InSilicoVA - An Algorithm for Automated COD Classification using Verbal Autopsy Data

Samuel J. Clark

Department of Sociology, The Ohio State University

Lead the openVA Team

affiliated with:

MRC/Wits Rural Public Health and Health Transitions Research Unit (Agincourt), School of Public Health, Faculty of Health Sciences, University of the Witwatersrand

Institute for Population Research, The Ohio State University

Translational Data Analytics Institute, The Ohio State University

Department of Statistics, University of Washington

February 5, 2024

Office of Population Research, Princeton University

Overview

Preliminaries

Motivation: global burden of disease and cause of death

Verbal autopsy
- VA algorithms and InterVA
- InSilicoVA
- VA Algorithm Validation Study
- ALPHA Network Cause of Death Study
- Software, and users

Beyond InSilicoVA

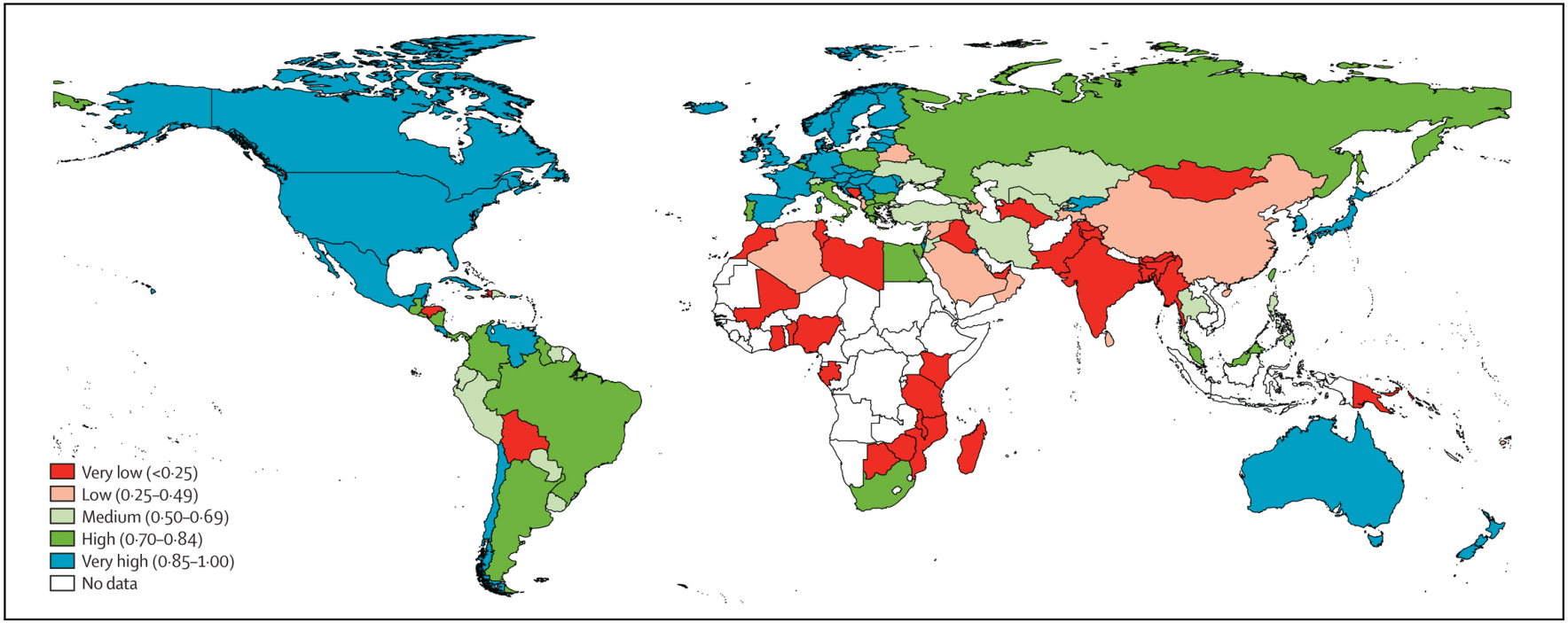
Acknowledgements

- ALPHA Network HDSS Sites
- Karonga HDSS site, Malawi
- Agincourt HDSS site, South Africa
- South African Medical Research Council, Burden of Disease Unit
- Bill and Melinda Gates Foundation
- National Institute for Child Health and Human Development
- Bloomberg Data for Health Initiative: Vital Strategies and CDC Foundation

Richard Li

Tyler McCormick

Global vital statistics performance index (Mikkelsen et al. [40])



Global Vital Statistics Performance Index

Missing

- about 60% of deaths - 36.5 million in 2023
- about 35% of births - 46.9 million in 2023

Burden of disease and cause of death determination

Burden of disease - BOD

- BOD is the distribution of deaths by cause
- BOD is fundamental population health metric
- Little *empirical knowledge* of the BOD for Africa and other resource-constrained settings

COD determination

- Registering and establishing a cause for all deaths are important population health priorities
- Traditional methods for COD determination are not feasible in resource-constrained settings
- The only realistic alternative is verbal autopsy - VA

Verbal autopsy – **VA**

Aim: Assign a cause to a death with VA – classify the death using an abbreviated VA cause list

Data:

1. Data from VA interview with knowledgeable caregiver of decedent
   - quantitative questions on signs, symptoms, diagnoses, durations, etc.
   - respondent's free-form narrative account of period leading up to death
2. Symptom-cause information (**SCI**) that describes the relationships between VA signs/symptoms and causes included in the VA cause list

Classification:

1. Physicians review VA data and assign causes: PCVA
2. Automated statistical/computational algorithms assign causes using VA data *and* SCI: CCVA

VA is an imperfect and frustrating approach

Advantages

- FEASIBLE compared to traditional COD determination: autopsy, medical review, etc.
- Comparatively cheap
- Comparatively tractable – logistics, skills, etc.
- With computer coding:
  - does not require advanced skills
  - produces reproducible cause assignments in a timely fashion
  - no physician opportunity costs
- Capable of providing highly useful COD and BOD information for public health assessment and planning

Disadvantages

- Much less accurate compared to traditional COD determination: autopsy, medical review, etc.
- Abbreviated cause list that does not easily mesh with full ICD cause lists, large catch-all causes
- Inherently low-information with many potential sources of error and bias: **classification is difficult**

Why I work on VA

- VA is a challenging approach that often produces underwhelming – but still useful – results
- Computer-coded VA is the only feasible solution for large-scale COD determination in resource-constrained settings without functioning vital statistics systems
- The remainder of this talk will be a technical discussion of the **InSilicoVA** automated cause coding algorithm for VA data developed by myself, **Richard Li**, and **Tyler McCormick**

VA Algorithms

VA cause-coding algorithms have three separable components

1. The VA data themselves
2. SCI that describes the relationship between VA symptoms and VA causes
3. The logic of the algorithm itself – mathematical, computational, statistical

The performance of each algorithm depends on both its logic and the SCI it uses

SCI can be swapped in/out and updated

This means that the performance of an algorithm can evolve and be adapted to a particular population

We will focus on algorithm logic and come back to SCI

Foundation for InSilicoVA: InterVA's origin, usage, and future

InSilicoVA originates with and fixes many weaknesses of InterVA, so we start with a discussion of InterVA

- InterVA (for 'Interpret VA') was developed by Peter Byass and his colleagues over many years, e.g. [8, 7, 17, 19, 10, 49, 21, 6, 9]
- InterVA is widely used and has been validated in a variety of ways – many substantive publications have relied on InterVA, e.g. [4, 44, 17, 3, 49, 11, 20, 21, 41, 12, 24, 50, 2, 37, 18, 46, 45, 6, 52, 48, 31, 47, 1, 42, 29, 16, 32, 5, 22, 43, 39]
- The last version is InterVA-5 [9]
- Peter Byass passed away during the pandemic; the openVA Team maintains openVA software that implements InterVA-5 but does not intend to produce any new updates

InterVA

InterVA is a computational algorithm designed

1. To distribute a single death across a number of causes with more weight on causes that are more consistent with the signs/symptoms associated with the death
2. Generate a population-level distribution of causes by summing up the fractions of a death associated with each cause across all individuals in the population

InterVA details 1

Notation

- J deaths: y_j
- N causes of death: c_n
- Death y_j with cause c_n: y_{jn}
- K sign/symptoms: $s_k \in \{0, 1\}$
- Vector of signs/symptoms for an individual death: $\vec{S}_j$
- Cause-specific mortality fractions (CSMF): f_n

InterVA details 2

Data

- For each death y_j, the VA interview produces a binary-valued vector of signs/symptoms

$$\vec{S_j} = \{s_{j1}, s_{j2}, \ldots s_{jK}\}$$

- Symptom-cause information in the form of 'probbase': a $K \times N$ matrix of conditional probabilities

$$\begin{bmatrix} \Pr(s_1|c_1) & \Pr(s_1|c_2) & \cdots & \Pr(s_1|c_N) \\ \Pr(s_2|c_1) & \Pr(s_2|c_2) & \cdots & \Pr(s_2|c_N) \\ \vdots & \vdots & \ddots & \vdots \\ \Pr(s_K|c_1) & \Pr(s_K|c_2) & \cdots & \Pr(s_K|c_N) \end{bmatrix}$$

InterVA details 3

Using Bayes' Rule we can derive an expression for what we want: the probability of a death with cause c_n, given that a specific set of symptoms were present or not present

$$\Pr(y_{jn}, \vec{S}_j) = \Pr(y_{jn}|\vec{S}_j)\Pr(\vec{S}_j) = \Pr(\vec{S}_j|y_{jn})\Pr(y_{jn})$$

$$\Pr(y_{jn}|\vec{S}_j) = \frac{\Pr(\vec{S}_j|y_{jn})\Pr(y_{jn})}{\Pr(\vec{S}_j)} \tag{1}$$

InterVA details 4

Assuming signs/symptoms are independent given cause, the probability of a specific sign/symptom vector is the product of the applicable probbase $\Pr(s_k|c_n)$ values when a sign/symptom exists and their complements when the sign/symptoms do not exist, so equation 1 can be expanded,

$$\Pr(y_{jn}|\vec{S}_j) = \frac{\Pr(y_{jn}) \prod_{k=1}^{K} \Pr(s_k|c_n)^{s_{jk}} \left[1 - \Pr(s_k|c_n)\right]^{(1-s_{jk})}}{\sum_{n'=1}^{N} \left(\Pr(y_{jn'}) \prod_{k=1}^{K} \Pr(s_k|c_{n'})^{s_{jk}} \left[1 - \Pr(s_k|c_{n'})\right]^{(1-s_{jk})} \right)}$$

Note

- Uses both presence and absence of symptom
- **NOT** how InterVA works!

