yes				X
va_bl_cough	Any coughing with blood	0 no,1 yes				
va_rapid_br	Any rapid breathing	0 no,1 yes			X	
va_exert_br	Any breathlessness on exertion	0 no,1 yes		X		X
va_lying_br	Any breathlessness lying flat	0 no,1 yes		X		X
va_chest_in	Any chest indrawing	0 no,1 yes	X	X	X	X
va_diff_br	Any difficulty breathing	0 no,1 yes		X		
va_wheeze	Any wheezing	0 no,1 yes	X			
va_cyanosis	Any cyanosis	0 no,1 yes	X	X	X	
va_abd_mass	Any abdominal mass	0 no,1 yes	X			X
va_abd_pain	Any abdominal pain	0 no,1 yes				
va_swe_abd	Any abdominal swelling	0 no,1 yes				

Variable name	Description	Coding	Data unavailable in questionnaire			
			Tanesa	Indepth	Tazama	Rakai
va_diarr_weeks	Diarrhoea duration	0 no diarrhoea 1 < 2 weeks 2 2-4 weeks 3 4+ weeks 4 had diarrhoea, duration not known				
va_bl_diarr	Any diarrhoea with blood	0 no,1 yes				
va_vomiting	Any vomiting	0 no,1 yes				
va_bl_vomit	Any vomiting with blood	0 no,1 yes				
va_yellow	Any yellowness/jaundice	0 no,1 yes				
va_urine	Any abnormality of urine	0 no,1 yes				
va_uri_ret	Any urinary retention	0 no,1 yes				
va_uri_haem	Any haematuria	0 no,1 yes				
va_swe_legs	Any swelling of ankles/legs	0 no,1 yes				
va_eye_sunk	Were eyes sunken	0 no,1 yes			X	X
va_rash	Any rash	0 no,1 yes				
va_measrash	Any measles rash	0 no,1 yes	X			
va_herpes	Any herpes zoster	0 no,1 yes		X		X
va_skin	Any skin lesions/ulcers	0 no,1 yes				
va_swe_breast	Any breast lump or lesion	0 no,1 yes	X			X
va_swe_gen	Any lump or lesion in groin or genitals	0 no,1 yes				X
va_swe_lump	Any other localised lump or lesion	0 no,1 yes		X		X
va_exc_drink	Any excessive water intake	0 no,1 yes		X	X	X
va_exc_urine	Any excessive urination	0 no,1 yes				
va_exc_food	Any excessive food intake	0 no,1 yes	X	X	X	X
va_fever_weeks	Fever duration	0 no fever 1 < 2 weeks 2 2+ weeks 3 had fever, duration not known				
va_night_sw	Any excessive night sweats	0 no,1 yes				X
va_swe_gland	Any enlarged/swollen glands	0 no,1 yes				X
va_swe_oth	Any facial swelling	0 no,1 yes				X
va_wt_loss	Any weight loss	0 no,1 yes				X
va_wasting	Any severe wasting [Note: Severe wasting is weight loss with other factors like anaemia, hair colour changes, swollen legs, burning feet]	0 no,1 yes				X
va_anaemia	Any anaemia/paleness	0 no,1 yes				X
va_asthma	Any medical diagnosis of asthma	0 no,1 yes	X			
va_epilepsy	Any medical diagnosis of epilepsy	0 no,1 yes	X			
va_diabetes	Any medical diagnosis of diabetes	0 no,1 yes				
va_heart_dis	Any medical diagnosis of heart disease	0 no,1 yes			X	X
va_kidney_dis	Any medical diagnosis of kidney disease	0 no,1 yes	X	X	X	X
va_sickle	Any medical diagnosis of haemoglobinopathy	0 no,1 yes		X	X	X
va_malaria	Any medical diagnosis of malaria	0 no,1 yes		X	X	
va_hiv_aids	Any medical diagnosis of HIV/AIDS	0 no,1 yes				
va_hypert	Any medical diagnosis of hypertension	0 no,1 yes				
va_tuber	Any medical diagnosis of TB	0 no,1 yes	X			
va_liver_dis	Any medical diagnosis of liver disease	0 no,1 yes	X	X	X	X
va_cancer	Any medical diagnosis of cancer	0 no,1 yes	X	X		X
va_stroke	Any medical diagnosis of stroke	0 no,1 yes	X	X	X	X
va_measles	Any medical diagnosis of measles	0 no,1 yes	X	X	X	X
va_antib_i	Was antibiotic injection required during final illness	0 no,1 yes	X			X
va_blood_tr	Was blood transfusion required during final illness	0 no,1 yes	X	X		X
va_surgery	Any surgery just before death	0 no,1 yes	X			X
va_disch	Was discharged from hospital very ill	0 no,1 yes	X	X	X	X
va_vaccin	Was s/he adequately vaccinated	0 no,1 yes	X	X	X	X
va_preg_status	Was she pregnant or did she deliver less than 6	0 reported not pregnant within last 6 weeks				

Variable name	Description	Coding	Data unavailable in questionnaire			
			Tanesa	Indepth	Tazama	Rakai
	weeks before she died	1 pregnant at time of death 2 died < 6 weeks after normal length pregnancy 3 died < 6 weeks after early pregnancy ending				
va_married	Was she married/partnered at death	0 no,1 yes	X			
va_ever_preg	Had she ever been pregnant	0 no,1 yes	X	X	X	
va_breast_fd	Was she breast feeding at death	0 no,1 yes	X	X	X	X
va_first_p	Did she die during/just after first pregnancy	0 no,1 yes		X		X
va_more4	Did she have more than 4 previous pregnancies	0 no,1 yes		X		X
va_trim1	Did she die after less than 3 months of pregnancy	0 no,1 yes	X	X	X	X
va_multip	Was this a multiple pregnancy	0 no,1 yes	X	X	X	X
va_preg_uw	Was this pregnancy unwanted	0 no,1 yes	X	X	X	X
va_term_att	Any attempt to terminate this pregnancy	0 no,1 yes	X			X
va_hyster	Hysterectomy shortly before death	0 no,1 yes	X	X	X	X
va_death_24	Death within 24 hrs of pregnancy ending	0 no,1 yes	X			X
va_bleed_1	Major bleeding during early pregnancy	0 no,1 yes				X
va_bleed_d	Major bleeding in late pregnancy/delivery	0 no,1 yes				X
va_placent_r	Did placenta remain inside	0 no,1 yes		X		X
va_bpr_preg	Was blood pressure raised during pregnancy	0 no,1 yes		X	X	X
va_fit_preg	Were fits only pregnancy-related	0 no,1 yes	X	X		X
va_baby_al	Did she deliver a live baby within 6 wks of death	0 no,1 yes			X	X
va_lab_24	Was labour prolonged > 24 hrs	0 no,1 yes	X			X
va_died_lab	Did she die in labour undelivered	0 no,1 yes	X	X		X
va_delivery	Where did delivery take place	0 at home 1 in transit 2 at health facility	No data on 'in transit'	No data on 'in transit'	No data on 'in transit'	X
va_prof_ass	Had professional assistance at delivery	0 no,1 yes	X	X		X
va_del_method	How was the baby delivered?	0 normal vaginal delivery, no instruments 1 vaginal delivery with forceps / Ventuse 2 delivery by Caesarean section	X	X		X
va_baby_pos	Was baby's delivery position abnormal	0 no,1 yes	X	X	X	X
va_baby_big	Was baby too big for delivery	0 no,1 yes	X	X	X	X
va_baby_part	Was part of the baby prolapsed	0 no,1 yes	X	X	X	X
va_disch_sm	Any foul smelling vaginal discharge	0 no,1 yes	X	X		X
va_cs_prev	Any previous Caesarean section	0 no,1 yes	X	X	X	X
va_coma_sudden	Did the coma come on suddenly	0 no,1 yes	X			X
va_transport_road	Was s/he in a road transport accident	0 no,1 yes				X
va_transport_oth	Was s/he in a non road transport accident	0 no,1 yes	X	X	X	X
va_burn	Was s/he burnt by heat, steam or fire	0 no,1 yes	X			X
va_bite	Any bite or sting by an animal	0 no,1 yes	X			X
va_poison_2	Any poisoning (not by an animal)	0 no,1 yes	X			X
va_inj_intent	Was h/she intentionally injured by another person or people	0 no,1 yes		X		X
va_nature	Was s/he injured by a force of nature	0 no,1 yes	X	X	X	X
va_assult	Injured in some kind of violence or assault by another person	0 no,1 yes				X
va_convul_time	Any convulsions or fits	0 no convulsions 1 < 5 minutes 2 ≥ 5 minutes 3 had convulsions, duration unknown	X	No duration	No duration	X

