DEMOGRAPHIC RESEARCH
*A peer-reviewed, open-access journal of population sciences*

***DEMOGRAPHIC RESEARCH***

**VOLUME 52, ARTICLE 8, PAGES 229–272**
**PUBLISHED 7 FEBRUARY 2025**
https://www.demographic-research.org/Volumes/Vol52/8
DOI: 10.4054/DemRes.2025.52.8

*Research Article*

# Studying multiple causes of death through verbal autopsies: The contribution of a similarity index

**Ariane Sessego**

**Géraldine Duthé**

**Bruno Lankoandé**

**Kassoum Dianou**

# Contents

*Demographic Research*: Volume 52, Article 8
*Research Article*

# Studying multiple causes of death through verbal autopsies: The contribution of a similarity index

**Ariane Sessego**[1]

**Géraldine Duthé**[2]

**Bruno Lankoandé**[3]

**Kassoum Dianou**[4]

## Abstract

**BACKGROUND**
The analysis of multiple causes of death was developed in high-income countries to study complex morbid processes leading to death. In other countries, such studies are severely limited by the lack of death certificates. Some cause-of-death statistics are produced at the local level through verbal autopsies (VAs): the collecting of information on medical history and symptoms reported by the final caregiver. Algorithmic models have been developed to estimate probable causes of death in a standardized and cost-effective manner. We investigate their potential to identify multiple causes.

**OBJECTIVE**
Bayesian models provide probabilities for all possible causes for each death. If multiple causes are probable, it could indicate multimorbidity or an uncertain diagnosis. In this paper, we aim to distinguish between these two possibilities.

**METHODS**
The INDEPTH Network provides a dataset of 72,300 adult deaths from 22 Health and Demographic Surveillance System (HDSS) sites in Asia and Africa, disaggregated by age, sex, and probable causes of death as determined by the InterVA-4 model. Using the model's probability matrix, we estimated the degree of similarity between causes and identified those with significant dissimilarities as probable multimorbidities. We test our approach using detailed VA data from the Ouagadougou HDSS (1,714 deaths).

[1] Centre Maurice Halbwachs, École des hautes études en sciences sociales (EHESS), France, and Institut national d'études démographiques (INED), France. Email: ariane.sessego@gmail.com.
[2] Institut national d'études démographiques (INED), France.
[3] Institut supérieur des sciences de la population (ISSP), Université Joseph Ki-Zerbo, Burkina Faso.
[4] Institut supérieur des sciences de la population (ISSP), Université Joseph Ki-Zerbo, Burkina Faso, and Université Catholique de Louvain-la-Neuve, Belgium.

**RESULTS**
InterVA-4 assigns at least two probable causes to 11% of deaths, but only 2% are identified as having multiple causes.

**CONCLUSIONS**
This proportion is low, but our approach remains conservative, as we cannot identify multimorbidity for similar causes.

**CONTRIBUTION**
This study advocates for better knowledge of multiple causes of death in low- and middle-income countries by providing a first approach to their identification through VAs.

## 1. Background

Multimorbidity has become a public health priority, especially in contexts where mortality depends mainly on adult deaths and non-communicable diseases (NCDs) (WHO 2022). Mostly chronic, NCDs may co-occur or interact (Désesquelles et al. 2015) and often require routine monitoring and costly treatment. In particular, multimorbidity significantly complicates care pathways and exposes individuals to poor health and disability (Skou et al. 2022).

Since the end of the 19$^{th}$ century, international efforts to harmonize cause-of-death statistics have emphasized the need to assign a single cause to each death – i.e., to identify the "underlying" cause that "initiated the series of events leading directly to death" (ICD-10-WHO 2019a: 29; Star and Bowker 1999). Following WHO recommendations, certifying physicians are asked to specify in death certificates diseases or conditions that were directly involved in the dying process to determine the underlying cause of death, as well as those that contributed to it. Recent studies emphasize the importance of considering all causes recorded on the death certificate to highlight associations and to re-evaluate the burden of diseases that are rarely considered the underlying cause of death, such as diabetes (Désesquelles et al. 2010, 2015, 2016; Barbieri et al. 2017). This multi-cause approach provides an opportunity to emphasize the role of multimorbidity in mortality (Multiple Causes-of-Death Network 2023).

This approach has mainly been applied in high-income countries. However, in low- and middle-income countries (LMICs), multimorbidity is becoming an increasing concern (Banerjee et al. 2020). Since the 1950s, LMICs have seen a significant aging of the population due to declining fertility rates and important health improvements (United Nations 2022). The age at death is increasing, as is the proportion of NCDs in mortality (Ahmed et al. 2023; Chu et al. 2024). This epidemiological transition (Omran 1971) is

taking place in a context of rapid urbanization and development, accompanied by a concerning increase in unhealthy behaviours (lack of physical activity, unbalanced diet), major risk factors for NCDs (obesity, hypertension, diabetes), and absence of adequate care (Duthé et al. 2024). Seventeen million people die each year from an NCD before the age of 70, and more than 85% of these "premature" deaths occur in LMICs (WHO 2022; United Nations 2015).

These trends represent a challenge to health care systems and public policies that are ill-equipped to prevent and manage NCDs (Kushitor and Boatemaa 2018; Barr et al. 2016; Martini and Figg 2010). This rise in NCDs is also taking place in a context where the prevalence of infectious diseases (HIV, malaria, and tuberculosis) remains high, leading to a double or cumulative burden of disease, particularly in sub-Saharan Africa (Boutayeb 2006; Kolčić 2012; Tabutin and Masquelier 2017; Ciccacci 2020). Infectious diseases also play a role in multimorbidity, especially if they are chronic or episodic. People living with HIV, for example, require continuous HIV care and are susceptible to opportunistic diseases (especially tuberculosis). People living with NCDs are also more susceptible to infectious or parasitic diseases (e.g., malaria, COVID-19) (Remais et al. 2013; Désesquelles et al. 2015). These interactions between infectious and non-communicable, acute, and chronic diseases are important in high-mortality contexts but remain poorly understood.

Unfortunately, in many LMICs studies are limited by the lack of cause-of-death statistics. Mikkelsen et al. (2015) estimate that between 2010 and 2012, six out of ten deaths worldwide were not properly recorded in vital statistics databases, and this proportion is even higher for causes of death, as they require medical certification. According to the WHO (2019b), only 6% of deaths in Africa and 10% in Southeast Asia were registered with a cause of death in 2017. The Global Burden of Diseases (GBD) provides global and national estimates for cause-specific mortality, but where data are not available, they are mostly based on models and expert opinion, limiting precise investigations (Foreman et al. 2012; GBD 2019 Diseases and Injuries Collaborators 2020). In some countries where health statistics are not available, Health and Demographic Surveillance System (HDSS) sites provide cause-of-death statistics at the local level using the verbal autopsy (VA) method (Chandramohan et al. 2021). HDSS sites monitor, among other things, births, deaths, and migration within a geographically defined population. When a death is registered, a VA (postmortem interview) is conducted with an informant who was involved in the care of the deceased to collect information on the medical history (symptoms and treatment) leading up to the death. This method is primarily aimed at determining probable causes for deaths that occurred outside of health care facilities. In the past, VA questionnaires were evaluated by physicians to determine causes of death. Since the early 2000s, however, algorithms have been developed to determine probable causes in a faster, more standardized, and more

cost-effective manner. While a variety of methods exist (McCormick et al. 2016), Bayesian algorithms are currently the most widely used, particularly InterVA (Byass et al. 2019). InterVA is based on a set of probabilities developed by a committee of experts, who associate each piece of information reported in the VA (age, sex, symptoms) with each cause of a preexisting classification. For a given death, the algorithm estimates the probability of each cause and selects up to three most probable causes and their associated probabilities (see Appendix 1 for details). This method allows the estimation of cause-specific mortality fractions at the population level by summing the probabilities for each cause. The algorithm is based on the routinely revised standard VA questionnaire proposed by the WHO (2024) and is regularly updated (Byass, Huong, and Hoang 2003; Fantahun et al. 2006; Byass et al. 2012).

This Bayesian approach was developed in the context of imperfect and incomplete information. The assignment of up to three causes for a given death is primarily intended to reflect this uncertainty. But could this approach also be used to identify multiple causes of death? In this paper, we aim to investigate the extent to which this tool can identify multimorbid processes leading to death.