InterVA details 5

To get InterVA

- keep pieces in blue that correspond to the presence of a sign/symptom
- drop pieces in red that correspond to absence of a sign/symptom

$$\Pr(y_{jn}|\vec{S}_j) = \frac{\Pr(y_{jn}) \prod_{k=1}^{K} \Pr(s_k|c_n)^{s_{jk}} [1 - \Pr(s_k|c_n)]^{(1-s_{jk})}}{\sum_{n'=1}^{N} \left(\Pr(y_{jn'}) \prod_{k=1}^{K} \Pr(s_k|c_{n'})^{s_{jk}} [1 - \Pr(s_k|c_{n'})]^{(1-s_{jk})} \right)}$$

Although not explained explicitly in the literature, it appears that the absent-symptom factors were dropped because most VA data available at the time did not have information on missing symptoms

InterVA details 6

With these deletions, InterVA calculates something different

$$\frac{\Pr(y_{jn}) \prod_{k=1}^{K} \Pr(s_k|c_n)^{s_{jk}}}{\sum_{n'=1}^{N} \left(\Pr(y_{jn'}) \prod_{k=1}^{K} \Pr(s_k|c_{n'})^{s_{jk}} \right)} = \Pr(y_{jn}|\vec{S}_j')$$

where $\vec{S}_j'$ is a vector that contains the subset of the elements of $\vec{S}_j$ whose values are all equal to 1

In general

$$\Pr(y_{jn}|\vec{S}_j) \neq \Pr(y_{jn}|\vec{S}_j')$$

except in the unique case when $\vec{S}_j$ and $\vec{S}_j'$ have the same number of elements and all the elements of both are equal to 1; in general not true

InterVA details 7

- We refer to the InterVA result quantities $\Pr(y_{jn}|\vec{S}_j')$ as 'cause-specific propensities'
- InterVA reports the three causes with the largest propensities that exceed 0.4
- If there are no propensities with magnitudes larger then 0.4, the cause is reported as 'indeterminate'
- 0.4 is an arbitrary threshold for which we cannot find any justification
- InterVA calculates CSMFs by summing cause-specific propensities across all deaths for each cause:

$$f_n = \sum_{j=1}^{J} \Pr(y_{jn}|\vec{S}_j')$$

InterVA: Implications

InterVA

- Does not compute the probability $\Pr(y_{jn}|\vec{S}_j)$ – **not** a probabilistic algorithm
- Computes probable cause conditional *only* on symptoms that were present, the propensity $\Pr(y_{jn}|\vec{S}'_j)$
  - **Critically**, this is fundamentally a different quantity for each death and therefore **not comparable** across deaths
  - Does not differentiate between observed absence of a sign/symptom and unobserved or missing
  - Sensitive to noise in data and obscures data issues related to symptoms that do not exist
- CSMFs are deterministic rescaling of individual-level cause determination
- Lots of other issues we don't have time for ...

InSilicoVA Setup

InSilicoVA (for on-chip/computational VA) is a statistical-computational algorithm that

1. At the individual level identifies a distribution of probabilities associated with each cause
2. At the population level identifies a distribution of counts of deaths for each cause
3. Links the two so that they are consistent with each other
4. Through these distributions quantifies uncertainty at both levels, again so that each is consistent with the other
5. Utilizes information from all deaths to help classify each death
6. Is comparatively robust to noise in data [15]

Uncertainty: if we are not sure about the cause for a death at the individual level, then in a proportionate way, we cannot be sure about that death's contribution to the cause-specific mortality fractions at the population level

InSilicoVA algorithm in words

Imagine a joint distribution of CSMF vectors and cause-specific probabilities for each death

Repeat the following steps many times, and for each loop, record the CSMFs and cause-specific probabilities for each death

1. For each death, use the current CSMFs (anything to start), knowledge of VA sign/symptom-cause relationships (SCI), and the VA data to calculate the probability of each cause
2. Again for each death, using those cause-specific probabilities, draw a cause from a multinomial distribution
3. Sum the deaths assigned to each cause to create a set of cause-specific death counts
4. Using the cause-specific death counts, draw a new set of CSMFs from a Dirichlet distribution

This procedure has two random steps that introduce uncertainty at the individual and population levels, and the cause-assignment step keeps the two levels linked together so that they are consistent with each other

The CSMFs aggregate information from all deaths and pass it along to each individual death in the next step

InSilicoVA algorithm – equations 1

Notation for InSilicoVA, similar to InterVA

- J deaths: y_j
- N causes of death: c_n
- Deaths, each with 1 assigned cause: $y_{jn} \in \{0, 1\}$; $\forall j : \sum_n y_{jn} = 1$ and $\sum_j \sum_n y_{jn} = J$
- $J \times N$ matrix of cause assignments for each death: $\mathbf{Y}$
- K sign/symptoms: $s_k \in \{0, 1\}$
- Vector of signs/symptoms for individual j: $\vec{S}_j$; elements s_{jk}
- For individual j, probability of cause n: ℓ_{jn}
- $J \times N$ matrix of cause-specific probabilities for each death: $\mathbf{L}$
- Cause-specific death count (CSDC): m_n
- Cause-specific mortality fraction (CSMF): f_n
- Vector of CSMFs: $\vec{F}$; $\sum_n f_n = 1$

InSilicoVA algorithm – equations 2

Following is a *minimal description* of InSilicoVA that illustrates the main ideas only – the full, published model has lots of refinements and nuances, see [38]

Data

- For each death y_j, VA interview produces a binary-valued vector of signs/symptoms
$$\vec{S_j} = \{s_{j1}, s_{j2}, \ldots s_{jk}\}$$
- Symptom-cause information as a $K \times N$ matrix of conditional probabilities $\Pr(s_k|c_n)$ – the same as InterVA

InSilicoVA algorithm – equations 3

Sketch of the model/algorithm

- We are interested in the joint distribution $(\vec{F}, \mathbf{L})$ – both unknown quantities
- So, we introduce a data augmentation procedure and use simulated cause assignments to stitch the two together
- Model CSMFs $\vec{F}$ conditional on cause assignments $\mathbf{Y}$

$$\vec{F}|\mathbf{Y} \sim \text{Dirichlet}(\vec{\alpha} + \vec{M}) \tag{2}$$

where the elements of $\vec{M}$ are

$$m_n = \sum_{j=1}^{J} y_{jn}$$

This ties the CSMFs $\vec{F}$ to the cause assignments y_{jn}

InSilicoVA algorithm – equations 4

- Model cause-assignments y_{jn} conditional on CSMFs $\vec{F}$ and the **data** – signs/symptoms $\vec{S}_j$

$$y_{jn}|\vec{S}_j, \vec{F} \sim \text{Multinomial}(1, \vec{L}_j) \tag{3}$$

The n components of $\vec{L}_j$ are

$$\begin{aligned}
\ell_{jn} &= \Pr(y_{jn}|\vec{S}_j, \vec{F}) \\
&= \frac{\Pr(\vec{S}_j|y_{jn}, \vec{F}) \Pr(y_{jn}|\vec{F})}{\Pr(\vec{S}_j|\vec{F})} \\
&\propto \Pr(\vec{S}_j|y_{jn}, \vec{F}) \Pr(y_{jn}|\vec{F})
\end{aligned}$$

Assume signs/symptoms are independent given cause and therefore independent of $\vec{F}$

$$\ell_{jn} \propto \Pr(\vec{S}_j|y_{jn}) \Pr(y_{jn}|\vec{F})$$

InSilicoVA algorithm – equations 5

- The cause-specific mortality fraction for cause n f_n is $\Pr(y_{jn}|\vec{F})$ and again, assuming signs/symptoms are independent given cause,

$$\Pr(\vec{S_j}|y_{jn}) = \prod_{k=1}^{K} \Pr(s_k|c_n)^{s_{jk}} \left[1 - \Pr(s_k|c_n)\right]^{(1-s_{jk})}$$

So,

$$\ell_{jn} \propto f_n \prod_{k=1}^{K} \Pr(s_k|c_n)^{s_{jk}} \left[1 - \Pr(s_k|c_n)\right]^{(1-s_{jk})}$$