Variable name	Description	Coding	Data unavailable in questionnaire			
			Tanesa	Indepth	Tazama	Rakai
va_convul_coma	Became unconscious immediately after convulsions	0 no,1 yes	X	X	X	X
va_stiff_neck_time	Any stiff or painful neck	0 no stiff neck 1 < 1 week 2 ≥ 1 week 3 stiff neck, duration unknown	No duration	No duration		X
va_cough_long_2_wk	How long did cough last	0 no cough 1 < 2 weeks 2 ≥ 2 weeks 3 had cough, duration not known				X
va_whoop	Any distinctive whoop (associated with characteristic whooping sound of pertussis)	0 no,1 yes	X	X	X	X
va_rapid_br_time	Any rapid breathing	0 no rapid breathing 1 < 2 weeks 2 ≥ 2 weeks 3 rapid breathing, duration unknown	X	X	X	X
va_breathless	Any breathlessness	0 no breathlessness 1 < 2 weeks 2 ≥ 2 weeks 3 breathlessness, duration unknown	No duration			X
va_abd_prob	Any abdominal problem	0 no,1 yes				X
va_abd_mas_time	Any abdominal mass	0 no abdominal mass 1 < 2 weeks 2 ≥ 2 weeks 3 abdominal mass, duration unknown	X			X
va_abd_pain_time	Any abdominal pain	0 no abdominal pain 1 < 2 weeks 2 ≥ 2 weeks 3 abdominal pain, duration unknown	No duration			X
va_swe_abd_time	Any abdominal swelling	0 no abdominal swelling 1 < 2 weeks 2 ≥ 2 weeks 3 abdominal swelling, duration unknown	No duration			X
va_swe_ankles	Any swelling of both feet/ankles	0 no,1 yes				X
va_ulc_feet	Any ulcers/ abscesses or sores on the feet	0 no,1 yes	X	X		X
va_ulc_oth	Any ulcers/ abscesses or sores on body, apart from feet	0 no,1 yes				X
va_rash_time	Any non measles rash	0 no non measles rash 1 < 1 week 2 ≥ 1 week 3 non measles rash, duration unknown				X
va_swe	Any localised lump or lesion	0 no,1 yes				X
va_swe_mouth	Any lump or lesion in mouth	0 no,1 yes				X
va_swe_armpit	Any lump or lesion in armpit	0 no,1 yes				X
va_swe_neck	Any lumps/swelling in neck	0 no,1 yes				X
va_drink_diff	Any difficulty or pain in swallowing liquids	0 no,1 yes				X
va_malaria_pos	Positive malaria test within one week of death	0 no,1 yes	X	X	X	X
va_malaria_neg	Negative malaria test within one week of death	0 no,1 yes	X	X	X	X
va_copd	Any medical diagnosis of chronic obstructive pulmonary disease	0 no,1 yes	X	X	X	X
va_depress	Any medical diagnosis of depression	0 no,1 yes	X	X	X	X
va_dementia	Any medical diagnosis of dementia	0 no, 1 yes	X	X	X	X
va_confusion	Any medical diagnosis of memory loss or mental confusions	0 no, 1 yes	X	X	X	X

Variable name	Description	Coding	Data unavailable in questionnaire			
			Tanesa	Indepth	Tazama	Rakai
va_confuse_3	Did the symptoms of mental confusion last 3 months or more?	0 no, 1 yes	X	X	X	X
va_bleed	Was the any bleeding from mouth, nose and anus	0 no, 1 yes	X	X		X
va_menstrual	Was there any bleeding between menstrual periods (women aged 12-50 only)	0 no, 1 yes	X	X		X
va_ menstr_stop	Had the woman's normal vaginal bleeding stopped naturally (women 40+)	0 no, 1 yes		X	X	X
va_menstr_post	Had the woman's normal vaginal bleeding stopped naturally but they later experienced vaginal bleeding	0 no, 1 yes		X	X	X
va_treatment	Treatment for final illness from a health facility	0 no, 1 yes		X		X
va_rehydrat	Was oral rehydration required during final illness	0 no, 1 yes		X		X
va_nose	Was treatment/food required through nose during final illness	0 no, 1 yes	X	X		X
va_iv	Was an IV drip required during final illness	0 no, 1 yes	X	X		X
va_operation	Was there an operation within one month of death	0 no, 1 yes	X			X
va_early_preg	Was the woman at an early stage of pregnancy within 6 weeks of her death, but the pregnancy had ended in a spontaneous or induced abortion at a stage before the foetus was viable?	0 no, 1 yes	X			X
va_rec_abort	Any recent abortion	0 no, 1 yes	X			X
va_bleed_m	Mother had excessive vaginal bleeding in pregnancy/postpartum period	0 no, 1 yes		X		X
va_bleed_preg	Major bleeding in first 6 months of pregnancy	0 no, 1 yes		X	X	X
va_bleed_pre_lab	Major bleeding shortly before labour	0 no, 1 yes		X		X
va_bleed_lab	Major bleeding during labour, before delivering the baby	0 no, 1 yes	X	X		X
va_bleed_post_lab	Major bleeding after delivering the baby	0 no, 1 yes	X	X		X
va_vis_bl_preg	Any blurred vision during the last 3 months of preg	0 no, 1 yes	X	X		X

Appendix 4: Resolution of issues encountered in translating the raw Kisesa VA data into Spec 8.1