We develop an innovative approach to distinguish between multiple or co-occurring causes that result from multimorbidity and associations of causes that could result from uncertainty. We assume that two causes that have distinct sets of symptoms are likely to be co-occurring, as the risk of confusion is low. In contrast, associated causes with similar sets of symptoms have a high risk of confusion. This does not mean that causes with similar symptomatologies cannot co-occur; it means that given the information provided, we are unable to distinguish co-occurring causes from causes that are co-assigned due to uncertainty. Following this hypothesis, we constructed a similarity index to characterize the degree of similarity between causes based on the predefined conditional probabilities provided by experts in the InterVA software. We first apply this approach to a large dataset of adult deaths collected in 22 HDSS sites in sub-Saharan Africa and Asia where the probable causes of death were assigned by InterVA-4 but the detailed reported symptoms are not available. Using detailed VA data from the Ouagadougou HDSS, we then test the robustness of this theoretical index – based on the general definition of each cause, including all potentially reported symptoms – with an individual-based index – based on detailed data from each individual VA.

# 2. Data

## 2.1 Cause-of-death statistics collected in 22 HDSS sites

The INDEPTH Network, compiled in 2014, is an open-access dataset of deaths from 22 HDSS sites in Africa and Asia (INDEPTH 2014). For each death, the dataset provides the age group, sex, HDSS site, year of death, and one to three causes of death with associated probabilities, determined using InterVA-4. To protect personal data, the dataset does not include the detailed set of symptoms reported in the VA, as is usually the case with VA cause-specific mortality datasets. The deaths occurred in 13 countries in sub-Saharan Africa and Asia, most of them in the first decade of the 2000s (INDEPTH 2014). The dataset is described and analysed in a special issue of *Global Health Action* dedicated to cause-specific mortality in LMICs (Streatfield et al. 2014a). Streatfield et al. (2014a, 2014b) show that the proportion of infectious and parasitic diseases remains high, even among adults, and that the cumulative burden of infectious diseases and NCDs is particularly important in rural Africa. We selected 72,330 deaths of adults aged 15 and older. Of these deaths, only one-third are assigned a cause with a probability of 1, and 5% are considered indeterminate (Appendix 1, Table A-1.3). Nearly 11% of the deaths are attributed more than one probable cause (Table 1; see Appendix 1 for details on the criteria for attributing more than one cause). The difference by age group is small, ranging from 9.0% for deaths of those aged 15 to 49 to 12.1% for deaths of those over age 65 (Appendix 2). As the percentage of deaths with a third cause is negligible (0.5%), we excluded the third causes from the analysis and focused on the first two most probable causes.[5] These causes are analysed regardless of their order of attribution (i.e., regardless of which one has the highest probability), since we are primarily interested in associations of causes. The analysis was conducted on the 7,734 adult deaths characterized by at least two causes in the dataset.

**Table 1: Number of probable causes assigned to each death by InterVA**

| Number of causes | Frequency | % |
|---|---|---|
| 1 cause | 64,596 | 89.3 |
| More than 1 cause | 7,734 | 10.7 |
| 2 causes | 7,389 | 10.2 |
| 3 causes | 345 | 0.5 |

*Scope*: 72,330 adult deaths from 22 HDSS sites (1992–2012).
*Source*: INDEPTH Network 2014.

[5] If the probabilities of the causes selected by InterVA do not add up to 1, the residual is considered indeterminate to account for uncertainty (see Appendix 1). We do not consider this indeterminate residual to be an additional cause of death.

## 2.2 Detailed VA data from the Ouagadougou HDSS

To test the method, the same approach is applied to the detailed information collected in the VAs. We use detailed VA data from the Ouagadougou HDSS collected between 2010 and 2019. This site monitors inhabitants of five neighbourhoods located on the outskirts of the capital of Burkina Faso. As this is an urban setting, the prevalence of NCDs is higher there than in our first dataset (Streatfield et al. 2014a). Of the 1,714 deaths in this dataset, 186 (10.8%) were assigned more than one cause. Given the relatively small sample size, deaths of all age groups are included in the analysis.

## 2.3 Method

### 2.3.1 Elaboration of a theoretical similarity index to isolate probable multiple causes

In the InterVA algorithm, each cause is defined by a vector of probabilities, one for each symptom or characteristic (age, sex, treatments, etc.) potentially reported in the VA. In this vector, each indicator is associated to a probability of being reported given that cause of death. These probabilities were defined by a committee of experts using letter grades and then transformed into numerical values according to Byass (2012) (see Appendix 1). We use these numerical values here. We evaluate the similarity between two causes by measuring the distance between the probabilities they associate with each indicator. To this end, we have created a similarity index that calculates, for each pair of causes, the distance between their respective conditional probability vectors normalized to the sum of the vectors. Formally, let *A* and *B* be the vectors of probabilities associated with cause *a* and cause *b*, respectively. The similarity index between *a* and *b* is:

$$\boldsymbol{I_{a,b}} = \boldsymbol{I_{b,a}} = \frac{\|\boldsymbol{A}-\boldsymbol{B}\|}{\|\boldsymbol{A}+\boldsymbol{B}\|}. \qquad (1)$$

The more similar the causes, the closer the index is to 0, and the more different the causes, the closer the index is to 1. We applied this formula with two different norm functions, the classical Euclidean norm[6] and the absolute norm,[7] to check the robustness of the index, and we calculated this index for all possible pairs of causes. Both indices

[6] Defined as follows: $\|A - B\| = \sqrt{\sum_{i \in I}\left(P(s_i|a) - P(s_i|b)\right)^2}$, where $P(s_i|c)$ is the conditional probability of presenting the indicator $s_i$ (symptom or characteristic reported in the VA) given cause *c* and *I* is the total number of indicators possibly reported in a VA.

[7] Defined as follows: $\|A - B\| = \sum_{i \in I}|P(s_i|a) - P(s_i|b)|$.

are very similar and highly correlated (Appendix 3). We decided to use the Euclidean norm index because it is a standard measure of distance. Two other possible normalizations and refinements of this method, in particular age- and sex-specific indices, were also tested, with very high correlations (Appendix 3).

### 2.3.2 Calculation of an individual-based index according to symptoms reported in VAs

The theoretical index is calculated from all the information potentially reported in the VA questionnaire. However, to make its diagnosis, InterVA uses only probabilities related to the signs and symptoms positively reported in the VA and does not consider probabilities related to the absence of symptoms or to unknown information (Appendix 1). This approach could potentially change the similarity between two causes, depending on the symptoms reported. For this reason, we created an individual-based index built on the same formula as the theoretical index but taking into account only indicators that were actually reported in each VA. We used the Euclidean norm to measure the distances between vectors as follows:

With $\boldsymbol{P(s_i|a)}$, the conditional probability of the presence of the indicator *i* given cause is *a*. Let J be the set of reported indicators, with $\boldsymbol{i \in J \Leftrightarrow s_i = 1}$ (i.e., $\boldsymbol{s_i}$ reported in the VA). The individual-based index of similarity is:

$$\boldsymbol{I_{a,b} = I_{b,a} = \frac{\sqrt{\sum_{i \in J}\left(P(s_i|a) - P(s_i|b)\right)^2}}{\sqrt{\sum_{i \in J}\left(P(s_i|a) + P(s_i|b)\right)^2}}}. \qquad (2)$$

In this index, the symptoms that are not reported positive in the VA are not included in the calculation. Each death with more than one cause is thus associated with an individualized index value – i.e., a given combination of causes can be associated with different index values depending on the symptoms reported for each death. We used the Ouagadougou HDSS data to compare the two indices: the theoretical one, using the cause-of-death statistics assigned by InterVA-4, and the individualized one, using detailed VA data.

The scripts to calculate the theoretical and individual-based indices, as well as the other indices tested, are openly available in *.R format in a dedicated Github repository,[8] along with the resulting datasets and fictitious data to apply the code when needed.

[8] https://github.com/ariane-sessego/Studying_multiple_CoD_through_VAs_similarity_index.