This ties the cause assignments y_{jn} to the CSMFs $\vec{F}$

InSilicoVA algorithm – computation

Computation

- Take sample from joint distribution $(\vec{F}, \mathbf{L})$ using a Gibbs sampler
- The model defines the conditional distributions we need
- Initialize the CSMF vector to a reasonable set of values and execute the following 2 steps many times
  1. Use equation 3 to draw a cause for each death
  2. Use those cause assignments in equation 2 to draw a new CSMF vector
- After both $\vec{F}$ and $\mathbf{L}$ have settled into a stationary distribution, record the values of both in a (large) number of steps
- This set of values approximates the joint distribution $(\vec{F}, \mathbf{L})$
- Summarize the margins of the sample as necessary to produce desired outputs, usually distributions and a measure of their central tendencies for
  - elements of $\vec{F}$
  - for each death, the cause-specific elements of $\vec{\ell}_j$

InSilicoVA summary

InSilicoVA advances and advantages

- Provides mutually consistent estimates of individual-level probabilities of dying and population-level CSMFs
- Both are reported as distributions so there is a metric of uncertainty for both
- Builds on InterVA and utilizes the same SCI – *immediately usable*
- Provides information on all causes for all deaths, no 'indeterminate' cause - wide, flat distributions instead
- Uses information from all deaths to strengthen each individual death classification – equation 2 incorporates information from all deaths into the CSMF that then informs the cause assignments for each death
- Compared to InterVA, much more robust to noise in data (reporting errors) [15]
- Computationally feasible – barely, we are developing faster versions of the software using C/C++

InSilicoVA main publication [38]

JOURNAL OF THE AMERICAN STATISTICAL ASSOCIATION
2016, VOL. 111, NO. 515, 1036–1049, Applications and Case Studies
http://dx.doi.org/10.1080/01621459.2016.1152191

Taylor & Francis
Taylor & Francis Group

Probabilistic Cause-of-Death Assignment Using Verbal Autopsies

Tyler H. McCormick[a,*], Zehang Richard Li[b,*], Clara Calvert[c], Amelia C. Crampin[c], Kathleen Kahn[d], and Samuel J. Clark[e,*]

[a]Department of Statistics and Sociology, University of Washington, Seattle, WA, USA; [b]Department of Statistics, University of Washington, Seattle, WA, USA; [c]London School of Hygiene and Tropical Medicine, London, UK; [d]MRC/Wits Rural Public Health and Health Transitions Unit, School of Public Health, University of the Witwatersrand, Johannesburg, South Africa; [e]Department of Sociology, the Ohio State University, MRC/Wits Rural Public Health and Health Transitions Unit, School of Public Health, University of the Witwatersrand, Johannesburg, South Africa, ALPHA Network, London School of Hygiene and Tropical Medicine, London, UK, and INDEPTH Network, Accra, Ghana

ABSTRACT

In regions without complete-coverage civil registration and vital statistics systems there is uncertainty about even the most basic demographic indicators. In such regions, the majority of deaths occur outside hospitals and are not recorded. Worldwide, fewer than one-third of deaths are assigned a cause, with the least information available from the most impoverished nations. In populations like this, verbal autopsy (VA) is a commonly used tool to assess cause of death and estimate cause-specific mortality rates and the distribution of deaths by cause. VA uses an interview with caregivers of the decedent to elicit data describing the signs and symptoms leading up to the death. This article develops a new statistical tool known as *InSilicoVA* to classify cause of death using information acquired through VA. InSilicoVA shares uncertainty between cause of death assignments for specific individuals and the distribution of deaths by cause across the population. Using side-by-side comparisons with both observed and simulated data, we demonstrate that InSilicoVA has distinct advantages compared to currently available methods. Supplementary materials for this article are available online.

ARTICLE HISTORY
Received November 2014
Revised December 2015

KEYWORDS
Bayesian methods; cause of death; Demography; Verbal autopsy; Vital records

Miscellaneous observations of work on InSilicoVA

- We replicated work on InterVA and Tariff 1.0/2.0 algorithms and implemented both in openVA
- Replication is **very** hard – literature is woefully incomplete in terms of describing methods
- Both algorithms were available as proprietary implementations with no source code – this made it even harder/impossible to really understand what was going on
- **Lots** of not-described data pre-processing, automated data cleaning, etc.
- We conduct ourselves in the opposite way, hence the **openVA Team**
- As a result of our activities, all VA algorithms are now available as open source software
- We are thoroughly committed to a transparent, open source approach to creating, disseminating, and supporting methods
- We are working on several new approaches, but none are mature yet, see [34, 30, 33, 36]

South Africa VA validation study

We recently validated InSilicoVA, InterVA, and Tariff 2.0 using a high quality VA study in South Africa

South Africa Medical Research Council conducted the South African National Cause-of-Death Validation study [26]

A component of the study created a national sample of deaths with verbal autopsy

- Fieldwork 2017 – 2019
- 2016 WHO Standard VA
- 5,387 respondents consented and VA completed
- Multiple physician cause coding
- Underlying cause determined by Iris [25]
- 4,535 VA deaths received an valid underlying cause
- HIV is a major cause of death: 22.8% of deaths

VA algorithm validation using South Africa VA validation study

Working with the South African team, the openVA Team

- Applied openVA algorithms InterVA-5, InSilicoVA, and IHME algorithm Tariff 2.0 to the validation deaths
- Compared causes assigned by algorithms to reference causes identified by physicians and Iris
- Calculated a variety of comparison metrics
- Published in December, 2023 [23]

Algorithm comparison 1

Algorithm	Individual-Level Agreement: Overall Agreement Top Cause (95% CI)	Overall Agreement Top 3 Causes (95% CI)	Kappa Top cause (95% CI)	Chance Corrected Concordance Top Cause	Population-Level: CSMF Accuracy	Spearman Rank correlation (95% CI)
Total Sample N = 4,534						
InterVA-5	48.2 (46.7 – 49.7)	70.9 (69.6 – 72.2)	0.43 (0.42 – 0.44)	0.39	0.81	0.64 (0.62 – 0.65)
InSilicoVA	51.6 (50.2 – 53.1)	73.8 (72.5 -75.1)	0.47 (0.46 – 0.48)	0.42	0.84	0.68 (0.67 – 0.70)
Tariff 2.0	51.2 (49.8 – 52.7)	*	0.46 (0.45 – 0.47)	0.38	0.82	0.66 (0.65 – 0.68)
Neonate (0-27 days) N = 82						
InterVA-5	78.5 (67.5 – 86.4)	78.0 (67.5 – 86.4)	-0.05 (-0.14 – 0.04)	0.13	0.90	0.02 (-0.02 – 0.24)
InSilicoVA	79.3 (68.9 – 87.4)	79.3 (68.9 – 87.4)	-0.04 (-0.14 – 0.05)	0.13	0.84	0.02 (-0.02 – 0.24)
Tariff 2.0	47.6 (36.4 – 58.9)	*	0.01 (-0.04 – 0.06)	0.06	0.83	-0.06 (-0.27 – 0.16)
Child (28 days – 11 years) N = 165						
InterVA-5	36.4 (29.0 – 44.2)	50.3 (42.4 – 58.2)	0.32 (0.28 – 0.36)	0.32	0.66	0.43 (0.30 – 0.55)
InSilicoVA	40.6 (33.0 – 48.5)	56.4 (48.4 – 64.1)	0.36 (0.32 – 0.40)	0.40	0.64	0.65 (0.55 – 0.73)
Tariff 2.0	28.5 (21.7 – 36.0)	*	0.24 (0.20 – 0.28)	0.20	0.66	0.50 (0.37 – 0.60)

InSilicoVA rankings in performance comparisons

Population Group	N	First	Second	Third
Total Sample	4,534	6	0	0
Neonate (0-27 days)	82	4	2	0
Child (28 days – 11 years)	165	5	0	1
Adult (12 – 49 years)	1,812	2	4	0
Elder (50+ years)	2,475	4	2	0
Male	2,400	3	3	0
Female	2,134	6	0	0
Died in Health Facility	2,591	4	2	0
Died Out of Health Facility	1,943	6	0	0
Total		40	13	1

Looks good. But, examination of results reveals that none of the algorithms performed exceptionally well and the magnitude of InSilicoVA's lead is often small

Still work to do

South Africa VA validation study [23]

GLOBAL HEALTH ACTION
2023, VOL. 16, 2285105
https://doi.org/10.1080/16549716.2023.2285105

Taylor & Francis
Taylor & Francis Group

RESEARCH ARTICLE

OPEN ACCESS Check for updates

Agreement between cause of death assignment by computer-coded verbal autopsy methods and physician coding of verbal autopsy interviews in South Africa

Pam Groenewald [a], Jason Thomas [b], Samuel J Clark [b], Diane Morof [c,d], Jané D. Joubert [a], Chodziwadziwa Kabudula [e], Zehang Li [f] and Debbie Bradshaw [a,g]

[a]Burden of Disease Research Unit, South African Medical Research Council, Cape Town, South Africa; [b]Department of Sociology, The Ohio State University, Columbus, Ohio, USA; [c]Division of Global HIV & TB, Centers for Disease Control and Prevention, Durban, South Africa; [d]United States Public Health Service Commissioned Corps, Rockville, Maryland, USA; [e]MRC/Wits Rural Public Health and Health Transitions Research Unit (Agincourt), University of Witwatersrand, Johannesburg, South Africa; [f]Department of Statistics, University of California Santa Cruz, Santa Cruz, California, USA; [g]Division of Public Health Medicine, School of Public Health, University of Cape Town, Cape Town, South Africa

ABSTRACT

Background: The South African national cause of death validation (NCODV 2017/18) project collected a national sample of verbal autopsies (VA) with cause of death (COD) assignment by physician-coded VA (PCVA) and computer-coded VA (CCVA).

Objective: The performance of three CCVA algorithms (InterVA-5, InSilicoVA and Tariff 2.0) in assigning a COD was compared with PCVA (reference standard).