Questionnaire	Issue	Action
Tanesa	Datasets contained two variables, “epilepsy” and “epilep”; I presume one referred to the question about epileptic seizures and one referred to the question about fits during pregnancy. Both variables contained responses by men and women.	va_convul and va_fit_preg not created in Spec 8.1
	Questionnaire contains a single question on rash that does not discriminate between measles and non-measles rash	va_rash created from the variable on rash ("skindise"); va_measrash not created
	The coding of the responses to the question "Did NAME die during pregnancy or childbirth or within six weeks after giving birth?" gave unclear indication of the timing of some deaths.	I created an additional code for va_preg_status, code "4", meaning the death was pregnancy-related in timing but not specifying when. In the input to InterVA, I coded these deaths "Yes" for the question "Was she pregnant at the time of death?".
	There was one question about ulcers, with the location of the ulcer not specified.	All reports of ulcers treated as non-feet ulcers (va_ulc_oth), except where diabetes was reported; va_ulc_feet not created
Indepth	There was one question about ulcers, with the location of the ulcer not specified.	All reports of ulcers treated as non-feet ulcers (va_ulc_oth), except where diabetes was reported; va_ulc_feet not created
Tazama	The length of pregnancy was not recorded for women who died after giving birth.	For women who had died after giving birth, I coded va_preg_status to "2" (died after normal length pregnancy), as most deliveries occur in term pregnancies. For women who died after an abortion (spontaneous or induced), I coded va_preg_status to "3" (died after early pregnancy ending).
	The questionnaire asks in one question (q1003_drip_treatment) whether the deceased received oral rehydration or an intravenous drip.	Where q1003_drip_treatment was answered positively, I coded va_rehydrat positively if the respondent also reported diarrhoea, and va_iv if diarrhoea was not reported.

Appendix 5: Descriptions (as provided) and ICD-10 codes assigned in reviews for which the assigned ICD-10 code was obviously erroneous, and cause groups assigned

"Severe alamal intoxication" was assumed to mean "Severe alcohol intoxication".

Description in review	ICD-10 code in review	Corrected ICD-10 code	Cause group assigned
Severe malaria	B24 [Unspecified HIV disease]	B54 [Unspecified malaria]	Malaria
Epilepsy	R40.9 [code non-existent]	G40.9 [Unspecified epilepsy]	Epilepsy
Cardiomyopathy	E42 [Marasmic kwashiorkor]	I42 [Cardiomyopathy]	Other/unspecified cardiac diseases
Liver cirrhosis	O74.6 [Other complications of spinal and epidural anaesthesia during labour and delivery]	K74.6 [Other and unspecified cirrhosis of liver]	Other/unspecified non-communicable diseases
Severe alamal intoxication	I51 [Complications and ill-defined descriptions of heart disease]	T51 [Toxic effect of alcohol]	Other/unspecified external causes

Appendix 6: Descriptions (as provided) or ICD-10 codes assigned in incomplete reviews, and cause groups assigned

Review description	Review ICD-10 code	Cause group assigned
–	W20 [Struck by thrown, projected or falling object]	Other/unspecified external causes
Melanoma	–	Other/unspecified neoplasms
Uterine fibroid uterine fibroid	–	Reproductive neoplasms
Malignant neoplasia of tahe colorectum	–	Digestive neoplasms
Pyogenic urethralis	–	Other/unspecified infectious diseases
Liver cirrhosis	–	Other/unspecified non-communicable diseases
Pyogenic urethratis	–	Other/unspecified infectious diseases
Chronic abdominal pain	–	Acute abdomen
Alcohol intoxication	–	Other/unspecified external causes
Road traffic accident	–	Road traffic collision

Appendix 7: Distribution of reviews by two reviewing physicians at the level of cause groups

This appendix presents the cause groups to which I assigned the respective physician reviews in the 460 records with two physician reviews. Deaths assigned to the same cause group by both reviewers are marked bold. Broad categories of cause of death are highlighted by colour.

First physician / Second physician	HIV/AIDS-related	Acute respiratory infections/pneumonia	Diarrhoeal diseases	Malaria	Meningitis/ encephalitis	Pulmonary tuberculosis	Other/unspecified infectious diseases	Digestive neoplasms	Breast neoplasms	Reproductive neoplasms	Other/unspecified neoplasms	Severe anaemia	Diabetes mellitus	Acute cardiac diseases	Stroke	Sickle cell with crisis	Other/unspecified caridiac diseases	Chronic obstructive pulmonary disease	Asthma
HIV/AIDS-related	**122**	2	1	1		4	6			2	3			1					
Acute respiratory infections/pneumonia		**6**												1					
Diarrhoeal diseases			**3**	1															
Malaria			1	**10**	3		3												
Meningitis/encephalitis				1	**8**						1								
Pulmonary tuberculosis	2	1				**17**					1						1	1	
Other/unspecified infectious diseases	3						**6**	1			1								
Digestive neoplasms								**4**			1								
Breast neoplasms									**1**										
Reproductive neoplasms	1									**7**									
Other/unspecified neoplasms	3						2			1	**7**								
Severe anaemia												**1**							
Diabetes mellitus													**5**						
Acute cardiac diseases																	1		
Stroke															**2**				
Sickle cell with crisis																**2**			
Other/unspecified cardiac diseases	1	1				1								2			**10**		
Chronic obstructive pulmonary disease	1																		1
Asthma																			

First physician / Second physician	Acute abdomen	Renal failure	Epilepsy	Other/unspecified non-communicable diseases	Abortion-related death	Pregnancy-induced hypertension	Obstructed labour	Anaemia of pregnancy	Other/unspecified maternal causes	Road traffic collision	Accidental drowning	Exposure to smoke/fire	Venomous plant/animal	Accidental poisoning	Intentional self-harm	Assault	Other/unspecified external causes	Cause of death unknown
HIV/AIDS-related death				10					1									
Acute respiratory infections/pneumonia																		
Diarrhoeal diseases																		
Malaria	1			2					1									
Meningitis/encephalitis				1														
Pulmonary tuberculosis				2														
Other/unspecified infectious diseases	1			8														
Digestive neoplasms				1														
Breast neoplasms																		
Reproductive neoplasms																		
Other/unspecified neoplasms			1	3														
Severe anaemia				1				1										
Diabetes mellitus																		
Acute cardiac diseases				1														
Stroke																		
Sickle cell with crisis																		
Other/unspecified cardiac diseases				1														
Chronic obstructive pulmonary disease																		
Asthma																		

First physician / Second physician	HIV/AIDS-related death	Acute respiratory infection/pneumonia	Diarrhoeal diseases	Malaria	Meningitis/encephalitis	Pulmonary tuberculosis	Other/unspecified infectious diseases	Digestive neoplasms	Breast neoplasms	Reproductive neoplasms	Other/unspecified neoplasms	Severe anaemia	Diabetes mellitus	Acute cardiac diseases	Stroke	Sickle cell with crisis	Other/unspecified cardiac diseases	Chronic obstructive pulmonary disease	Asthma
Acute abdomen				1															
Renal failure											1								
Epilepsy																			
Other/unspecified NCDs	5			2			3	2		2	1					1	2		
Abortion-related death																			
Pregnancy-induced hypertension																			
Obstructed labour																			
Anaemia of pregnancy																			
Other/unspecified maternal causes																			
Road traffic collision																			
Accidental drowning																			
Exposure to smoke/fire																			
Venomous plant/animal																			
Accidental poisoning																			
Intentional self-harm	1																		
Assault																			
Other/unspecified external causes	1																		
Cause of death unknown	1																		