Interactive visualizations of the theoretical indices tested are also available online as supplementary materials.[9]

## 3. Results

### 3.1 Combinations of causes

Table 2 presents the 19 most frequent combinations of causes (for more than 100 deaths) among the 7,734 deaths assigned at least two probable causes by InterVA. The algorithm can theoretically identify co-occurring diseases at the time of death based on the reported symptoms. Tuberculosis and HIV-related deaths (Table 4, no. 1), stroke and heart disease (no. 3), and stroke and diabetes (no. 10) are among the most common associations and appear to be plausible comorbidities. Tuberculosis is indeed one of the most common opportunistic diseases leading to death for people living with HIV, and diabetes and heart disease are risk factors for strokes. In these cases, it seems likely that both causes contributed to the process leading to death.

However, the Bayesian approach is also used to deal with the uncertainty generated by the lack of information. Some associations – which we refer to as competing – seem likely to result from the difficulty of deciding between several probable causes that are most likely mutually exclusive. For example, pneumonia and malaria – one of the most frequent associations (Table 2, no. 6) – are notoriously difficult to distinguish without a biometric test (Källander, Nsungwa-Sabiiti, and Peterson 2004). In this case, their association may be more related to uncertainty than to multimorbidity. The same applies to all diseases that have a similar set of symptoms and affect the same organ(s) or are localized in the same part of the body, such as tuberculosis and respiratory neoplasms (Table 2, no. 4).

[9] *VACauseSimilarity* Shiny app here.

**Table 2: Most frequent association of causes assigned by InterVA**

| No. | Cause A | Cause B | Frequency | %* |
|---|---|---|---|---|
| 1 | Pulmonary tuberculosis | HIV/AIDS-related death | 392 | 5.1 |
| 2 | Acute respiratory infection including pneumonia | Sepsis (non-obstetric) | 339 | 4.4 |
| 3 | Acute cardiac disease | Stroke | 319 | 4.1 |
| 4 | Pulmonary tuberculosis | Respiratory neoplasms | 309 | 4.0 |
| 5 | Reproductive neoplasms (male and female) | Digestive neoplasms | 268 | 3.5 |
| 6 | Acute respiratory infection including pneumonia | Malaria | 208 | 2.7 |
| 7 | Acute cardiac disease | Other and unspecified cardiac disease | 184 | 2.4 |
| 8 | Acute respiratory infection including pneumonia | Pulmonary tuberculosis | 179 | 2.3 |
| 9 | Digestive neoplasms | Other and unspecified neoplasms | 179 | 2.3 |
| 10 | Diabetes mellitus | Stroke | 148 | 1.9 |
| 11 | Road traffic accident | Assault | 148 | 1.9 |
| 12 | Acute respiratory infection including pneumonia | Other and unspecified cardiac disease | 141 | 1.8 |
| 13 | Pulmonary tuberculosis | Other and unspecified cardiac disease | 123 | 1.6 |
| 14 | HIV/AIDS-related death | Intentional self-harm | 117 | 1.5 |
| 15 | Acute abdomen | Digestive neoplasms | 115 | 1.5 |
| 16 | Other and unspecified cardiac disease | Chronic obstructive pulmonary disease | 115 | 1.5 |
| 17 | Pulmonary tuberculosis | Chronic obstructive pulmonary disease | 109 | 1.4 |
| 18 | Acute abdomen | Diarrhoeal diseases | 103 | 1.3 |
| 19 | Stroke | Other and unspecified cardiac disease | 100 | 1.3 |

*Among deaths attributed more than one cause.
*Notes*: Associations with frequency >/= 100. Associations are considered irrespective of the order of the associated probabilities.
*Scope*: 72,330 adult deaths in 22 HDSS sites (1992–2012).
*Source:* INDEPTH Network 2014.

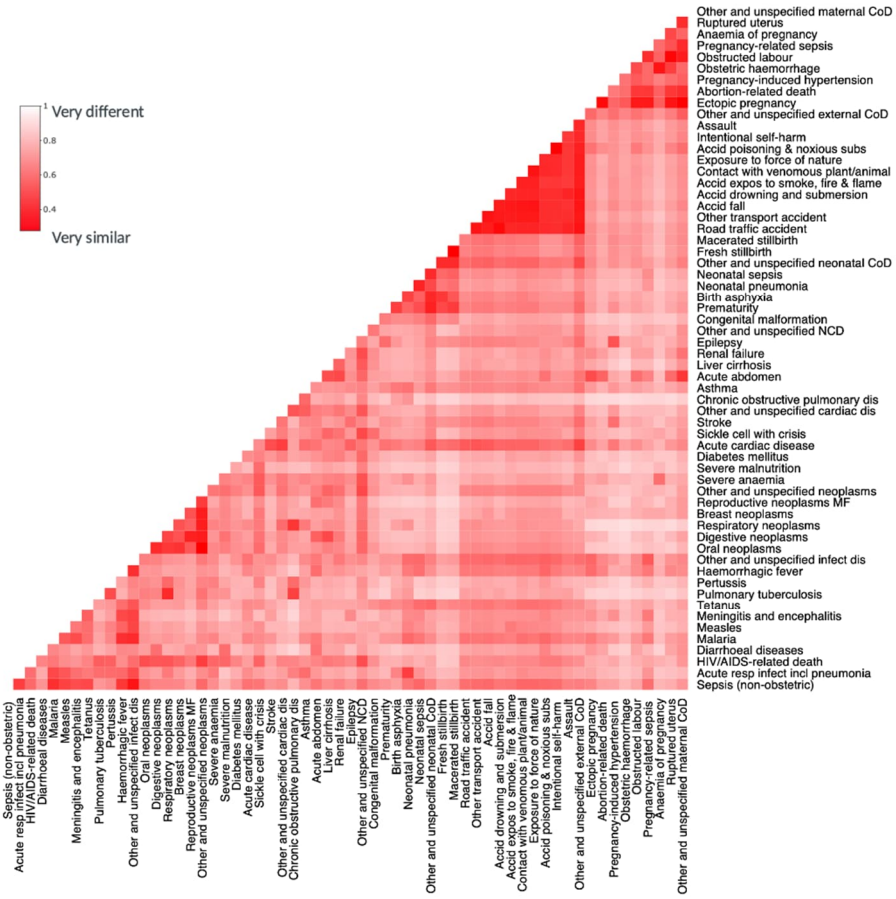
### 3.2 Theoretical similarity between causes

To distinguish the co-occurring causes from the probable competing causes, we used the theoretical similarity index. Figure 1 shows the heat map of similarity between all causes as defined by the theoretical index. Some groups of causes stand out as particularly different. We expect this for maternal and neonatal causes, as they are specific to certain groups of individuals (according to age group and sex) and are related to specific events (pregnancy and birth). However, this is also the case for other diseases, such as chronic obstructive pulmonary disease, pulmonary tuberculosis, malnutrition, and diarrhoeal diseases. Within groups of causes, there are strong similarities, especially within cancers, maternal causes, and external causes (accidents and violent deaths).

We then tested various thresholds to distinguish between co-occurring causes and associations that may result from uncertainty, taking into account the distribution of the index and the plausibility of confusion between causes according to our knowledge. We retained the value of 0.65, which led to selecting approximately 75% of all possible

associations as co-occurring causes while selecting combinations that appear sufficiently dissimilar to be interpreted as probable multimorbidity (Appendices 4 and 5).

**Figure 1: Heat map of similarity between causes as classified by InterVA-4**




*Note*: An interactive version of this figure is available in the supplementary material.
*Source*: InterVA-4's probability matrix (probbase; malaria: VL [very low]; HIV: VL) by cause (our calculations using the Euclidean norm index).
Link to interactive version.

### 3.3 Estimation of probable multiple causes

Of the 7,734 deaths characterized by at least two probable causes, 1,591 (20.6%) met our criterion. The comparison between Figures 2 and 3 shows which associations were selected, cause by cause (see also Appendix 5). Overall, the identified multimorbidity accounts for only 2.2% of all adult deaths. This percentage increases with age, which contributes to the credibility of the method; 1.3% of deaths between the ages of 15 and 49, 2.3% of deaths between the ages of 50 and 64, and 3% of deaths in the 65+ age group are identified as having multiple causes (p-value < 0.000; see Appendix 6). The proportion is also higher in women (2.4%) than men (2%) (p-value < 0.000; see Appendix 6), a result that seems to be consistent with the higher rate of multimorbidity in women (Oksuzyan, Brønnum-Hansen and Jeune 2010).