Methods: Seven performance metrics assessed individual and population level agreement of COD assignment by age, sex and place of death subgroups. Positive predictive value (PPV), sensitivity, overall agreement, kappa, and chance corrected concordance (CCC) assessed individual level agreement. Cause-specific mortality fraction (CSMF) accuracy and Spearman's rank correlation assessed population level agreement.

Results: A total of 5386 VA records were analysed. PCVA and CCVAs all identified HIV/AIDS as the leading COD. CCVA PPV and sensitivity, based on confidence intervals, were comparable except for HIV/AIDS, TB, maternal, diabetes mellitus, other cancers, and some injuries. CCVAs performed well for identifying perinatal deaths, road traffic accidents, suicide and homicide but poorly for pneumonia, other infectious diseases and renal failure. Overall agreement between CCVAs and PCVA for the top single cause (48.2–51.6) indicated comparable weak agreement between methods. Overall agreement, for the top three causes showed moderate agreement for InterVA (70.9) and InSilicoVA (73.8). Agreement based on kappa (–0.05–0.49)and CCC (0.06–0.43) was weak to none for all algorithms and groups. CCVAs had moderate to strong agreement for CSMF accuracy, with InterVA-5 highest for neonates (0.90), Tariff 2.0 highest for adults (0.89) and males (0.84), and InSilicoVA highest for females (0.88), elders (0.83) and out-of-facility deaths (0.85). Rank correlation indicated moderate agreement for adults (0.75–0.79).

Conclusions: Whilst CCVAs identified HIV/AIDS as the leading COD, consistent with PCVA, there is scope for improving the algorithms for use in South Africa.

ARTICLE HISTORY
Received 14 September 2023
Accepted 14 November 2023

RESPONSIBLE EDITOR
Stig Wall

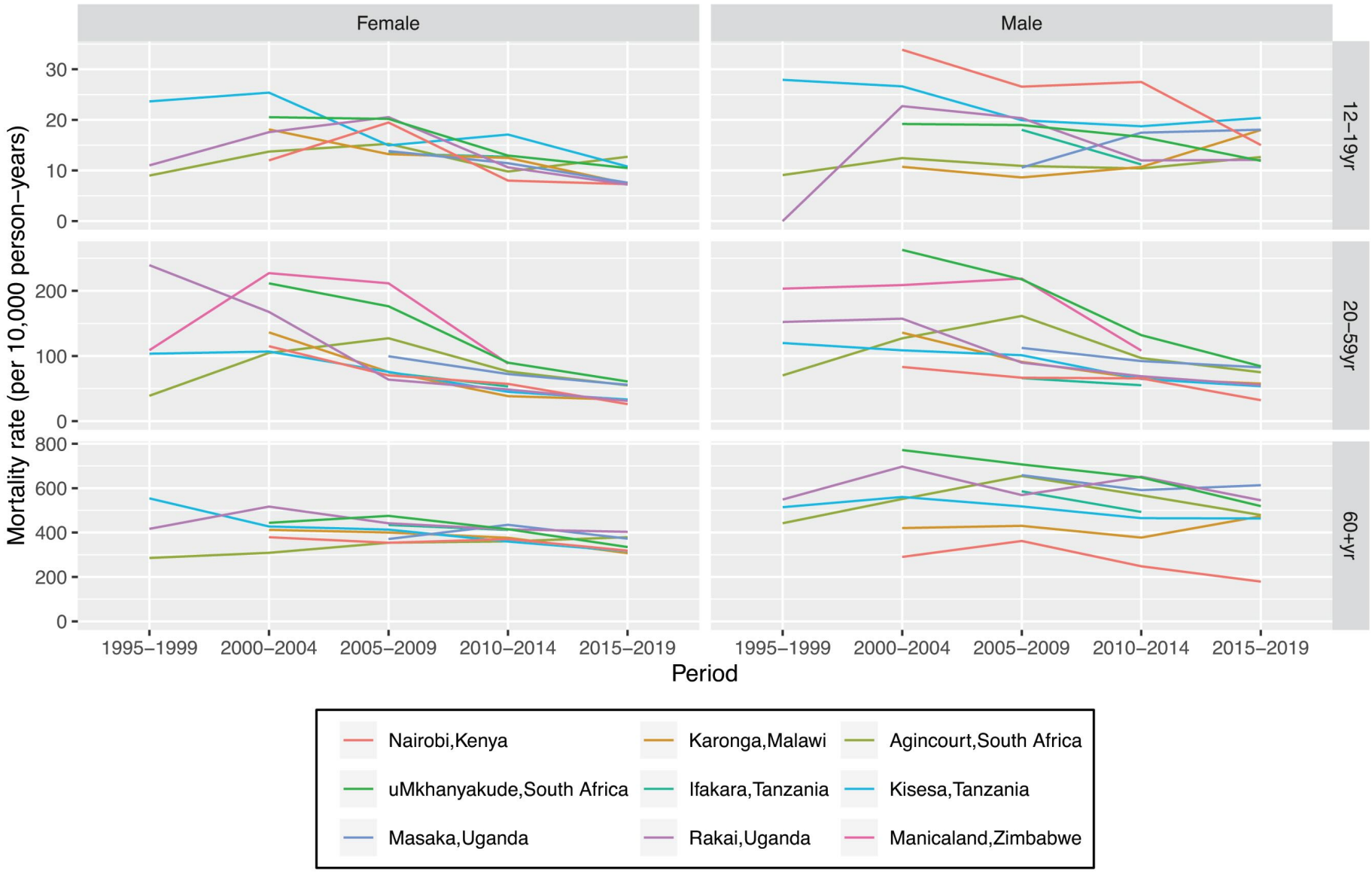
KEYWORDS
Physician-coded; computer-coded; verbal autopsy; agreement

ALPHA Network HDSS cause-specific mortality study

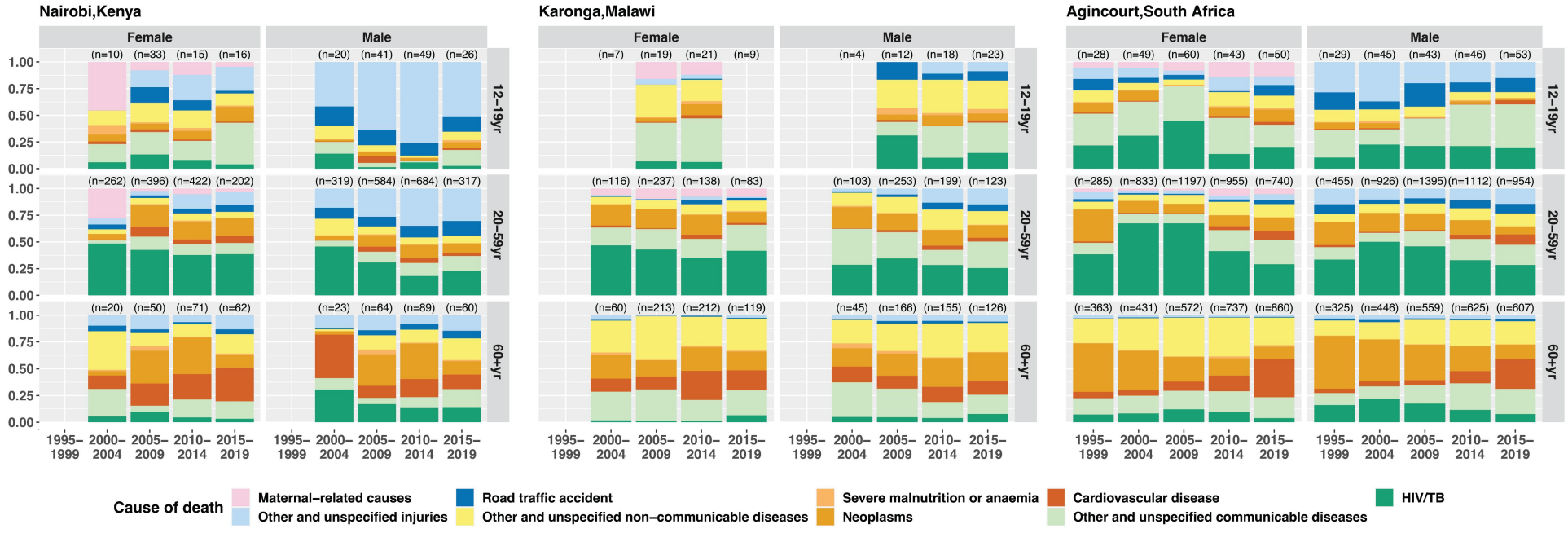
Work with Clara Calvert, Yue Chu, Milly Marston and ALPHA Network HDSS sites [13]

- Harmonized all exposure, death, and VA data from 9 health and demographic surveillance system sites (HDSS) in East and Southern Africa – *all are high HIV prevalence populations*
- Apply InSilicoVA to ascertain cause of death from VA
- Calculate trends in all-cause and cause-specific mortality
- Manuscript in review; hopefully out soon

ALPHA Network all-cause mortality trends



ALPHA Network cause-specific mortality fractions 1



openVA Suite

The openVA Team has developed and supports a range of software for VA, including InSilicoVA

- openVA: https://cran.r-project.org/package=openVA
- InSilicoVA: https://cran.r-project.org/package=InSilicoVA
- interVA5: https://cran.r-project.org/package=InterVA5
- interVA4: https://cran.r-project.org/package=InterVA4
- Tariff 1: https://cran.r-project.org/package=Tariff
- CrossVA: https://cran.r-project.org/package=CrossVA
- pyCrossVA: https://pypi.org/project/pycrossva/0.92/
- openVA Pipeline: https://pypi.org/project/openva-pipeline/
- Python openVA – in final testing, release planned mid 2024
- Others: https://github.com/verbal-autopsy-software
- User-oriented description and tutorial – **The openVA Toolkit for Verbal Autopsies** [35]

The openVA Suite is the reference implementation of VA algorithms that support WHO VA standards and is used by a wide variety of researchers and CRVS organizations globally

openVA Toolkit [35]

The R Journal HOME ISSUES NEWS CURRENT SUBMIT

The openVA Toolkit for Verbal Autopsies

Abstract:
Verbal autopsy (VA) is a survey-based tool widely used to infer cause of death (COD) in regions without complete-coverage civil registration and vital statistics systems. In such settings, many deaths happen outside of medical facilities and are not officially documented by a medical professional. VA surveys, consisting of signs and symptoms reported by a person close to the decedent, are used to infer the COD for an individual, and to estimate and monitor the COD distribution in the population. Several classification algorithms have been developed and widely used to assign causes of death using VA data. However, the incompatibility between different idiosyncratic model implementations and required data structure makes it difficult to systematically apply and compare different methods. The openVA package provides the first standardized framework for analyzing VA data that is compatible with all openly available methods and data structure. It provides an open-source, R implementation of several most widely used VA methods. It supports different data input and output formats, and customizable information about the associations between causes and symptoms. The paper discusses the relevant algorithms, their implementations in R packages under the openVA suite, and demonstrates the pipeline of model fitting, summary, comparison, and visualization in the R environment.