First physician / Second physician	Acute abdomen	Renal failure	Epilepsy	Other/unspecified non-communicable diseases	Abortion-related death	Pregnancy-induced hypertension	Obstructed labour	Anaemia of pregnancy	Other/unspecified maternal causes	Road traffic collision	Accidental drowning	Exposure to smoke/fire	Venomous plant/animal	Accidental poisoning	Intentional self-harm	Assault	Other/unspecified external causes	Cause of death unknown
Acute abdomen	**4**			3														
Renal failure	1	**2**		1														
Epilepsy			**13**	1														
Other/unspecified non-communicable diseases	1			**31**							1							1
Abortion-related death					**4**													
Pregnancy-induced hypertension						**2**												
Obstructed labour							**1**											
Anaemia of pregnancy																		
Other/unspecified maternal causes									**2**									
Road traffic collision										**16**								
Accidental drowning											**3**							
Exposure to smoke/fire												**1**						
Venomous plant/animal													**1**					
Accidental poisoning														**2**				
Intentional self-harm				1											**4**			
Assault																**22**	1	
Other/unspecified external causes																	**8**	1
Cause of death unknown				1		1											1	**1**

Appendix 8: Cause groups assigned to deaths on which the physicians evidently agree about cause but which have ICD-10 codes indicating different cause groups. Descriptions as provided in dataset. OU=Other/unspecified

Review 1			Review 2				
Description	*ICD-10 code*	*Cause group indicated by WHO VA standards*	*Description*	*ICD-10 code*	*Cause group indicated by WHO VA standards*	**Cause group assigned to death**	**Rationale**
Bacterial food poison	A05.9	Diarrhoeal diseases	Food poisoning	T62.9	OU external causes	OU external causes	Food poisoning is the common factor and is an external cause.
Acute alcohol intoxication	F10	OU non-communicable diseases	Alcohol intoxication	T51.9	OU external causes	OU external causes	T51.9 is akin to an acute poisoning, and the specification 'acute' in Review 1 suggests this is a non-chronic alcohol problem.
Alcoholic encephalo-pathy	F10.5	OU non-communicable diseases	Alcohol intoxication	T51.9	OU external causes	OU non-communicable diseases	'Encephalopathy' implies something more chronic than an acute poisoning.
Hypertension	I11	OU cardiac diseases	Hypertensio with stroke	I64	OU cardiac diseases	OU cardiac diseases	Agreement is evident on this being hypertension, so it is agreed to be due to cardiac diseases.
Hypertension	I67	Stroke	Hypertension	I10	OU cardiac diseases	OU cardiac diseases	Agreement is evident on this being hypertension, so it is agreed to be due to cardiac diseases.
Pueperial sepsis	O85	Pregnancy-related sepsis	Pueperal complication	O90	OU maternal causes	OU maternal causes	O90 contains many non-sepsis conditions, so agreement is limited to the broader cause group.
Burn injury	T31.9	OU external causes	Burning while in a building	X09	Exposure to smoke/fire	Exposure to smoke/fire	Chosen cause group is more specific.
Poisoned	T51.9	OU external causes	Poisoned	X49	Accidental poisoning	Accidental poisoning	Chosen cause group is more specific.
Venom snake bite	T63.0	OU external causes	Snake bite	X20	Venomous plant/animal	Venomous plant/animal	Chosen cause group is more specific.
Alcohol intoxication	T51.9	OU external causes	Acute alcohol intoxication	F10	OU non-communi-cable diseases	OU external causes	T51.9 is akin to an acute poisoning, and the specification 'acute' in Review 2 suggests this is a non-chronic alcohol problem.

Appendix 9: Descriptions and ICD-10 codes assigned by reviewing physicians, and cause groups assigned to reviews, for 134 deaths where I assigned the respective physician reviews to discordant cause groups. Descriptions are as provided, I have not edited them. OU = Other/unspecified; NCD = non-communicable diseases

First physician review			Second physician review		
Description (as provided)	*ICD-10 code*	*Cause group assigned*	*Description (as provided)*	*ICD-10 code*	*Cause group assigned*
15 deaths in which one review was assigned to the cause group "HIV/AIDS-related death" and the other indicated a condition defining stage 3/4 HIV disease					
Pulmonary tuberculosis	A169	Pulmonary tuberculosis	AIDS	B24	HIV/AIDS-related death
Pneumonia	J189	ARI/pneumonia	HIV	B24	HIV/AIDS-related death
Chronic lung infection	J188	ARI/pneumonia	HIV AIDS	B20	HIV/AIDS-related death
HIV disease	B20	HIV/AIDS-related death	Tuberculosis	A169	Pulmonary tuberculosis
HIV disease	B22	HIV/AIDS-related death	Encephalopathy	G934	OU NCD
Cutaneous abcess of the trunk	L02	OU infectious diseases	HIV resulting into bacteria infection	B207	HIV/AIDS-related death
Pyomyositis	M600	OU infectious diseases	HIV disease with multiple infection	B207	HIV/AIDS-related death
HIV resulting into neoplasia	B210	HIV/AIDS-related death	Cellulitis	L039	OU infectious diseases
Pulmonary tuberculosis	A169	Pulmonary tuberculosis	HIV disease resulting into tuberculosis	B200	HIV/AIDS-related death
HIV disease	B24	HIV/AIDS-related death	Pulmonary tuberculosis	A169	Pulmonary tuberculosis
HIV resulting into pulmonary TB	B200	HIV/AIDS-related death	Tuberculosis lymphoadenopathy	A18	OU infectious diseases
Multiple abscesses	L02	OU infectious diseases	HIV disease with unspecified infection	B209	HIV/AIDS-related death
Pulmonary tuberclulsis	A169	Pulmonary tuberculosis	HIV disease	B24	HIV/AIDS-related death
Pulmonary tuberculosis	A16	Pulmonary tuberculosis	HIV resulting tuberculosis	B20	HIV/AIDS-related death
Multiple abscess	L02	OU infectious diseases	HIV disease with unspecified in fection	B209	HIV/AIDS-related death
35 deaths in which the cause groups assigned to the two physician reviews were different and one mentioned HIV					
HIV disease resulting into encephalopathy	B22	HIV/AIDS-related death	Self hanging	X70	Intentional self-harm