**Table 3: Most frequent co-occurring causes identified through the theoretical index**

| No. | Cause A | Cause B | Frequency | % |
|---|---|---|---|---|
| 1 | Diabetes mellitus | Stroke | 148 | 9.3 |
| 2 | Acute respiratory infection including pneumonia | Other and unspecified cardiac disease | 141 | 8.9 |
| 3 | Pulmonary tuberculosis | Other and unspecified cardiac disease | 123 | 7.7 |
| 4 | Acute respiratory infection including pneumonia | Acute cardiac disease | 49 | 3.1 |
| 5 | Stroke | Acute abdomen | 48 | 3.0 |
| 6 | Acute respiratory infection including pneumonia | Stroke | 44 | 2.8 |
| 7 | Diabetes mellitus | HIV/AIDS-related death | 43 | 2.7 |
| 8 | Other and unspecified cardiac disease | HIV/AIDS-related death | 43 | 2.7 |
| 9 | Acute abdomen | HIV/AIDS-related death | 42 | 2.6 |
| 10 | Stroke | Digestive neoplasm | 42 | 2.6 |
| 11 | Acute respiratory infection including pneumonia | Acute abdomen | 40 | 2.5 |
| 12 | Stroke | Other and unspecified NCD | 31 | 1.9 |

*Notes*: Associations with frequency >/= 30. Associations are considered irrespective of the order of the probabilities associated.
*Scope*: 1,591 adult deaths with probable co-occurring causes.
*Source*: INDEPTH Network 2014.

Table 3 shows the most frequent co-occurring causes. The most common combination is diabetes and stroke (9.3%), a known multimorbidity, as diabetes is a risk factor for stroke. Combinations of infectious respiratory disease (acute respiratory infection, including pneumonia and pulmonary tuberculosis) and cardiac disease (unspecified, acute, or stroke) account for almost a quarter of multiple causes. Associations involving HIV/AIDS are also relatively common, especially with diabetes (2.7%), unspecified cardiac disease (2.7%), and acute abdomen (2.6%). In general, co-

occurring causes are often a combination of acute and chronic diseases (e.g., diabetes and stroke) or acute respiratory infections and non-acute cardiac disease).

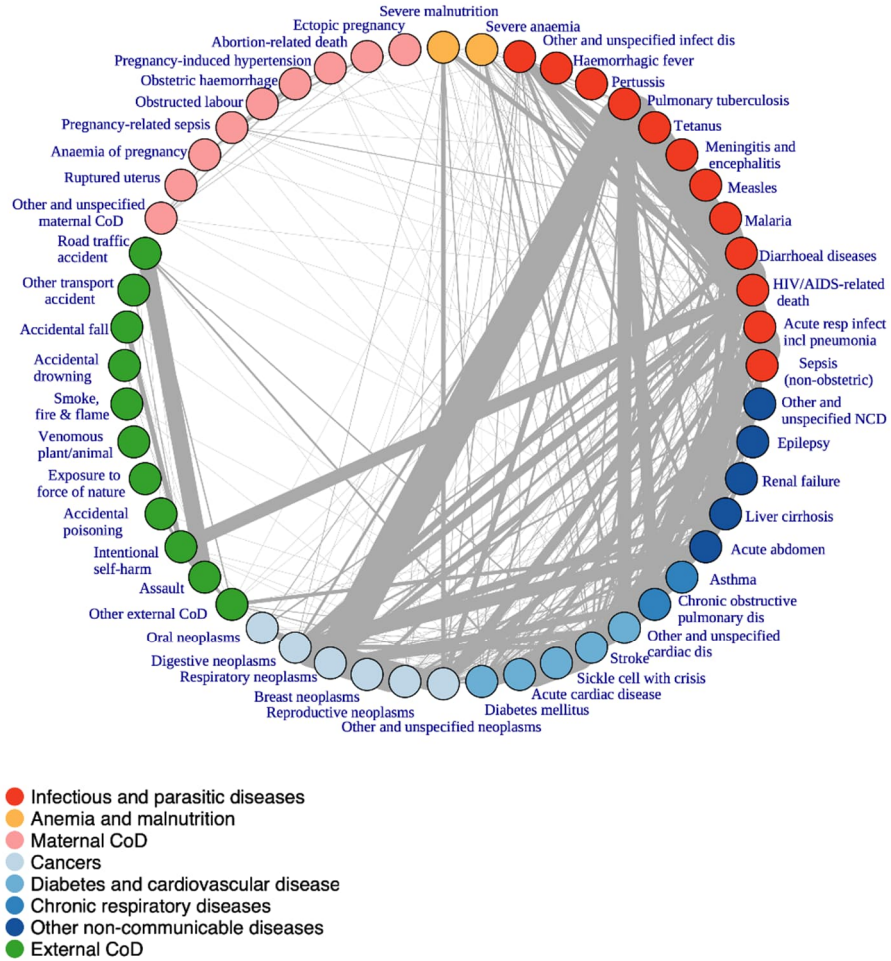
Overall, 61% of deaths from multiple causes involve a combination of an infectious, maternal, or neonatal disease and an NCD (Table 4). Nearly one-third involve two NCDs, only 6% involve two infectious maternal and neonatal diseases, and 2% involve an external cause and an NCD (Table 6). Diabetes and cardiovascular diseases are involved in a high proportion of multimorbidity (about 70%); this is partly explained by the association of these chronic diseases with acute infectious diseases (41%) and by the importance of the association of diabetes and cardiovascular disease with other NCDs (29%).

**Table 4: Distribution of co-occurring causes by group**

| Group A | Group B | Frequency | % |
|---|---|---|---|
| **Non-communicable diseases** | **Infectious, maternal, and neonatal causes** | **976** | **61.3** |
| Diabetes and cardiovascular diseases | Infectious and parasitic diseases | 648 | 40.7 |
| Infectious and parasitic diseases | Other non-communicable diseases | 189 | 11.9 |
| Cancers | Infectious and parasitic diseases | 60 | 3.8 |
| Anaemia and malnutrition | Diabetes and cardiovascular diseases | 31 | 1.9 |
| Chronic respiratory diseases | Infectious and parasitic diseases | 25 | 1.6 |
| **Non-communicable diseases** | **Non-communicable diseases** | **486** | **30.5** |
| Diabetes and cardiovascular diseases | Diabetes and cardiovascular diseases | 165 | 10.4 |
| Diabetes and cardiovascular diseases | Other non-communicable diseases | 132 | 8.3 |
| Cancers | Diabetes and cardiovascular diseases | 111 | 7.0 |
| Chronic respiratory diseases | Diabetes and cardiovascular diseases | 39 | 2.5 |
| Cancers | Other non-communicable diseases | 28 | 1.8 |
| **Infectious, maternal, and neonatal causes** | **Infectious, maternal, and neonatal causes** | **97** | **6.1** |
| Infectious and parasitic diseases | Infectious and parasitic diseases | 69 | 4.3 |
| Non-communicable diseases | Injuries and violent deaths | 29 | 1.8 |