AUTHORS
Zehang Richard Li
Jason Thomas
Eungang Choi
Tyler H. McCormick
Samuel J Clark

AFFILIATIONS
University of California, Santa Cruz
The Ohio State University
The Ohio State University
University of Washington
The Ohio State University

PUBLISHED
Feb. 24, 2023

CITATION
Li, et al., 2023

Issues that InSilicoVA doesn't solve

Limitations of InSilicoVA

- Assumes conditional independence of symptoms – lots of information lost
- Does not utilize free text account part of VA data – physicians place great weight on this, more information lost
- Does not recognize or account for different epidemiological domains

Additional challenge: SCI is not sufficiently informative

- Current SCI – $\Pr(s|c)$ – is inadequate, out of date, static, and related to limited epidemiological settings
- Two options to create SCI: elicite from physicians or calculate from reference deaths
- Physicians are comparatively easy and cheap but provide less nuance and essentially no information on dependence among symptoms
- Reference deaths are very hard to accumulate in sufficient numbers but potentially provide much more information, e.g. dependence among symptoms and domain-specific information

The key role of SCI

- We have demonstrated that with current algorithm logics, the SCI is at least, and often more, important than the algorithm logic in explaining the differences in algorithm performance [14]
- We have also shown that SCI dramatically affects algorithm performance and that, as expected, SCI is domain-specific [14, 38]
- Consequently, improving SCI is arguably more important than improving algorithm logic
- Two options for better SCI
  - Improved physician-elicited $\Pr(s|c)$, or
  - Large collection of reference deaths with VA and independent reference cause to infer/estimate/calculate new, more informative SCI

WHO 2022 VA and 'probbase' update – update to physician-elicited SCI

- The WHO VA Reference Group and Data for Health partners created an streamlined and strengthened, post-CV19 update to the WHO Standard VA
- The 2022 VA is much shorter, smoother, and has CV19 as a new cause
- This requires a big update to the algorithms and completely new SCI – the $\Pr(s|c)$ in the probbase
- Supported by the Data for Health Initiative through the CDC Foundation, I am currently leading a team of about 70 people, mostly physicians, to create a new physician-based SCI for the 2022 standard VA
- The openVA algorithm code has been updated and is ready to go
- Anticipate testing the new 2022-compatible algorithms in late 2024

Reference Death Archive

The openVA Team with many partners is currently creating a global reference death archive for VA

- Reference deaths from many sites around the world, many with reference deaths informed by pathology through minimally-invasive tissue sample (MITS – autopsy-light)
  - Cover wide variety of epidemiological domains and develop/test domain-adaptive algorithms
  - Updated through time
  - Include enough deaths to estimate dependencies among symptoms and include those in new algorithms
- Hosted at WHO in Geneva, globally available
- Many reference deaths from mortality surveillance units in Brazil who are conducting traditional autopsy and WHO 2022 VA
- Supported by Bill and Melinda Gates Foundation; right now mired in bureaucratic setup activities around the world!

Interview

- We have demonstrated that reporting error can dramatically reduce algorithm performance [15], InSilicoVA is more robust to this
- For this reason, the quality and consistency of the VA interview is critical
- Clarissa Surek-Clark – sociolinguist/translator/interpreter, Nicole Angotti – sociologist/demographer, and soon Brian Houle – sociologist/demographer are conducting qualitative studies of the VA interview, language usage, and translation/interpretation issues aiming to
  - improve the interview experience for respondent and interviewer – *whole topic for a different talk*
  - standardize the interview design and conduct of interviews
  - standardize the way languages are handled, processed, etc.
  - develop a standard protocol for narrative account elicitation
- The hope is to greatly improve the conduct of the interviews and the quality and consistency of the VA data they produce
- A key goal is to greatly improve the free-text narrative account so that it is respondent-friendly and maximally useful for machine-based text processing

Incorporating free-form text from account

Physicians place great weight on the narrative account when assigning causes to VA deaths

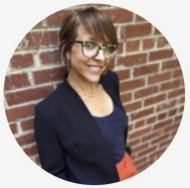
- InSilicoVA does not use any part of the account
- We are exploring various ways of producing consistent, useful information from the accounts using off-the-shelf NLP methods – not going very well
- Attempts have been made to classify VA deaths based solely on automated text processing of accounts – doesn't work well yet, many studies exaggerate performance by truncating/aggregating the cause list, e.g. [28, 51, 27]
- This effort is combined with work on the interview to ensure that elicitation of text-based accounts is both maximally meaningful and consistent with the needs of automated text processing methods
- Various approaches to incorporating text into algorithms are being considered; anticipate significant improvement in algorithm performance when text-based information is available

openVA Team

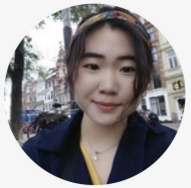
Research Team

Sam Clark

Principal Investigator



Nicole Angotti



Yoonyoung Choi

Collins Ochieng

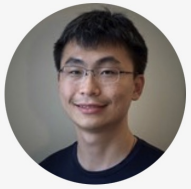
Isaac Lyatuu

Sherry Zhao

Yue Chu

Jason Thomas

Tyler McCormick



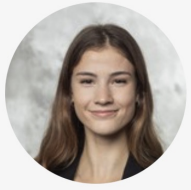
Zehang Richard Li

Clarissa Surek-Clark

Zhenke Wu

Emeritus Members

Eungang Peter Choi



Melina Raglin

Download slides: https://samclark.net/talks – Contact: sam@samclark.net

References I

[1] Raymond A. Aborigo, Pascale Allotey, Paulina Tindana, Daniel Azongo, and Cornelius Debpuur. Cultural imperatives and the ethics of verbal autopsies in rural Ghana. *Global Health Action*, 6:1–11, 2013.

[2] Evasius Bauni, Carolyne Ndila, George Mochamah, Gideon Nyutu, Lena Matata, Charles Ondieki, Barbara Mambo, Maureen Mutinda, Benjamin Tsofa, Eric Maitha, Anthony Etyang, and Thomas N. Williams. Validating physician-certified verbal autopsy and probabilistic modeling (InterVA) approaches to verbal autopsy interpretation using hospital causes of adult deaths. *Population Health MetricsS*, 9, AUG 5 2011.

[3] Jacqueline S. Bell, Moctar Ouedraogo, Rasmane Ganaba, Issiaka Sombie, Peter Byass, Rebecca F. Baggaley, Veronique Filippi, Ann E. Fitzmaurice, and Wendy J. Graham. The epidemiology of pregnancy outcomes in rural Burkina Faso. *Tropical Medicine & International Health*, 13(1):31–43, JUL 2008.

[4] P Byass, DL Huong, and HV Minh. A probabilistic approach to interpreting verbal autopsies: methodology and preliminary validation in Vietnam. *Scandinavian Journal of Public Health*, 31(62):32–37, 2003.

[5] Peter Byass. Usefulness of the Population Health Metrics Research Consortium gold standard verbal autopsy data for general verbal autopsy methods. *BMC Medicine*, 12, FEB 4 2014.

[6] Peter Byass, Daniel Chandramohan, Samuel J. Clark, Lucia D'Ambruoso, Edward Fottrell, Wendy J. Graham, Abraham J. Herbst, Abraham Hodgson, Sennen Hounton, Kathleen Kahn, Anand Krishnan, Jordana Leitao, Frank Odhiambo, Osman A. Sankoh, and Stephen M. Tollman. Strengthening standardised interpretation of verbal autopsy data: the new InterVA-4 tool. *Global Health Action*, 5:1–8, 2012.

[7] Peter Byass, Edward Fottrell, Dao Lan Huong, Yemane Berhane, Tumani Corrah, Kathleen Kahn, Lulu Muhe, et al. Refining a probabilistic model for interpreting verbal autopsy data. *Scandinavian journal of public health*, 34(1):26–31, 2006.

[8] Peter Byass, Dao Lan Huong, and Hoang Van Minh. A probabilistic approach to interpreting verbal autopsies: methodology and preliminary validation in vietnam. *Scandinavian Journal of Public Health*, 31(62 suppl):32–37, 2003.