AIDS	B24	HIV/AIDS-related death	Liver disease	K769	OU NCD
HIV disease with mycobacterial disease	B200	HIV/AIDS-related death	Heart disease	I519	OU cardiac diseases
HIV disease with mycosis	B205	HIV/AIDS-related death	Malignant neoplasia of the cervix	C539	Reproductive neoplasms
HIV disease with unspecified infection or par	B209	HIV/AIDS-related death	Secondary malignant neoplasia of the breast	C798	OU neoplasms
HIV disease resulting into	B205	HIV/AIDS-related death	Obstructive lung disease	J449	Chronic obstructive pulmonary disease
HIV disease with unspecified infection	B209	HIV/AIDS-related death	Genital prolapse	N819	OU NCD
HIV disease	B20	HIV/AIDS-related death	Intra abdomina malignancy	C76	OU neoplasms
HIV disease with unspecified infection	B209	HIV/AIDS-related death	Alcoholic liver disease	K703	OU NCD
HIV disease resulting into mycobacterial dise	B20	HIV/AIDS-related death	Malignacy neoplasia of urinary bladder	C679	OU neoplasms
AIDS	B24	HIV/AIDS-related death		W20	OU external causes
HIV disese	B24	HIV/AIDS-related death	Liver diseses	K769	OU NCD
HIV disease with encephalopathy	B22	HIV/AIDS-related death	Undermined	R97	Cause of death unknown
Tentative slim disease	B222	HIV/AIDS-related death	Tentative chronic lung disease	J988	OU infectious diseases
Angina pectoria with ischaemic heart disease	I259	Acute cardiac disease	HIV disease	B24	HIV/AIDS-related death
Cholera	A009	Diarrhoeal diseases	AIDS	B24	HIV/AIDS-related death
Severe malaria	B54	Malaria	HIV disease	B24	HIV/AIDS-related death
Viral hepatitis unspecified	B199	OU infectious diseases	HIV disease	B24	HIV/AIDS-related death
Renal disease	N159	OU infectious diseases	HIV disease with unspecified infection	B209	HIV/AIDS-related death
Gonodotrophoblastic disease (GTDS)	O019	OU maternal cause	HIV disease with mycobacterial disease	B20	HIV/AIDS-related death
Chronic respiratory	J989	OU NCD	HIV with mycobacterial disease	B200	HIV/AIDS-related death
Liver cirrhosis unspecified	K746	OU NCD	HIV resulting into TB	B200	HIV/AIDS-related death
Liver disease	K769	OU NCD	HIV disease with multiple infection	B207	HIV/AIDS-related death

Liver cirrhosis	K746	OU NCD	HIV resulting into multiple infection	B227	HIV/AIDS-related death
Hepatocellular disease	K769	OU NCD	AIDS	B20	HIV/AIDS-related death
Chest pain	R07	OU NCD	HIV disease	B24	HIV/AIDS-related death
Chronic liver disease	K769	OU NCD	HIV disease	B24	HIV/AIDS-related death
Liver disease	K76	OU NCD	HIV disease with unspecified infection	B20	HIV/AIDS-related death
Chronic	J989	OU NCD	HIV disease with unspecified infection	B20	HIV/AIDS-related death
Chronic liver disease	K769	OU NCD	HIV disese	B24	HIV/AIDS-related death
Malignancy neoplasia involving eye	C69	OU neoplasms	HIV disease	B24	HIV/AIDS-related death
Malignant melanonma	C43	OU neoplasms	HIV disease	B21	HIV/AIDS-related death
Brain tumor	C71	OU neoplasms	HIV disease with encephalopathy	B24	HIV/AIDS-related death
Ca cenix	C53	Reproductive neoplasms	HIV with unspecified infection	B209	HIV/AIDS-related death
Neoplasia of the utlrus	C559	Reproductive neoplasms	HIV with unspecified infector	B238	HIV/AIDS-related death
84 deaths in which the cause groups assigned to the two physician reviews were different and neither mentioned HIV					
Bronchio usthma	J45	Asthma	chrunic obstructive lung disease	J44	Chronic obstructive pulmonary disease
Intra-abdominal malignancy (viscern)	C762	OU neoplasms	Stomach malignancy	C169	Digestive neoplasms
Grand mal epilepsy	G406	Epilepsy	Brain tumour (benign)	D332	OU neoplasms
Senility	R54	OU NCD	Cerebral malaria	B50	Malaria
Pneumonia	J189	ARI/pneumonia	Heart failure	I50	OU cardiac diseases
Rectal prolapse	K623	OU NCD	Intestinal worms	B839	OU infectious diseases
Acute myocardial infarction unspecified	I219	Acute cardiac disease	Heart disease	I519	OU cardiac diseases
Cryptococcal meningitis	B451	OU infectious diseases	Unexplain headche	R51	OU NCD
Chronic liver disease	K769	OU NCD	Tuberculosis	A169	Pulmonary tuberculosis
Unexplained intra-abdominal mass	R19	OU NCD	Intra-abdominal mass	C772	OU neoplasms
TB of the spine	A18	OU infectious diseases	Malignacy (viscern)	C496	OU neoplasms

Urinary schistosomisis	B659	OU infectious diseases	Urinary bladder	C679	OU neoplasms
Chronic hepatitis	K739	OU NCD	Hamolytic anemia of unknown cause	D539	Severe anaemia
Heart disease	I519	OU cardiac diseases	Renal disease	N289	OU NCD
Stomach maligneney	C169	Digestive neoplasms	Schistosomiasis	B659	OU infectious diseases
Abd malignany unspecific	C762	OU neoplasms	Abdomenal tuberculosis	K93	OU infectious diseases
Open wound of head part unspecified	S01	OU external causes	Undetermined	R97	Cause of death unknown
Hypertension	I10	OU cardiac diseases	Hypertension	I21	Acute cardiac disease
Liver disease	K769	OU NCD	Hepatoma	C229	Digestive neoplasms
Malaria	B54	Malaria	Typhoid fever	A01	Diarrhoeal diseases
Dervical malignancy	C539	Reproductive neoplasms	Vesical - vaginal fistula	N820	OU NCD
Post-partum haemonhear	O72	Obstetric haemorrhage	Prolonged labour	O639	Obstructed labour
Anemia of unknown cause	D649	OU NCD	Chronic abdominal	R10	Acute abdomen
Chronic obstructure lung disease	J449	Chronic obstructive pulmonary disease	Pulmonary tuberculosis	A16	Pulmonary tuberculosis
Skin lesions	L089	OU infectious diseases	Steven johnson syndrome	L28	OU NCD
Pharyngeal ca	C329	Respiratory neoplasms	Oesophageal malignancy	C159	Digestive neoplasms
Acute abdomen	R10	Acute abdomen	Severe malaria	B54	Malaria
Acute abdomen	R10	Acute abdomen	Renal failure	N19	Renal failure
Puerperal infection	O86	OU maternal causes	Malaria	B54	Malaria
Spiral disease	G951	OU NCD	Tuberculosis of the spine	A18	OU infectious diseases
Acute myovardial infarction	I219	Acute cardiac disease	Heart disease	I519	OU cardiac diseases
Anaemia in pregnancy	O990	Anaemia of pregnancy	Severe anemia	D649	Severe anaemia
Chronic hepatitis	K74	OU NCD	Severe malaria	B54	Malaria
Peritonium lumour	D201	Digestive neoplasms	Liver disease	K769	OU NCD
Acute hepatitic	K759	OU NCD	Acute viral hepatitis with coma	B190	OU infectious diseases