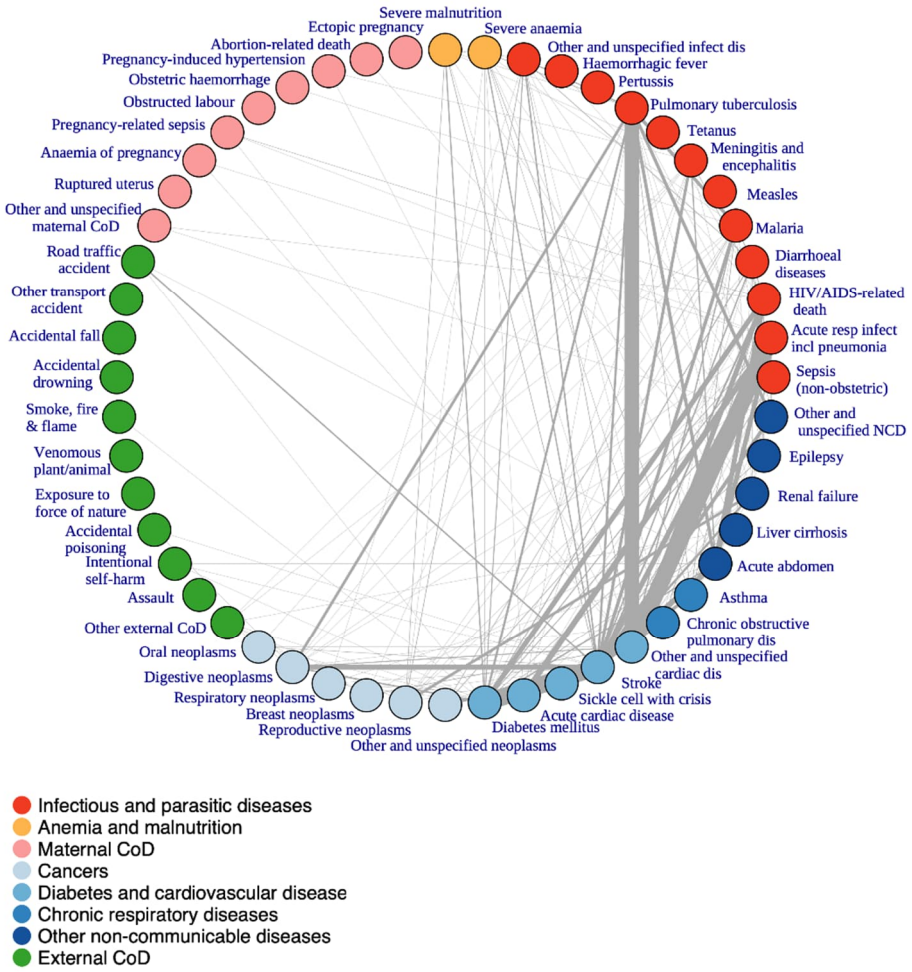
*Note*: Associations are considered irrespective of the order of the probabilities associated.
*Scope*: 1,591 (1992–2013) adult deaths with probable multiple causes (Euclidean index >/= 0.65).
*Source*: INDEPTH Network 2014.

**Figure 2: Network of all associations of causes estimated by InterVA**




*Notes*: Nodes represent causes of death. Edges represent associations between causes. The width of the edge is proportional to the frequency of the association.
Scope: 7,734 VAs of adults with more than one cause of death.
Source: INDEPTH Network, 1992–2012.
Link to the interactive version.

## Figure 3: Network of probable co-occurring causes




*Notes*: Nodes represent causes of death. Edges represent associations between causes. The width of the edge is proportional to the frequency of the association.
Scope: 7,734 VAs of adults with more than one cause of death.
Source: INDEPTH Network, 1992–2012.
Link to the interactive version.

### 3.4 Confronting the approach to detailed VA data: the individual-based index

Using the detailed information collected in VAs in the Ouagadougou HDSS, we compare the similarity of all potentially reported symptoms – the theoretical similarity index – with the similarity of the symptoms actually reported in each VA – the individual-based index. On average, only 21 indicators were reported in the VAs, compared to the 245 possible indicators used in our first index (Table 5). In fact, not all questions are asked to respondents in VAs. The number of questions asked depends on age, sex, and circumstances of death (reported accident, cough, etc.).[10] But the limited number of indicators reported also reflects the limited information present in VAs; respondents do not know everything about the disease and its treatment. This bias could potentially alter the similarity between two causes depending on the reported symptoms, thus limiting the validity of the theoretical index.

**Table 5: Number of demographic and symptomatic indicators in the Ouagadougou VAs**

| | Number of indicators |
|---|---|
| Minimum | 3.0 |
| 1st quartile | 14.0 |
| Median | 19.0 |
| Mean | 21.3 |
| 3rd quartile | 27.0 |
| Maximum | 61.0 |

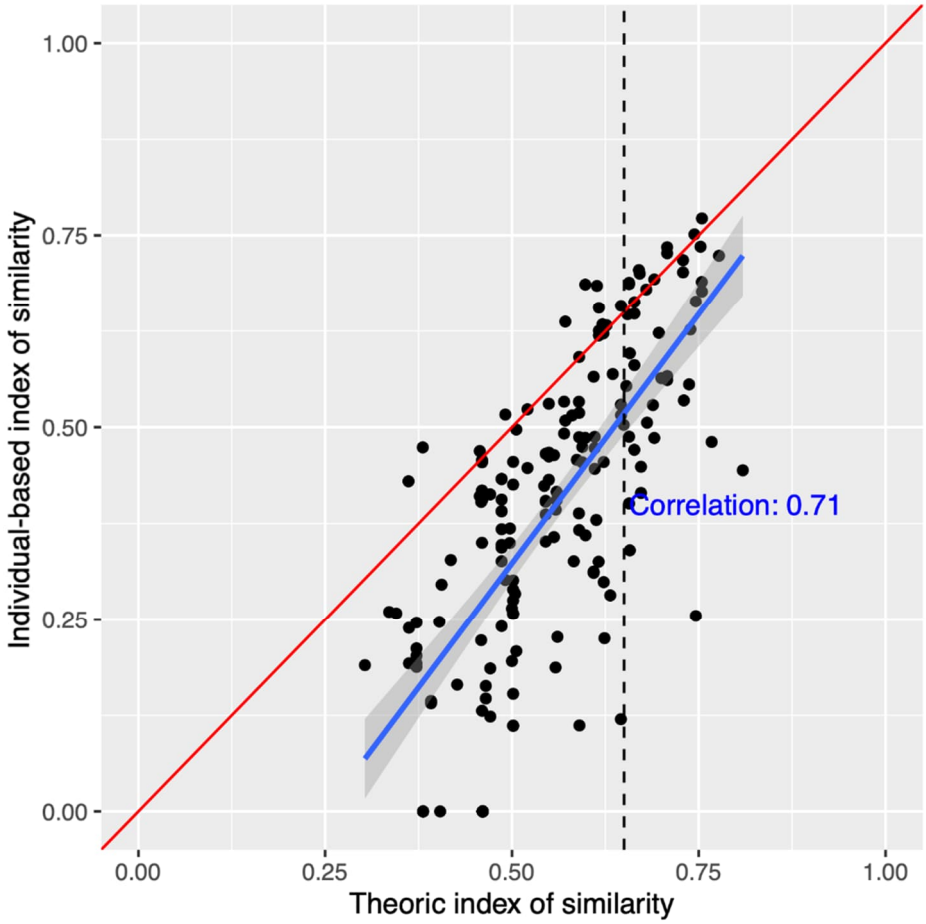
*Scope*: 1,714 deaths of all ages (2010–2019).
*Source*: VAs, Ouagadougou HDSS, Burkina Faso.

Figure 4 shows the association between the two indices. The vast majority of points are located below the diagonal; the value of the individual-based index is generally lower than that of the theoretical one, possibly reflecting more confusion. However, these two indices are not directly comparable, as they were not calculated with the same number and type of indicators.[11] The correlation between the two indices is a more accurate representation of their relationship, and the Pearson coefficient is quite high (0.71). This

[10] To control for age and sex, an age- and sex-specific index was also tested. Very highly correlated with the theoretical index, especially for adults (> 15 years), it shows that they do not constitute the main limitation (see Appendix 3).
[11] In particular, all indicators pertaining to age groups, sex, and circumstances irrelevant to that death have been removed from the calculation. However, when they are irrelevant to both causes, these indicators are often very similar (corresponding to a baseline prevalence in the population). This decreases automatically the overall value of the index.

means that, given the much smaller amount of information actually available in VAs, the relative similarity of causes does not change substantially.

**Figure 4: Correlation between the theoretical index and the individual-based index**




*Scope*: 186 deaths attributed to more than one cause by InterVA-4 (2010–2019, HDSS Ouagadougou).
*Source*: InterVA-4's probability matrix (malaria: VL; HIV: VL) and detailed VAs, HDSS Ouagadougou.

## 4. Discussion

Of the 72,300 adult deaths documented in VAs within the INDEPTH Network in various sites in LMICs from 1992 to 2012, one in ten has more than one probable cause of death according to InterVA-4. However, we could not directly consider these deaths as having multiple causes; some associations could be the result of an uncertain diagnosis. To rule out associations with a high risk of confusion – causes with similar sets of symptoms – we developed a method to estimate the similarity between causes and to select causes sufficiently different to be co-occurring.

Our first approach, the theoretical one, is based on InterVA's conditional probability matrix and aims to evaluate the similarity between causes according to the probabilities associated with all potentially reported symptoms in the VA. We identified 2% of adult deaths that were likely to result from co-occurring causes. The results highlight the prominent role of chronic cardiovascular diseases in this multimorbidity (diabetes and cardiovascular diseases are involved in 70% of the identified multiple causes) and the importance of associations between infectious diseases and NCDs (61% of multiple causes). This illustrates interactions between acute and chronic diseases, underlining the potential of the multiple-cause-of-death approach to investigate cause-specific mortality trends in context with a cumulative burden of diseases.