[9] Peter Byass, Laith Hussain-Alkhateeb, Lucia D'Ambruoso, Samuel J. Clark, Justine Davies, Edward Fottrell, Jon Bird, Chodziwadziwa Kabudula, Stephen Tollman, Kathleen Kahn, Linus Schiöler, and Max Petzold. An Integrated Approach to Processing WHO-2016 Verbal Autopsy Data: the InterVA-5 Model. *BMC Medicine*, 17(102), 2019.

[10] Peter Byass, Kathleen Kahn, Edward Fottrell, Mark A Collinson, and Stephen M Tollman. Moving from data on deaths to public health policy in agincourt, south africa: approaches to analysing and understanding verbal autopsy findings. *PLoS medicine*, 7(8):1073, 2010.

[11] Peter Byass, Kathleen Kahn, Edward Fottrell, Mark A. Collinson, and Stephen M. Tollman. Moving from Data on Deaths to Public Health Policy in Agincourt, South Africa: Approaches to Analysing and Understanding Verbal Autopsy Findings. *PLoS Medicine*, 7(8), AUG 2010.

[12] Peter Byass, Kathleen Kahn, Edward Fottrell, Paul Mee, Mark A. Collinson, and Stephen M. Tollman. Using verbal autopsy to track epidemic dynamics: the case of HIV-related mortality in South Africa. *Population Health MetricsS*, 9, AUG 5 2011.

References II

[13] Yue Chu, Milly Marston, Albert Dube, Charles Festo, Eveline Geubbels, Simon Gregson, Kobus Herbst, Chodziwadziwa Kabudula, Kathleen Kahn, Tom Lutalo, Louisa Moorhouse, Robert Newton, Constance Nyamukapa, Rondad Makenga, Emma Slaymaker, Mark Urassa, Abdhalah Ziraba, Clara Calvert, and Samuel J. Clark. Temporal Changes in Cause of Death among Adolescents and Adults in Six Countries in Eastern and Southern Africa: a Multi-country Cohort Study using Verbal Autopsy Data. *(In Review)*, 2024.

[14] Samuel J. Clark, Zehang R Li, and Tyler H McCormick. Quantifying the Contributions of Training Data and Algorithm Logic to the Performance of Automated Cause-Assignment Algorithms for Verbal Autopsy. *arXiv Preprint*, (arXiv:1803.07141v2), 2018.

[15] Samuel J. Clark, Tyler McCormick, Zehang Li, and Jon Wakefield. InSilicoVA: A Method to Automate Cause of Death Assignment for Verbal Autopsy. *Center for Statistics and the Social Sciences (CSSS) Working Paper Series*, (133), 2013.

[16] Nikita Desai, Lukasz Aleksandrowicz, Pierre Miasnikof, Ying Lu, Jordana Leitao, Peter Byass, Stephen Tollman, Paul Mee, Dewan Alam, Suresh Kumar Rathi, Abhishek Singh, Rajesh Kumar, Faujdar Ram, and Prabhat Jha. 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 Medicine*, 12, FEB 4 2014.

[17] M Fantahun, E Fottrell, Y Berhane, S Wall, U Hogberg, and P Byass. Assessing a new approach to verbal autopsy interpretation in a rural Ethiopian community: the InterVA model. *Bulletin of the World Health Organization*, 84(3):204–210, MAR 2006.

[18] Cleusa P. Ferri, Daisy Acosta, Mariella Guerra, Yueqin Huang, Juan J. Llibre-Rodriguez, Aquiles Salas, Ana Luisa Sosa, Joseph D. Williams, Ciro Gaona, Zhaorui Liu, Lisseth Noriega-Fernandez, A. T. Jotheeswaran, and Martin J. Prince. Socioeconomic Factors and All Cause and Cause-Specific Mortality among Older People in Latin America, India, and China: A Population-Based Cohort Study. *PLoS Medicine*, 9(2), FEB 2012.

[19] Edward Fottrell, Peter Byass, Thomas W Ouedraogo, Cecile Tamini, Adjima Gbangou, Issiaka Sombié, Ulf Högberg, Karen H Witten, Sohinee Bhattacharya, Teklay Desta, et al. Revealing the burden of maternal mortality: a probabilistic model for determining pregnancy-related causes of death from verbal autopsies. *Popul Health Metr*, 5(1):10–1186, 2007.

[20] Edward Fottrell, Kathleen Kahn, Nawi Ng, Benn Sartorius, Dao Lan Huong, Hoang Van Minh, Mesganaw Fantahun, and Peter Byass. Mortality measurement in transition: proof of principle for standardised multi-country comparisons. *Tropical Medicine & International Health*, 15(10):1256–1265, OCT 2010.

[21] Edward Fottrell, Kathleen Kahn, Stephen Tollman, and Peter Byass. Probabilistic Methods for Verbal Autopsy Interpretation: InterVA Robustness in Relation to Variations in A Priori Probabilities. *PLoS ONE*, 6(11), NOV 3 2011.

[22] Judith R. Glynn, Clara Calvert, Alison Price, Menard Chihana, Lackson Kachiwanda, Sebastian Mboma, Basia Zaba, and Amelia C. Crampin. Measuring causes of adult mortality in rural northern Malawi over a decade of change. *Global Health Action*, 7, 2014.

[23] Pam Groenewald, Jason Thomas, Samuel J. Clark, Diane Morof, Jane Joubert, Chodziwadziwa W. Kabudula, Zehang Li, and Debbie Bradshaw on behalf of the South African National Cause-of-Death Validation Project Team. Agreement between Cause Of Death Assignment by Computer-coded Verbal Autopsy Methods and Physician Coding of Verbal Autopsy Interviews in South Africa. *Global Health Action*, 16(2285105):1–13, 2023.

References III

[24] Abraham J. Herbst, Tshepiso Mafojane, and Marie-Louise Newell. Verbal autopsy-based cause-specific mortality trends in rural KwaZulu-Natal, South Africa, 2000-2009. *Population Health MetricsS*, 9, AUG 5 2011.

[25] Iris Institute. *Iris automated coding system for causes of death*. http://https://www.bfarm.de/EN/Code-systems/Collaboration-and-projects/Iris-Institute/_node.html, accessed 2024-01-31 2024.

[26] Iris Institute. *South African National Cause-of-Death Validation*. https://www.samrc.ac.za/research-reports/south-african-national-cause-death-validation, accessed 2024-01-31 2024.

[27] Serena Jeblee, Mireille Gomes, and Graeme Hirst. Multi-task learning for interpretable cause-of-death classification using key phrase prediction. *BioNLP 2018 - SIGBioMed Workshop on Biomedical Natural Language Processing, Proceedings of the 17th BioNLP Workshop*, pages 12–17, 2018. ISBN: 9781948087339.

[28] Serena Jeblee, Mireille Gomes, Prabhat Jha, Frank Rudzicz, and Graeme Hirst. Automatically determining cause of death from verbal autopsy narratives. *BMC medical informatics and decision making*, 19(1):127, July 2019.

[29] Chifundo Kanjala, Denna Michael, Jim Todd, Emma Slaymaker, Clara Calvert, Raphael Isingo, Alison Wringe, Basia Zaba, and Mark Urassa. Using HIV-attributable mortality to assess the impact of antiretroviral therapy on adult mortality in rural Tanzania. *Global Health Action*, 7:1–8, 2014.

[30] Tsuyoshi Kunihama, Zehang Richard Li, Samuel J. Clark, and Tyler H McCormick. Bayesian Factor Models for Probabilistic Cause of Death Assessment with Verbal Autopsies. *Annals of Applied Statistics*, 14(1):241–256, 2020.

[31] JC Leitao, D Chandramohan, P Byass, R Jakob, K Bundhamcharoen, C Choprapowan, et al. Revising the WHO verbal autopsy instrument to facilitate routine cause-of-death monitoring. *Global Health Action*, 6(21518), 2013.

[32] Jordana Leitao, Nikita Desai, Lukasz Aleksandrowicz, Peter Byass, Pierre Miasnikof, Stephen Tollman, Dewan Alam, Ying Lu, Suresh Kumar Rathi, Abhishek Singh, Wilson Suraweera, Faujdar Ram, and Prabhat Jha. 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 Medicine*, 12, FEB 4 2014.

[33] Zehang (Richard) Li, Tyler H McCormick, and Samuel J. Clark. Bayesian Joint Spike-and-Slab Graphical Lasso. In *Proceedings of the 36th International Conference on Machine Learning*, volume 97 of *Proceedings of Machine Learning Research*, pages 3877–3885. International Conference on Machine Learning, Long Beach, California, USA, 6 2019.

[34] Zehang (Richard) Li, Tyler H McCormick, and Samuel J. Clark. Using Bayesian Latent Gaussian Graphical Models to Infer Symptom Associations in Verbal Autopsies. *Bayesian Analysis*, 15(3):781–807, 2020.

[35] Zehang Richard Li, Jason Thomas, Eungang Choi, Tyler H. McCormick, and Samuel J. Clark. The openVA Toolkit for Verbal Autopsies. *The R Journal*, 2023.

[36] Zehang Richard Li, Zhenke Wu, Irena Chen, and Samuel J. Clark. Bayesian Nested Latent Class Models for Cause-of-death Assignment using Verbal Autopsies across Multiple Domains. *Annals of Applied Statistics*, 2024 *in-press, preprint available*.

References IV

[37] Rafael Lozano, Michael K. Freeman, Spencer L. James, Benjamin Campbell, Alan D. Lopez, Abraham D. Flaxman, Christopher J. L. Murray, and PHMRC. Performance of InterVA for assigning causes of death to verbal autopsies: multisite validation study using clinical diagnostic gold standards. *Population Health MetricsS*, 9, AUG 5 2011.