Hemolytic anemia	D589	OU NCD	Jaundice with febrile illness	B159	OU infectious diseases
BPH	D291	Reproductive neoplasms	Benign prostale hypertrophy	N40	OU NCD
Alcoholic liver diseases	K709	OU NCD	Found dead	R98	Cause of death unknown
Malaria	B54	Malaria	Fever unspecified	R509	OU NCD
Leukemia unspecified	C959	OU neoplasms	Liver disease	K769	OU NCD
Hepatoma	C229	Digestive neoplasms	Gross splenomegaly	R16	OU NCD
Idiopathic convulsion	R568	OU NCD	Meningitis	G039	Meningitis/encephalitis
Pyrexia	R509	OU NCD	Abdominal pain	R100	Acute abdomen
Heart disease	I519	OU cardiac diseases	Chest pain	R074	OU NCD
Alcohol	T519	OU external causes	Raped to death under heary alcohol intake	Y05	Assault
Pulmonary tuberculosis	A169	Pulmonary tuberculosis	Heart disease	I519	OU cardiac diseases
Malaria	B54	Malaria	Meningistis	G030	Meningitis/encephalitis
Malaria	B54	Malaria	Acute abdomen	R10	Acute abdomen
Severe malaria	B54	Malaria	Chronic hepatitis	K73	OU NCD
Drowning in natural water	W70	Accidental drowning	Mental illness	F99	OU NCD
Intestinal obstruction	K566	OU NCD	Acute abdomen	R10	Acute abdomen
Undetermined	R97	Cause of death unknown	Headache	R51	OU NCD
Sickle cell disease	D571	Sickle cell with crisis	Splenomegally	R161	OU NCD
Viral hepatitis	B199	OU infectious diseases	Severe malaria	B54	Malaria
Chronic lung disease	J989	OU NCD	Pulmonary tuberculosis	A162	Pulmonary tuberculosis
Undetermined	R97	Cause of death unknown	Poisoned, unspecified	T519	OU external causes
Interstinal obstruction	K566	OU NCD	Viral hepatisis	B199	OU infectious diseases
Viral hepatitis	B199	OU infectious diseases	Chronic liver disease	K769	OU NCD
Pre-cordial pain	R07	OU NCD	Ischaemic heart disease	I249	Acute cardiac disease

Meningitis unspecified	G039	Meningitis/encephalitis	Severe malaria	B54	Malaria
Abdominal pain	R10	Acute abdomen	Pyogenic urethratis		OU infectious diseases
Psychosis	F29	OU NCD	Epilepsy	G40	Epilepsy
Urine retention	N11	OU NCD	Schistosomiasis	B659	OU infectious diseases
Meningitis	G03	Meningitis/encephalitis	Severe malaria	B54	Malaria
Mental illness	F99	OU NCD	Hanging herself	X70	Intentional self-harm
Brain tumour inspecified	D432	OU neoplasms	Meningitis	G039	Meningitis/encephalitis
Acute myocardial inferction	I21	Acute cardiac disease	Pneumonia	J18	ARI/pneumonia
Severe post-partum anemia	D62	OU NCD	Female interal organ tumour	D399	OU neoplasms
Intestinal obstruction	K56	OU NCD	Pyogenic urethralis		OU infectious diseases
Female genital organ	D399	Reproductive neoplasms	Intra abdominal malignancy	C767	OU neoplasms
Acute pyonephritis	N10	OU infectious diseases	Malaria	B54	Malaria
Neftrotic syndrome	N04	OU NCD	Renal tumour	C64	OU neoplasms
Cardine failure	I50	OU cardiac diseases	Pulmonary tuberculos	A16	Pulmonary tuberculosis
Meningitis	G039	Meningitis/encephalitis	Malaria	B54	Malaria
Typhoid fever	A011	Diarrhoeal diseases	Severe malaria	B54	Malaria
Nephrotic syndrome	N04	OU NCD	Vulvular heart diseases	I38	OU cardiac diseases
Tentative acute viral hepatitis	B199	OU infectious diseases	Tentantive: malaria	B54	Malaria
Pneumonia	J189	ARI/pneumonia	Pulmonary tuberculosis	A169	Pulmonary tuberculosis
Malignant brain neiplasia	C719	OU neoplasms	Pulmonary tuberculosis	A169	Pulmonary tuberculosis
Nephrotic syndrome	N04	OU NCD	Renal failure	N18	Renal failure
Neoplasis of the urinary system	C68	OU neoplasms	Unspecified renal disease	N19	Renal failure
Liver disease	K769	OU NCD	Viral hepatitis viral hepatitis	B199	OU infectious diseases
Abdominal pain	R104	Acute abdomen	Spleenic disease	D739	OU NCD

Sever eph ghestosis	O149	Pregnancy-induced hypertension	Undeterminaal	R97	Cause of death unknown

OU = Other/unspecified; NCD = non-communicable diseases

Appendix 10: InterVA Conditional probabilities of indicators where the cause of death is "HIV-related" (indicators with conditional probability ≥2%)

Conditional probability 0.8

Duration of final illness 3 weeks or more
History of HIV/AIDS
Not pregnant within 6 weeks of death

Conditional probability 0.5

Age 15-49 years
Male
Female
Wet season
Dry season
Cough of any kind
Diarrhoea lasting 4 weeks or more
Headache
Shingles/herpes zoster
Received vaccines as appropriate for age at death
Became very thin or wasted
Discharged from hospital very ill
Received treatment for illness that lead to death

Conditional probability 0.2

Age 50-64 years
Fever lasting 2 weeks or more
Cough lasting 3 weeks or more
Diarrhoea of any kind
Vomiting
Weight loss
Sores or white patches in the mouth or tongue
Lumps/swelling in the neck

Conditional probability 0.1

Age 1-4 years
Fever of any kind
Breathing problem of any kind
Fast breathing lasting 2 weeks or more
Breathlessness lasting 2 weeks or more
Ulcers, abscess, sores anywhere except feet
Skin rash lasting 1 week or more
Sunken eyes
Child was not growing normally or as expected

Conditional probability 0.05

Recent negative test for malaria
Productive cough with sputum
Fast breathing
Breathlessness
Any skin problems
Any skin rash (non-measles)
Unconscious for at least 24 hours before death
Both feet or ankles swollen
Hair colour changed to yellowish or reddish
Received (or needed) oral rehydration therapy
Received (or needed) IV drip
Received (or needed) treatment/food through nose
Received (or needed) IV or IM antibiotics
In the final illness, travelled to a hospital/health facility
Used motorised transport to get to the hospital
Problems during admission to the hospital
Problems in the way of being treated in the hospital
Problems getting medications or tests in the hospital
Takes more than 2 hours to get to the nearest hospital
In the final illness, doubts if medical care was needed
In the final illness, traditional medicine was used
At the time of death, used a phone to call for help
Total costs of care and treatment prohibitive

Conditional probability 0.02

Coughed blood
Diarrhoea lasting at least 2 weeks but <4 weeks
Severe abdominal pain
Severe abdominal pain lasting <2 weeks
Severe abdominal pain lasting 2 weeks or more
Abdominal mass lasting 2 weeks or more
Convulsions
Convulsions lasted less than 10 minutes
Convulsions lasted 10 minutes or more
Pale (thinning of blood) or pale palms/soles or nail beds