The proportion of multimorbidity identified is relatively low compared to that in other studies conducted in LMICs (4% to 20% in Southeast Asia in the meta-analysis by Pati et al. 2015 and up to 50% in South Africa [Wade et al. 2021; Wong et al. 2021]). However, these studies use different definitions of multimorbidity – as the concept is not harmonized (Fortin et al. 2012) – and focus on morbidity rather than mortality. Our estimates are also very low compared to studies of multiple causes in high-income countries, where death certificates list on average more than two causes (Désesquelles et al. 2016). Deaths in high-income countries occur on average at older ages, which explains the higher proportion of multiple causes. However, it is also important to keep in mind that our results are likely to largely underestimate the proportion of multiple causes for several reasons.

First, our data source has unavoidable limitations because it is produced in contexts where health information is scarce; it is very different from death certificates filled out by a physician. The information is reported by relatives and is therefore subject to bias; the respondent is not necessarily aware of all the diseases and symptoms of the deceased. In absence of medical care, some of the diseases present may be entirely unknown (especially "silent" diseases such as diabetes).

Second, we are not able to identify multiple causes of death when causes have similar symptoms. In fact, causes with similar symptomatology, such as cancers that spread or combined infections, may co-occur. However, given the information available

through VAs, we are not able to distinguish them from uncertain diagnoses. This certainly leads to underestimation and is probably the main limitation of our approach. The situation is exacerbated by certain features of the algorithm. The conditional probabilities defined by experts to identify a single underlying cause of death sometimes take into account possible opportunistic diseases or risk factors. This is the case in particular for HIV/AIDS, where symptoms of opportunistic infectious diseases are used to attribute the diagnosis, leading to a high degree of similarity with frequently associated diseases, such as tuberculosis or pneumonia, which cannot be interpreted as multiple causes. In addition, InterVA's output relies on a strict and somewhat arbitrary rule to attribute more than one probable cause to a death [12] (see Appendix 1), which likely leads to further underestimation. This limitation could be alleviated in further studies by loosening this selection rule using a large dataset of detailed VAs, but this could not be done with the limited number of deaths available from the Ouagadougou HDSS.

Third, InterVA does not take into account the difference between the absence of symptoms and unknown information, which could be crucial to reducing the set of diseases with similar symptomatology. It would be interesting to test this approach with another Bayesian algorithm that takes this information into account, such as InSilicoVA (Clark et al. 2015).

It is interesting to note that "other and unspecified" causes, especially "other and unspecified cardiac diseases" and "other and unspecified NCDs," often appear in the results. This does not seem to be a direct result of the similarity index, as these unspecified causes are not particularly dissimilar from other causes according to the theoretical similarity index (see Figure 1). The importance of this category seems to reflect more the general prevalence of these unspecified causes, especially for NCDs (Streatfield et al. 2014a and 2014b). Mostly associated with the presence of risk factors (e.g., history of heart disease, hypertension) and general symptoms (e.g., swollen feet or ankles, illness of long duration), these unspecified non-communicable causes could be indicative of chronic NCDs that are not often considered underlying causes (such as hypertension or diabetes), which would explain their importance in the results.

We tested our first approach with an individual-based process by comparing the theoretical similarity, based on InterVA's general definition of causes, to actual information reported by VAs in the field. We argue here that the high correlation between the individual-based index and the theoretical index shows that the similarity identified by the two indices remains very similar, despite a significant difference in the number of indicators considered. The theoretical approach seems to be robust to potentially scarce information reported in real data and can directly be applied to any dataset with causes of death assigned by InterVA or other algorithms based on a similar probability matrix.

[12] The tool selects a second cause of death only if the second-most-likely cause is associated with a probability greater than half the probability of the most probable cause (see Appendix 1).

However, our comparison of the individual-based and theoretical indices remains limited by the sample size of the detailed VAs. The individual-based index could be the object of further analysis to test the robustness of the theoretical index for individual associations of causes and to help determine a more robust threshold for the theoretical index when detailed VAs are not available. When detailed VAs are available, the individual-based index could potentially be more precise but would require further investigation with a larger dataset to select an appropriate threshold.

Finally, the calculation of these indices also results from a series of deliberate choices. Similarity can be evaluated according to different norm functions. We selected here the Euclidean norm, and the high correlation with the other tested norms shows that this choice has only a limited impact on the results (Appendix 3). We tested only norms that measure the distance between probability vectors in an additive way. In the Bayesian formula, however, the probabilities are multiplied with one another. We could take into account this multiplicative property by measuring the distance between the logarithm of the two vectors. We chose not to include this approach because we thought it would give too much weight to indicators associated with very small probabilities (i.e., a distance between 0.01 and 0.005) that are very unlikely to be reported if the cause is considered probable and would skew the measure of similarity in the theoretical index. Nevertheless, this remains a possibility that could be considered in future research, particularly when dealing with individual-based indices. Further refinements, such as the choice of indicators used in the calculation of the index, can also be made. As some questions depend on age and sex (such as those involving women of reproductive age and neonates), the inclusion of only age- and sex-relevant indicators can provide intermediary indices between the theoretical and individual-based ones. Additional analysis shows that these indices correlate strongly with the general theoretical index, especially for adults (0.95 or more). (See Appendix 3 and supplementary material.) The difference is only important in relation to neonates (less than a month old; Pearson coefficient of 0.70 for females and 0.75 for males), who were not included in this study.

Moreover, the selection of the threshold between co-occurring and competing causes is to some extent arbitrary given the information available. The sensitivity analysis performed shows that variation of the threshold does not substantially change the most frequent associations considered co-occurring causes or the distribution of co-occurring causes among groups (associations within NCDs or between infectious diseases and NCDs; see Appendix 4). To the best of our knowledge, the chosen threshold excludes associations that appear most prone to confusion among the most common associations in the INDEPTH dataset. However, this would need to be validated through systematic review by physicians or by a gold standard dataset for multiple causes.

Despite these limitations, this approach is, to the best of our knowledge, the first attempt to estimate multiple causes of death in LMICs, where the vast majority of deaths

occur. We propose an innovative method for identifying possible multimorbidity at the time of death that can be readily applied to other Bayesian algorithms. We also address a key challenge in VAs – the high risk of confusion between causes – in a novel way and provide an approach to assess this risk that could be used more generally to interpret VA results. We aim in this paper to make the most of the available information, while highlighting the constraints imposed, to inform further research and data collection. This first step will require further research to be fully substantiated, particularly through comparisons with medical data sources such as comorbidity diagnoses or medical certification of multiple causes.

## 5. Conclusion

Estimating multiple causes of death in LMICs should be a major concern if we are to better understand current epidemiologic trends and inform public health policy. In the context of an increasing cumulative burden of disease, this approach could highlight the interaction between infectious diseases and NCDs and contribute to a reassessment of the burden of disease.

In contexts where health data are lacking, the standardized method of collecting and automatically assigning causes through VA algorithms has enormous potential to increase knowledge of cause-specific mortality on a larger scale. This method is used at the local level in many HDSS sites and is increasingly being considered for the production of cause-of-death statistics at the national level (Sankoh and Byass 2014; de Savigny et al. 2017; Firth et al. 2021; Chandramohan et al. 2021; Niang et al. 2023). Our paper aims to leverage this potential by proposing an approach to identifying multiple causes of death through these algorithms. It advocates for the consideration of multimorbidity in the production and interpretation of cause-of-death data through VAs and provides elements for this ongoing discussion.

## 6. Acknowledgments

We thank the INDEPTH Network for the online data repository with free-access archival datasets from HDSS sites that are members of INDEPTH. We thank the Institut supérieur des sciences de la population (ISSP) for providing access to the Ouagadougou HDSS data. We also acknowledge the French Development Agency (AFD) and the École Normale Supérieur de Paris (ENS Paris) for their funding during the initiation of this project. We thank the anonymous reviewers for their important insights on earlier

versions of this paper. Finally, we deeply thank Laurie Hurwitz for her important contribution of wording suggestions and proofreading.