[38] Tyler H. McCormick, Zehang (Richard) Li, Clara Calvert, Amelia C. Crampin, Kathleen Kahn, and Samuel J. Clark. Probabilistic Cause-of-Death Assignment using Verbal Autopsies. *Journal of the American Statistical Association*, 111(515):1036–1049, 2016.

[39] Yohannes Adama Melaku, Berhe Weldearegawi Sahle, Fisaha Haile Tesfay, Afework Mulugeta Bezabih, Alemseged Aregay, Semaw Ferede Abera, Loko Abreha, and Gordon Alexander Zello. Causes of Death among Adults in Northern Ethiopia: Evidence from Verbal Autopsy Data in Health and Demographic Surveillance System. *PLoS ONE*, 9(9), SEP 4 2014.

[40] Lene Mikkelsen, David E Phillips, Carla AbouZahr, Philip W Setel, Don De Savigny, Rafael Lozano, and Alan D Lopez. A global assessment of civil registration and vital statistics systems: monitoring data quality and progress. *The Lancet*, 386(10001):1395–1406, 2015.

[41] Christopher J. L. Murray, Alan D. Lopez, Robert Black, Ramesh Ahuja, Said Mohd Ali, et al. Population Health Metrics Research Consortium gold standard verbal autopsy validation study: design, implementation, and development of analysis datasets. *Population Health MetricsS*, 9, AUG 4 2011.

[42] Christopher J. L. Murray, Rafael Lozano, Abraham D. Flaxman, Peter Serina, David Phillips, Andrea Stewart, et al. Using verbal autopsy to measure causes of death: the comparative performance of existing methods. *BMC Medicine*, 12, JAN 9 2014.

[43] Carolyne Ndila, Evasius Bauni, Vysaul Nyirongo, George Mochamah, Alex Makazi, Patrick Kosgei, Gideon Nyutu, Alex Macharia, Sailoki Kapesa, Peter Byass, and Thomas N. Williams. Verbal autopsy as a tool for identifying children dying of sickle cell disease: a validation study conducted in Kilifi district, Kenya. *BMC Medicine*, 12, APR 22 2014.

[44] Samuel O. Oti and Catherine Kyobutungi. Verbal autopsy interpretation: a comparative analysis of the InterVA model versus physician review in determining causes of death in the Nairobi DSS. *Population Health MetricsS*, 8, 2010.

[45] Heribert Ramroth, Eva Lorenz, Johanna C. Rankin, Edward Fottrell, Maurice Ye, Florian Neuhann, Mark Ssennono, Ali Sie, Peter Byass, and Heiko Becher. Cause of death distribution with InterVA and physician coding in a rural area of Burkina Faso. *Tropical Medicine & International Health*, 17(7):904–913, JUL 2012.

[46] Johanna C. Rankin, Eva Lorenz, Florian Neuhann, Maurice Ye, Ali Sie, Heiko Becher, and Heribert Ramroth. 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. *Malaria Journal*, 11, FEB 21 2012.

[47] Sebsibe Tadesse. Validating the InterVA Model to Estimate the Burden of Mortality from Verbal Autopsy Data: A Population-Based Cross-Sectional Study. *PLoS ONE*, 8(9), SEP 13 2013.

[48] Sebsibe Tadesse and Takele 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*, 12, NOV 29 2012.

References V

[49] Biruk Tensou, Tekebash Araya, Daniel S. Telake, Peter Byass, Yemane Berhane, Tolcha Kebebew, Eduard J. Sanders, and Georges Reniers. Evaluating the InterVA model for determining AIDS mortality from verbal autopsies in the adult population of Addis Ababa. *Tropical Medicine & International Health*, 15(5):547–553, MAY 2010.

[50] Stefania Vergnano, Edward Fottrell, David Osrin, Peter N. Kazembe, Charles Mwansambo, Dharma S. Manandhar, Stephan P. Munjanja, Peter Byass, Sonia Lewycka, and Anthony Costello. 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. *Population Health MetricsS*, 9, AUG 5 2011.

[51] Zhaodong Yan, Serena Jeblee, and Graeme Hirst. Can character embeddings improve cause-of-death classification for verbal autopsy narratives? *BioNLP 2019 - SIGBioMed Workshop on Biomedical Natural Language Processing, Proceedings of the 18th BioNLP Workshop and Shared Task*, pages 234–239, 2019. ISBN: 9781950737284.

[52] Yazoume Ye, Marilyn Wamukoya, Alex Ezeh, Jacques Emina, and Osman Sankoh. Health and demographic surveillance systems: a step towards full civil registration and vital statistics system in sub-Sahara Africa? *BMC public health*, 12(1):741, September 2012.

Algorithm comparison 1

Algorithm	Individual-Level Agreement: Overall Agreement Top Cause (95% CI)	Overall Agreement Top 3 Causes (95% CI)	Kappa Top cause (95% CI)	Chance Corrected Concordance Top Cause	Population-Level: CSMF Accuracy	Spearman Rank correlation (95% CI)
Total Sample N = 4,534						
InterVA-5	48.2 (46.7 – 49.7)	70.9 (69.6 – 72.2)	0.43 (0.42 – 0.44)	0.39	0.81	0.64 (0.62 – 0.65)
InSilicoVA	51.6 (50.2 – 53.1)	73.8 (72.5 -75.1)	0.47 (0.46 – 0.48)	0.42	0.84	0.68 (0.67 – 0.70)
Tariff 2.0	51.2 (49.8 – 52.7)	*	0.46 (0.45 – 0.47)	0.38	0.82	0.66 (0.65 – 0.68)
Neonate (0-27 days) N = 82						
InterVA-5	78.5 (67.5 – 86.4)	78.0 (67.5 – 86.4)	-0.05 (-0.14 – 0.04)	0.13	0.90	0.02 (-0.02 – 0.24)
InSilicoVA	79.3 (68.9 – 87.4)	79.3 (68.9 – 87.4)	-0.04 (-0.14 – 0.05)	0.13	0.84	0.02 (-0.02 – 0.24)
Tariff 2.0	47.6 (36.4 – 58.9)	*	0.01 (-0.04 – 0.06)	0.06	0.83	-0.06 (-0.27 – 0.16)
Child (28 days – 11 years) N = 165						
InterVA-5	36.4 (29.0 – 44.2)	50.3 (42.4 – 58.2)	0.32 (0.28 – 0.36)	0.32	0.66	0.43 (0.30 – 0.55)
InSilicoVA	40.6 (33.0 – 48.5)	56.4 (48.4 – 64.1)	0.36 (0.32 – 0.40)	0.40	0.64	0.65 (0.55 – 0.73)
Tariff 2.0	28.5 (21.7 – 36.0)	*	0.24 (0.20 – 0.28)	0.20	0.66	0.50 (0.37 – 0.60)