Appendix 11: Likelihood ratios for all symptoms in the VA dataset, Kisesa

VA item	Sens	Spec	LR
Had herpes zoster	14.5	98.0	7.36
Had medical diagnosis of HIV/AIDS	21.7	96.5	6.14
Had bloody diarrhoea	6.3	98.8	5.32
Had oral candidiasis	34.8	93.3	5.20
Had abnormal hair colouring	22.2	94.9	4.34
Had ulcers/abscesses/sores on body not on feet	21.3	94.5	3.86
Had sunken eyes	27.1	92.5	3.62
Had a rash	25.1	92.5	3.36
Had difficulty drinking	29.0	91.3	3.35
Had medical diagnosis of TB	6.3	98.0	3.19
Had skin lesions/ulcers	24.2	92.1	3.07
Had excessive night sweats	13.5	94.9	2.64
Had rapid breathing	9.2	96.5	2.59
Had a lump or lesion in the mouth	32.4	87.0	2.49
Had a productive cough	32.9	86.6	2.45
Had severe wasting	52.7	78.3	2.43
Had weight loss	70.5	68.1	2.21
Required oral rehydration during final illness	9.2	95.3	1.94
Had lumps/swelling in neck	15.9	91.7	1.93
Coughing with blood	6.8	96.5	1.91
Had localised lump/lesion not otherwise specified	20.3	89.4	1.91
Had anaemia/paleness	44.0	76.4	1.86
Medical diagnosis of heart disease	1.4	99.2	1.84
Both-sided paralysis	4.8	97.2	1.75
Vomiting	19.8	88.2	1.68
Swelling of ankles/legs	27.5	83.5	1.67
Haematuria	3.9	97.6	1.64
Illness lasted at least 3 weeks	62.8	59.4	1.55
Difficulty breathing	36.2	76.0	1.51
Blurred vision	25.6	82.7	1.48
Facial swelling	15.9	89.0	1.45
Delivered at home	19.3	86.6	1.44
Convulsions or fits	3.4	97.6	1.43
Convulsions or fits, duration unknown	3.4	97.6	1.43
Abdominal pain ≥ 2 weeks	18.4	87.0	1.41
Rigidity/lockjaw	4.3	96.9	1.38
Abdominal pain, duration unknown	13.0	90.6	1.38
Chest pain	28.0	79.5	1.37
Headache	52.7	61.0	1.35
Required antibiotic injection during final illness	21.7	83.5	1.31
Enlarged/swollen glands	8.2	93.7	1.30
Any abdominal pain	37.2	71.3	1.29
Breast lump or lesion	1.0	99.6	1.23
Stiff or painful neck, less than 1 week's duration	1.0	99.6	1.23
Foul smelling vaginal discharge	0.5	100.0	1.23
Medical diagnosis of haemoglobinopathy	0.5	100.0	1.23
Measles rash	0.5	100.0	1.23
Vomiting with blood	1.4	98.8	1.23
Lump or lesion in armpit	1.4	98.8	1.23

VA item	Sens	Spec	LR
Pregnant at time of death	1.0	99.2	1.23
Wheezing	1.0	99.2	1.23
Bleeding from mouth, nose or anus	0.5	99.6	1.23
Bite or sting by an animal	0.5	99.6	1.23
Abdominal swelling lasting ≥ 2 weeks	7.7	93.7	1.16
Known to drink alcohol	45.9	60.2	1.15
Swelling of both feet/ankles	8.2	92.9	1.10
Abdominal problem	38.6	65.0	1.10
Treatment for final illness from a health facility	56.5	47.6	1.08
Breathlessness lasting <2 weeks	3.4	96.9	1.07
Required IV drip during final illness	9.7	91.3	1.07
Ulcers/abscesses/sores on the feet	2.9	97.6	1.06
Stiff neck	11.1	89.8	1.05
Yellowness/jaundice	15.5	85.4	1.04
Stiff neck, duration unknown	9.2	91.3	1.02
Breathlessness lying flat	4.3	95.7	1.00
Rash lasting ≥ 1 week	100.0	0.0	1.00
Fever lasting < 2 weeks	100.0	0.0	1.00
Diarrhoea lsating 2-4 weeks	100.0	0.0	1.00
Diarrhoeaduration unknown	100.0	0.0	1.00
Cough lasting at least 2 weeks	100.0	0.0	1.00
Fever, duration unknown	100.0	0.0	1.00
Diarrhoea lasting <2 weeks	100.0	0.0	1.00
Cough, duration unknown	100.0	0.0	1.00
Diarrhoea lasting at lesat 4 weeks	100.0	0.0	1.00
Fever lasting at least 2 weeks	100.0	0.0	1.00
Rash lasting <1 week	100.0	0.0	1.00
Rash, duration unknown	100.0	0.0	1.00
Cough lasting <2 weeks	100.0	0.0	1.00
Death in the dry season	65.2	31.1	0.95
Known to smoke	6.3	93.3	0.94
Required blood transfusion during final illness	4.3	95.3	0.92
Breathlessness lasting at least 2 weeks	6.8	92.5	0.90
Any abdominal swelling	13.0	85.0	0.87
Death in the wet season	11.1	87.0	0.86
Coma came on suddenly	7.7	90.9	0.85
Stiff neck lasting at least 1 week	1.0	98.8	0.82
Abdominal mass lasting at least 2 weeks	2.9	96.5	0.82
Excessive urination	1.9	97.6	0.82
Any abnormality of urine	9.2	88.6	0.80
Urinary retention	4.3	94.5	0.79
Breathlessness on exertion	7.2	90.6	0.77
Coma lasting >24 hours	7.7	89.8	0.76
Lump or lesion in groin or genitals	3.9	94.9	0.76
Had operation within one month of death	1.4	98.0	0.74
Abdominal pain lasting <2 weeks	4.8	93.3	0.72
Any abdominal mass	3.9	94.5	0.70
Married at death	10.6	84.3	0.67
Had more than 4 previous pregnancies	4.8	92.5	0.65
Medical diagnosis of cancer	0.5	99.2	0.61
Medical diagnosis of hypertension	0.5	99.2	0.61
Medical diagnosis of malaria	0.5	99.2	0.61
Professional assistance at delivery	1.0	98.4	0.61