## References

Ahmed, A., Nahian, A., Mahabubur Rahman, M., Alam, N., Nahar, Q., Streatfield, P.K., Moinuddin Haider, M., and Rahman, M. (2023). Adult mortality trends in Matlab, Bangladesh: An analysis of cause-specific risks. *BMJ Open* 13(9): e065146. doi:10.1136/bmjopen-2022-065146.

Banerjee, A., Hurst, J., Fottrell, E., and Miranda, J.J. (2020). Multimorbidity: Not just for the West. *Global Heart* 15(1): 45. doi:10.5334/gh.835.

Barbieri, M., Désesquelles, A., Egidi, V., Demuru, E., Frova, L., Meslé, F., and Pappagallo, M. (2017). Obesity-related mortality in France, Italy, and the United States: A comparison using multiple cause-of-death analysis. *International Journal of Public Health.* 62(6): 623–629. doi:10.1007/s00038-017-0978-1.

Barr, A.L., Young, E.H., Smeeth, L., Newton, R., Seeley, J., Ripullone, K., Hird, T.R. et al. (2016). The need for an integrated approach for chronic disease research and care in Africa. *Global Health, Epidemiology and Genomics* 1: e19. doi:10.1017/gheg.2016.16.

Boutayeb, A. (2006). The double burden of communicable and non-communicable diseases in developing countries. *Transactions of the Royal Society of Tropical Medicine and Hygiene* 100(3): 191–199. doi:10.1016/j.trstmh.2005.07.021.

Byass, P., Chandramohan, D., Clark, S.J., D'Ambruoso, L., Fottrell, E., Graham, W.J., Herbst, A.J., Hodgson, A., Hounton, S., Kahn, K., Krishnan, A., Leitao, J., Odhiambo, F., Sankoh, O.A., and Tollman, S.M. (2012). Strengthening standardized interpretation of verbal autopsy data: The new InterVA-4 tool. *Global Health Action* 5: 19281. doi:10.3402/gha.v5i0.19281.

Byass, P., Huong, D.L., and Minh, H.V. (2003). A probabilistic approach to interpreting verbal autopsies: Methodology and preliminary validation in Vietnam. *Scandinavian Journal of Public Health* 31(62_suppl): 32–37. doi:10.1080/14034950310015086.

Byass, P., Hussain-Alkhateeb, L., D'Ambruoso, L., Clark, S., Davies, J., Fottrell, E., Bird, J., Kabudula, C., Tollman, S., Kahn, K., Schiöler, L., and Petzold, M. (2019). An integrated approach to processing WHO-2016 verbal autopsy data: The InterVA-5 model. *BMC Medicine* 17(1): 102. doi:10.1186/s12916-019-1333-6.

Chandramohan, D., Fottrell, E., Leitao, J., Nichols, E., Clark, S. J., Alsokhn, A., Munoz, D. C., AbouZahr, C., Di Pasquale, A., Robert, R., Choi, E., Baiden, F., Thomas, J., Lyatuu, I., Li, Z., Patrick Larbi-Debrah, P., Chu, Y., Cheburet, S., Sankoh, O.,

Badr, A. M., Ma Fat, D., Setel, P. R., and de Savigny, D. (2021). Estimating causes of death where there is no medical certification: Evolution and state of the art of verbal autopsy. *Global Health Action* 14(sup1): 1982486. doi:10.1080/16549716.2021.1982486.

Chu, Y., Marston, M., Dube, A., Festo, C., Geubbels, E., Gregson, S., Herbst, K. et al. (2024). Temporal changes in cause of death among adolescents and adults in six countries in Eastern and Southern Africa in 1995–2019: A multi-country surveillance study of verbal autopsy data. *The Lancet Global Health* 12(8): e1278–1287. doi:10.1016/S2214-109X(24)00171-2.

Ciccacci, F., Orlando, S., Majid, N., and Marazzi, C. (2020). Epidemiological transition and double burden of diseases in low-income countries: The case of Mozambique. *Pan African Medical Journal* 37(49). doi:10.11604/pamj.2020.37.49.23310.

Clark, S.J., McCormick, T., Li, Z. and Wakefield, J. (2015). *InSilicoVA: A method to automate cause-of-death assignment for verbal autopsy*. arXiv: Other Statistics.

De Savigny, D., Riley, I., Chandramohan, D., Odhiambo, F., Nichols, E., Notzon, S., AbouZahr, C., Mitra, R., Cobos Muñoz, D., Firth, S., Maire, N., Sankoh, O., Bronson, G., Setel, P., Byass, P., Jakob, R., Boerma, T., and Lopez, A.D. (2017). Integrating community-based verbal autopsy into civil registration and vital statistics (CRVS): System-level considerations. *Global Health Action* 10: 1272882. http://dx.doi.org/10.1080/16549716.2017.1272882.

Désesquelles, A., Demuru, E., Pappagallo, M., Frova, L., Meslé, F., and Egidi, V. (2015). After the epidemiologic transition: A reassessment of mortality from infectious diseases among over-65s in France and Italy. *International Journal of Public Health* 60(8): 961–967. doi:10.1007/s00038-015-0704-9.

Désesquelles, A., Gamboni, A., Demuru, E., and the MultiCause Network (2016). We only die once... but from how many causes? *Population et Sociétés* 534(6): 1–4. doi:10.3917/popsoc.534.0001.

Désesquelles, A., Salvatore, M.A., Frova, L., Pace, M., Pappagallo, M., Meslé, F., and Egidi, V. (2010). Revisiting the mortality of France and Italy with the multiple cause-of-death approach. *Demographic Research* 23(28): 771–806. doi:10.4054/DemRes.2010.23.28.

Duthé, G., Vanhoutte, L., Ouedraogo, S., and Pison, G. (2024). Health transition and population aging: Challenges for the Global South. In: Lemaître, J.-F. and Pavard, S. (eds.). *Aging and longevity: A multi-disciplinary approach*. Cambridge University Press: 253–274. doi:10.1017/9781009007245.014.

Fantahun, M., Fottrell, E., Berhane, Y., Wall, S., Högberg, U., and Byass, P. (2006). Assessing a new approach to verbal autopsy interpretation in a rural Ethiopian community: The InterVA model. *Bulletin of the World Health Organization* 84: 204–210. doi:10.2471/BLT.05.028712.

Firth, S.M., Hart, J.D., Reeve, M., Li, H., Mikkelsen, L., Sarmiento, D.C., Bo, K.S., Kwa, V., Qi, J.-L., Yin, P., Segarra, A., Riley, I., and Joshi, R. (2021). Integrating community-based verbal autopsy into civil registration and vital statistics: Lessons learnt from five countries. *BMJ Global Health* 6(11): e006760. doi:10.1136/bmjgh-2021-006760.

Foreman, K.J., Lozano, R., Lopez, A.D., and Murray, C.J. (2012). Modeling causes of death: An integrated approach using CODEm. *Population Health Metrics* 10(1). doi:10.1186/1478-7954-10-1.

Fortin, M., Stewart, M., Poitras, M.-E., Almirall, J., and Maddocks, H. (2012). A systematic review of prevalence studies on multimorbidity: Toward a more uniform methodology. *The Annals of Family Medicine* 10(2): 142–151. doi:10.1370/afm.1337.

GBD 2019 Diseases and Injuries Collaborators (2020). Global burden of 369 diseases and injuries in 204 countries and territories, 1990–2019: A systematic analysis for the Global Burden of Disease Study 2019. *The Lancet* 396 (10258): 1204–1222. doi:10.1016/S0140-6736(20)30925-9.

INDEPTH Network (2014). *INDEPTH Network Cause-Specific Mortality – Release 2014*. Provided by the INDEPTH Network Data Repository. www.indepth-network.org.10.7796/INDEPTH.GH003.COD2014.v1.

Källander, K., Nsungwa-Sabiiti, J., and Peterson, S. (2004). Symptom overlap for malaria and pneumonia-policy implications for home management strategies. *Acta Tropica* 90(2): 211–214. doi:10.1016/j.actatropica.2003.11.013.

Kolčić, I. (2012). Double burden of malnutrition: A silent driver of double burden of disease in low- and middle-income countries. *Journal of Global Health* 2(2). doi:10.7189/jogh.02.020303.

Kushitor, M.K. and Boatemaa, S. (2018). The double burden of disease and the challenge of health access: Evidence from access, bottlenecks, cost and equity facility survey in Ghana. *PLoS ONE* 13(3): e0194677. doi:10.1371/journal.pone.0194677.