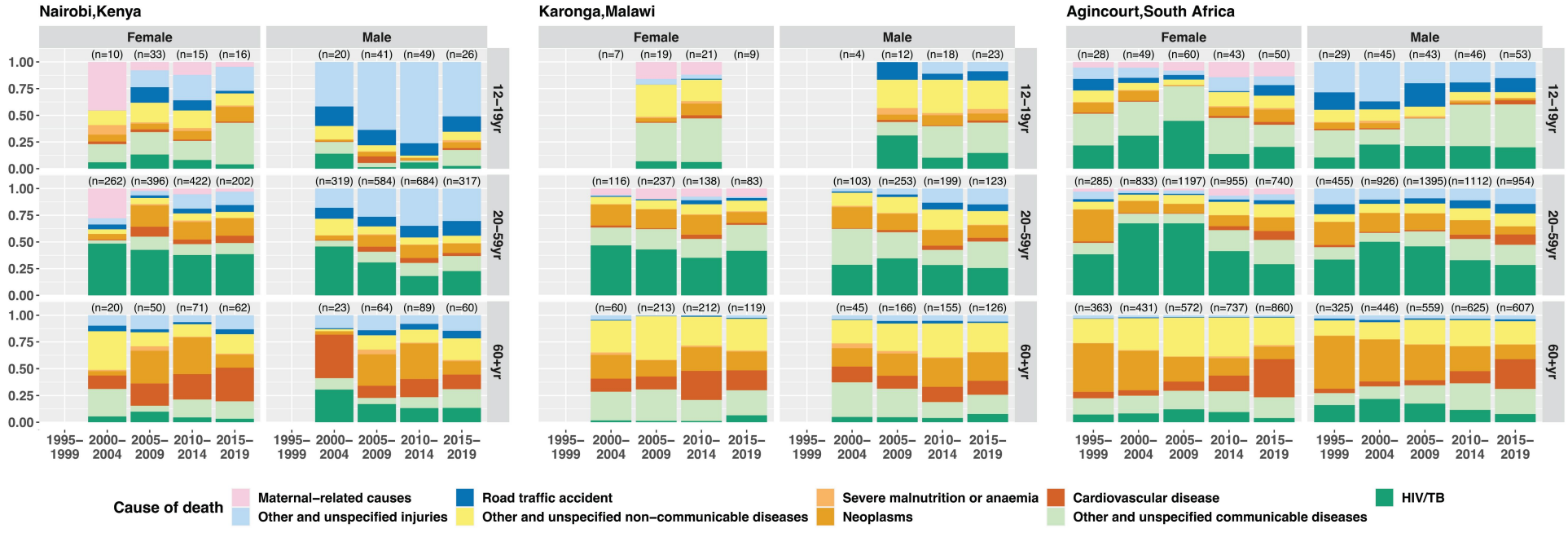
Algorithm comparison 2

Algorithm	Individual-Level Agreement: Overall Agreement Top Cause (95% CI)	Overall Agreement Top 3 Causes (95% CI)	Kappa Top cause (95% CI)	Chance Corrected Concordance Top Cause	Population-Level: CSMF Accuracy	Spearman Rank correlation (95% CI)
Adult (12 – 49 years) N = 1,812						
InterVA-5	58.3 (56.0 – 60.6)	77.2 (75.1 – 79.1)	0.49 (0.47 – 0.51)	0.34	0.80	0.75 (0.73 – 0.77)
InSilicoVA	62.3 (60.0 – 64.5)	79.7 (77.8 – 81.5)	0.54 (0.52 – 0.56)	0.37	0.84	0.79 (0.77 – 0.81)
Tariff 2.0	65.4 (63.2 – 67.6)	*	0.56 (0.55 – 0.58)	0.38	0.89	0.75 (0.73 – 0.75)
Elder (50+ years) N =2,475						
InterVA-5	40.6 (38.7 – 42.6)	67.5 (65.6 – 69.4)	0.35 (0.34 – 0.36)	0.34	0.80	0.47 (0.44 – 0.50)
InSilicoVA	43.7 (41.7 – 45.7)	70.4 (68.6 – 72.2)	0.38 (0.37 – 0.40)	0.37	0.83	0.51 (0.48 – 0.54)
Tariff 2.0	42.5 (40.5 – 44.5)	*	0.38 (0.37 – 0.39)	0.39	0.76	0.55 (0.52 – 0.58)
Male N = 2,400						
InterVA-5	47.7 (45.7 – 49.7)	71.1 (69.3 – 72.9)	0.43 (0.41 – 0.44)	0.39	0.76	0.68 (0.66 – 0.70)
InSilicoVA	52.0 (50.0 – 54.0)	74.1 (72.3 – 75.9)	0.47 (0.46 – 0.48)	0.43	0.80	0.73 (0.71 – 0.73)
Tariff 2.0	55.3 (53.2 – 57.3)	*	0.50 (0.49 – 0.52)	0.39	0.84	0.71 (0.69 – 0.73)

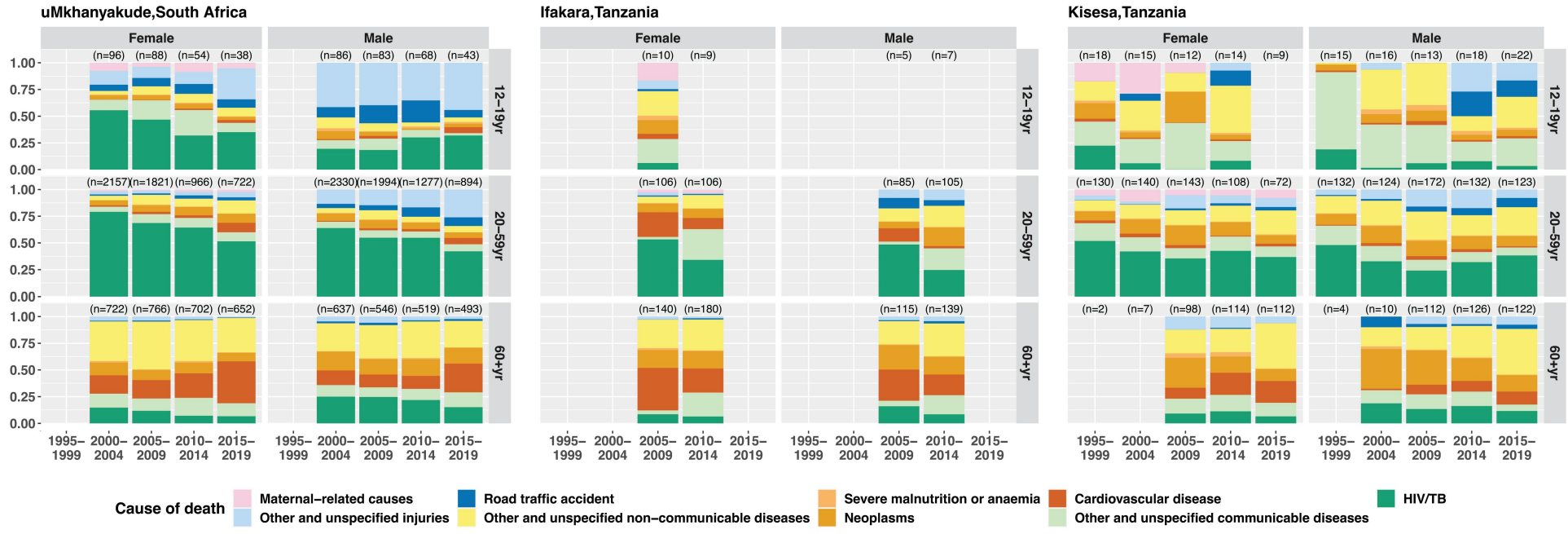
Algorithm comparison 3

Algorithm	Individual-Level Agreement: Overall Agreement Top Cause (95% CI)	Overall Agreement Top 3 Causes (95% CI)	Kappa Top cause (95% CI)	Chance Corrected Concordance Top Cause	Population-Level: CSMF Accuracy	Spearman Rank correlation (95% CI)
Female N =2,134						
InterVA-5	48.8 (46.6 – 50.9)	70.7 (68.7 – 72.6)	0.43 (0.41 – 0.44)	0.38	0.87	0.57 (0.54 – 0.60)
InSilicoVA	51.2 (49.1 – 53.4)	73.4 (71.5 – 75.2)	0.46 (0.44 – 0.47)	0.40	0.88	0.61 (0.59 – 0.64)
Tariff 2.0	46.7 (44.6 – 48.9)	*	0.41 (0.40 – 0.43)	0.34	0.79	0.60 (0.57 – 0.63)
Died in Health Facility N = 2,591						
InterVA-5	47.4 (45.4 – 49.3)	71.2 (69.4 – 72.9)	0.41 (0.40 – 0.42)	0.38	0.80	0.57 (0.54 – 0.59)
InSilicoVA	50.6 (48.7 – 52.6)	73.5 (71.8 – 75.2)	0.45 (0.43 – 0.46)	0.43	0.82	0.63 (0.60 – 0.65)
Tariff 2.0	51.6 (49.7 – 53.5)	*	0.45 (0.44 – 0.46)	0.36	0.82	0.65 (0.63 – 0.67)
Died Out of Health Facility N = 1,943						
InterVA-5	49.3 (47.1 – 51.6)	70.6 (68.5 – 72.6)	0.45 (0.44 – 0.46)	0.39	0.82	0.71 (0.69 – 0.73)
InSilicoVA	53.0 (50.7 – 55.2)	74.1 (72.1 – 76.0)	0.49 (0.48 – 0.50)	0.41	0.85	0.74 (0.72 – 0.76)
Tariff 2.0	50.7 (48.5 – 53.0)	*	0.47 (0.46 – 0.48)	0.40	0.81	0.66 (0.64 – 0.69)

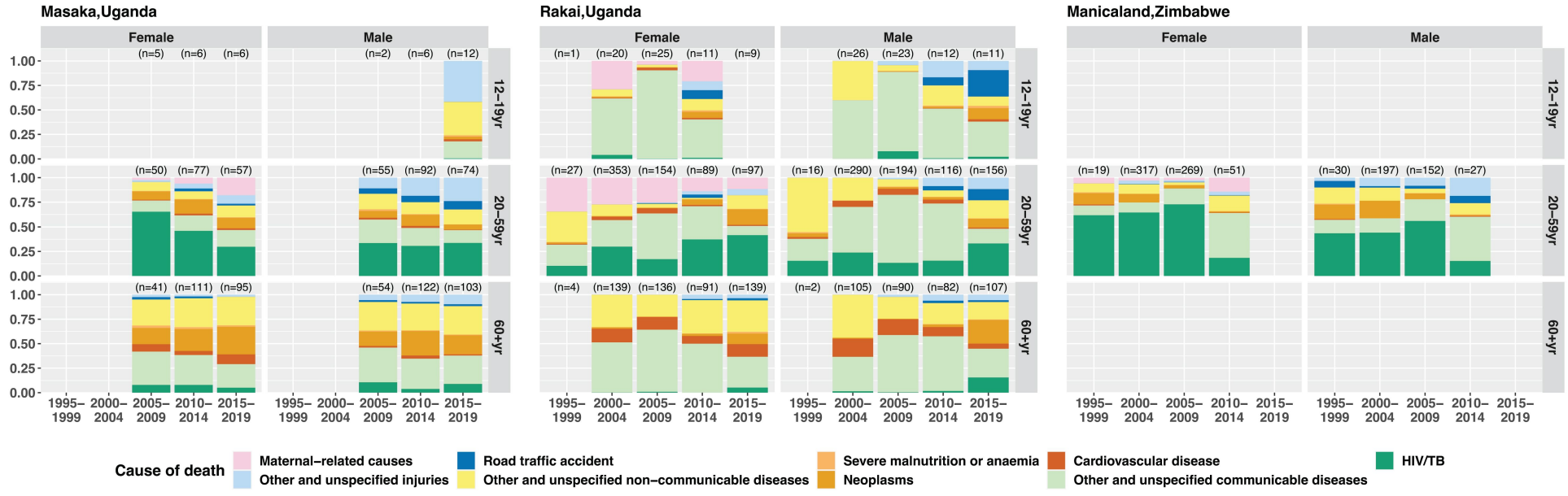
ALPHA Network cause-specific mortality fractions 1



ALPHA Network cause-specific mortality fractions 2



ALPHA Network cause-specific mortality fractions 3



ALPHA Network CSMFs

ALPHA Network cause-specific mortality rates