VA item	Sens	Spec	LR
Pregnancy was first pregnancy	1.0	98.4	0.61
Adbominal swelling lasting <2 weeks	1.4	97.2	0.53
Adbominal swelling, duration unknown	8.2	83.5	0.50
Bleeding between menstrual periods	1.0	98.0	0.49
Adbominal mass, duration unknown	1.0	98.0	0.49
Medical diagnosis of diabetes	0.5	98.8	0.41
One-sided paralysis	1.4	96.5	0.41
Required treatment/food through nose during final illness	1.0	97.6	0.41
Excessive water intake	1.0	97.6	0.41
Woman's normal vaginal bleeding had stopped naturally	1.4	96.1	0.37
Breathlessness, duration unknown	1.4	96.1	0.37
Pregnancy-related death, timing unknown	1.4	96.1	0.37
Final illness lasted <3 weeks	11.6	68.5	0.37
Surgery just before death	0.5	98.4	0.31
Normal vaginal delivery, no instruments	0.5	98.4	0.31
Delivery at health facility	0.5	98.0	0.25
Medical diagnosis of epilepsy	0.5	98.0	0.25
Death was very sudden or unexpected	1.4	82.3	0.08
Obvious recent injury	0.0	86.6	0.00
Intentionally injured by another person or people	0.0	94.9	0.00
Suggestion of homicide	0.0	95.3	0.00
Injured in some kind of violence or assault by another person	0.0	95.7	0.00
Road transport accident	0.0	95.7	0.00
Died < 6 weeks after normal length pregnancy	0.0	96.9	0.00
Major bleeding in late pregnancy/delivery	0.0	97.2	0.00
Transport accident	0.0	97.2	0.00
Suggestion of suicide	0.0	98.0	0.00
Major bleeding shortly before labour	0.0	98.0	0.00
Mother had excessive vaginal bleeding in pregnancy/postpartum period	0.0	98.0	0.00
Poisoning/bite/sting	0.0	98.4	0.00
Poisoning (not by an animal)	0.0	98.4	0.00
Labour prolonged >24 hours	0.0	98.8	0.00
Death within 24 hrs of pregnancy ending	0.0	98.8	0.00
Delivery by Caesarian section	0.0	99.2	0.00
Major bleeding during early pregnancy	0.0	99.2	0.00
Delivered a live baby within 6 weeks of death	0.0	99.2	0.00
Drowned	0.0	99.2	0.00
Fell recently	0.0	99.2	0.00
Vaginal bleeding after woman’s normal vaginal bleeding had stopped naturally	0.0	99.6	0.00
Major bleeding in first 6 months of pregnancy	0.0	99.6	0.00
Major bleeding during labour, before delivering the baby	0.0	99.6	0.00
Within 6 weeks of death, woman's pregnancy had ended in a spontaneous/induced abortion before foetus was viable	0.0	99.6	0.00
Major bleeding after delivering baby	0.0	99.6	0.00
Burnt by heat, steam or fire	0.0	99.6	0.00
Placenta remained inside	0.0	99.6	0.00

Appendix 12: Likelihood ratios for all symptoms in the VA dataset, Manicaland

VA item	Sens	Spec	LR
Had herpes zoster	22.3	95.4	4.80
Had severe wasting	11.3	96.7	3.41
Had a rash	30.2	89.4	2.85
Rash lasting >= 1 week	22.6	92.1	2.85
Coughing with blood	11.5	96.0	2.89
Had a productive cough	54.2	78.1	2.48
Had abnormal hair colouring	34.1	86.1	2.45
Had oral candidiasis	39.9	83.4	2.41
Rash, duration unknown	6.0	98.0	3.02
Had skin lesions/ulcers	27.6	88.1	2.31
Fever lasting at least 2 weeks	48.2	76.8	2.08
Had difficulty drinking	45.1	78.1	2.06
Cough lasting at least 2 weeks	62.4	68.9	2.00
Had medical diagnosis of HIV/AIDS	17.7	91.4	2.05
Had excessive night sweats	57.8	68.9	1.86
Diarrhoea lasting at lesat 4 weeks	34.1	81.5	1.84
Had weight loss	82.0	51.0	1.67
Breathlessness lasting at least 2 weeks	7.6	96.0	1.91
Had sunken eyes	40.6	75.5	1.66
Breathlessness on exertion	20.3	88.1	1.70
Breathlessness, duration unknown	1.1	100.0	
Vomiting	56.0	65.6	1.63
Illness lasted at least 3 weeks	90.5	40.4	1.52
Vomiting with blood	5.8	96.7	1.76
Had anaemia/paleness	60.1	59.6	1.49
Urinary retention	6.7	96.0	1.69
Headache	64.7	55.6	1.46
Diarrhoea lsating 2-4 weeks	7.2	95.4	1.56
Excessive water intake	53.0	61.6	1.38
Chest pain	13.4	90.7	1.45
Diarrhoeaduration unknown	2.7	98.7	2.00
Breathlessness lying flat	38.7	71.5	1.36
Had localised lump/lesion not otherwise specified	24.0	82.1	1.34
Swelling of both feet/ankles	23.1	82.8	1.34
Excessive urination	5.1	96.7	1.55
Death in the dry season	22.8	82.8	1.32
Lump or lesion in groin or genitals	20.1	84.8	1.32
Wheezing	11.7	91.4	1.35
Abdominal problem	66.4	47.7	1.27
Difficulty breathing	56.4	55.6	1.27
Abdominal pain, duration unknown	66.3	47.7	1.27
Any abdominal pain	66.3	47.7	1.27
Treatment for final illness from a health facility	70.8	43.7	1.26
Had bloody diarrhoea	10.6	92.1	1.33
Any abnormality of urine	14.7	88.7	1.30

VA item	Sens	Spec	LR
Coma lasting >24 hours	31.4	74.8	1.25
Rash lasting <1 week	1.6	99.3	2.40
Fever, duration unknown	6.4	95.4	1.37
Enlarged/swollen glands	3.9	97.4	1.47
Had lumps/swelling in neck	3.9	97.4	1.47
Stiff neck	11.5	90.7	1.24
Known to drink alcohol	45.4	58.9	1.11
Diarrhoea lasting <2 weeks	14.8	86.8	1.12
Rigidity/lockjaw	97.2	4.6	1.02
Swelling of ankles/legs	36.6	64.2	1.02
Stiff neck, duration unknown	14.7	85.4	1.01
Convulsions/fits, duration unknown	14.7	85.4	1.01
Convulsions or fits	14.7	85.4	1.01
Death in the wet season	74.9	21.9	0.96
Cough lasting <2 weeks	5.5	94.7	1.03
Blurred vision	5.3	94.7	1.00
Yellowness/jaundice	8.8	90.7	0.95
Fever lasting < 2 weeks	17.8	79.5	0.87
Both-sided paralysis	7.2	92.1	0.91
Cough, duration unknown	2.3	97.4	0.87
Any abdominal swelling	0.9	99.3	1.33
Adbominal swelling, duration unknown	0.9	99.3	1.33
Married at death	6.9	90.1	0.69
One-sided paralysis	3.0	96.0	0.76
Breathlessness lasting <2 weeks	3.0	95.4	0.65
Haematuria	1.4	97.4	0.53
Pregnant at time of death	1.8	96.0	0.44
Facial swelling	2.1	94.7	0.40
Final illness lasted <3 weeks	8.0	63.6	0.22
Blood pressure raised in pregnancy	0.2	98.7	0.13
Death very sudden or unexpected	0.2	91.4	0.02
Suggestion of homicide	0.0	94.7	0.00
Intentionally injured by other person	0.0	95.4	0.00
Transport accident	0.0	95.4	0.00
Poisoning/bite/sting	0.0	96.7	0.00
Drowned	0.0	99.3	0.00
Pregnancy was first pregnancy	0.0	98.7	0.00
Major bleeding during labour, before delivering the baby	0.0	97.4	0.00
Required oral rehydration in final illness	0.0	98.7	0.00
Poisoning (not by an animal)	0.0	96.7	0.00
Fell recently	0.0	99.3	0.00
Road transport accident	0.0	95.4	0.00
Burnt by heat, steam or fire	0.0	99.3	0.00
Suggestion of suicide	0.0	98.0	0.00
Obvious recent injury	0.0	85.4	0.00