Marshall, S.J. (2004). Developing countries face double burden of disease. *Bulletin of the WHO* 82:556

Martini, J. and Figg, A. (2010). L'émergence du diabète de type 2 en tant que problème de santé publique au Mali. In: Kerouedan, D. (ed.). *Santé internationale: Les enjeux de santé au sud*. Paris: Presses de Sciences Po: 105–118. doi:10.3917/scpo.kerou.2011.01.105.

McCormick, T.H., Li, Z.R., Calvert, C., Crampin, A.C., Kahn, K., and Clark, S.J. (2016). Probabilistic cause-of-death assignment using verbal autopsies. *Journal of the American Statistical Association* 111(515): 1036–1049. doi:10.1080/01621459.2016.1152191.

Mikkelsen, L., Phillips, D.E., AbouZahr, C., Setel, P.W., De Savigny, D., Lozano, R., and Lopez, A.D. (2015). A global assessment of civil registration and vital statistics systems: monitoring data quality and progress. *The Lancet* 386(10001): 1395–1406. doi:10.1016/S0140-6736(15)60171-4.

Multiple Causes-of-Death Network (2007). MultiCause [electronic resource]. Paris: Institut national d'études démographiques (INED). https://mcod.web.ined.fr/wiki/Accueil.

Niang, K., Fall, A., Ndiaye, S., Sarr, M., Ba, K., and Masquelier, B. (2023). Enhancing the value of death registration with verbal autopsy data: A pilot study in the Senegalese urban population in 2019. *Archives of Public Health* 81(45). doi:10.1186/s13690-023-01067-6.

Oksuzyan, A., Brønnum-Hansen, H., and Jeune, B. (2010). Gender gap in health expectancy. *European Journal of Ageing* 7(4): 213–218. doi:10.1007/s10433-010-0170-4.

Omran, A.R. (1971). The epidemiologic transition: A theory of the epidemiology of population change. *The Milbank Quarterly* 83(4): 731–757. doi:10.1111/j.1468-0009.2005.00398.x.

Osman, S. and Byass, P. (2014). Time for civil registration with verbal autopsy. *The Lancet Global Health* 2(12): e69394. doi:10.1016/S2214-109X(14)70340-7.

Pati, S., Swain, S., Hussain, M.A., van den Akker, M., Metsemakers, J., Knottnerus, J.A., and Salisbury, C. (2015). Prevalence and outcomes of multimorbidity in South Asia: A systematic review. *BMJ Open* 5(10): e007235. doi:10.1136/bmjopen-2014-007235.

Remais, J.V., Zeng, G., Li, G., Tian, L., and Engelgau, M.M. (2013). Convergence of non-communicable and infectious diseases in low- and middle-income countries. *International Journal of Epidemiology* 42(1): 221–227. doi:10.1093/ije/dys135.

Sankoh, O. and Byass, P. (2014). Time for civil registration with verbal autopsy. *The Lancet Global Health* 2(12): e693–e694. doi:10.1016/S2214-109X(14)70340-7.

Skou, S.T., Mair, F.S., Fortin, M., Guthrie, B., Nunes, B.P., Miranda, J.J., Boyd Pati, C.M.S., Mtenga, S., and Smith, S.M. (2022). Multimorbidity. *Nature Reviews Disease Primers* 8(1): 1-22. doi:10.1038/s41572-022-00376-4.

Star, S.L. and Bowker, G.G. (1999). *Sorting things out: Classification and its consequences*. Cambridge, MA: MIT Press.

Streatfield, P.K., Khan, W.A., Bhuiya, A., Alam, N., Sié, A., Soura, A.B., Bonfoh, B., Ngoran, E.K., Weldearegawi, B., Jasseh, M., Oduro, A., Gyapong, M., Kant, S., Juvekar, S., Wilopo, S., Williams, T.N., Odhiambo, F.O., Beguy, D., Ezeh, A., Kyobutungi, C., Crampin, A., Delaunay, V., Tollman, S.M., Herbst, K., Chuc, N.T.K., Sankoh, O.A., Tanner, M., and Byass, P. (2014a). Cause-specific mortality in Africa and Asia: Evidence from INDEPTH health and demographic surveillance system sites. *Global Health Action* 7(s7): 25362. doi:10.3402/gha.v7.25362.

Streatfield, P.K., Khan, W.A., Bhuiya, A., Hanifi, S.M.A., Alam, N., Bagagnan, C.H., Sié, A., Zabré, P., Lankoandé, B., Rossier, C., Soura, A.B., Bonfoh, B., Kone, S., Ngoran, E.K., Utzinger, J., Haile, F., Melaku, Y.A., Weldearegawi, B., Gomez, P., Jasseh, M., Ansah, P., Debpuur, C., Oduro, A., Wak, G., Adjei, A., Gyapong, M., Sarpong, D., Kant, S., Misra, P., Rai, S.K., Juvekar, S., Lele, P., Bauni, E., Mochamah, G., Ndila, C., Williams, T.N., Laserson, K.F., Nyaguara, A., Odhiambo, F.O., Phillips-Howard, P., Ezeh, A., Kyobutungi, C., Oti, S., Crampin, A., Nyirenda, M., Price, A., Delaunay, V., Diallo, A., Douillot, L., Sokhna, C., Gómez-Olivé, F.X., Kahn, K., Tollman, S.M., Herbst, K., Mossong, J., Chuc, N.T.K., Bangha, M., Sankoh, O.A., and Byass, P. (2014b). Adult non-communicable disease mortality in Africa and Asia: Evidence from INDEPTH Health and Demographic Surveillance System sites. *Global Health Action* 7: 12. doi:10.3402/gha.v7.25365.

Tabutin, D. and Masquelier, B. (2017). Mortality inequalities and trends in low- and middle-income countries, 1990–2015. *Population* 2: 221–296. doi:10.3917/popu.1702.0225.

United Nations (2015). *Transforming our world: The 2030 agenda for sustainable development*. United Nations (A/RES/70/1).

United Nations (2022). *World population prospects 2022*. New York: Population Division, Department of Economic and Social Affairs, United Nations. doi:10.18356/9789210014380.

Wade, A.N., Payne, C.F., Berkman, L., Chang, A., Gómez-Olivé, F.X., Kabudula, C., Kahn, K., Salomon, J.A., Tollman, S., Witham, M., and Davies, J. (2021). Multimorbidity and mortality in an older, rural black South African population cohort with high prevalence of HIV findings from the HAALSI Study. *BMJ Open* 11(9): e047777. doi:10.1136/bmjopen-2020-047777.

WHO (2019a). Volume 2: Instruction manual. In: *International classification of diseases tenth revision (ICD-10), Sixth Edition*. Geneva: World Health Organization.

WHO (2019b). Completeness of death data (%). Geneva: World Health Organization. https://www.who.int/data/gho/data/indicators/indicator-details/GHO/completeness-of-cause-of-death-data-(-).

WHO (2022). *Non-communicable diseases. Key facts.* Geneva: World Health Organization. https://www.who.int/news-room/fact-sheets/detail/noncommunicable-diseases.

WHO (2024). Verbal autopsy standards: ascertaining and attributing causes of death tool. Geneva: World Health Organization. https://www.who.int/standards/classifications/other-classifications/verbal-autopsy-standards-ascertaining-and-attributing-causes-of-death-tool.

Wong, E.B., Olivier, S., Gunda, R., Koole, O., Surujdeen, A., Gareta, D., Munatsi, D., Modise, Tshwaraganang H., Dreyer, J., Nxumalo, S., Smit, T.K., Ording-Jespersen, G., Mpofana, I.B., Khan, K., Sikhosana, Z.E.L., Moodley, S., Shen, Y.-J., Khoza, T., Mhlongo, N., Bucibo, Sanah, Nyamande, K., Baisley, K.J., Cuadros, D., Tanser, F., Grant, A.D., Herbst, K., Seeley, J., Hanekom, W.A., Ndung'u, T., Siedner, M.J., Pillay, D., Wong, E.B., Olivier, S., Gunda, R., Koole, O., Surujdeen, A., Gareta, D., Munatsi, D., Modise, Tswaraganang H., Dreyer, J., Nxumalo, S., Smit, T.K., Ording-Jespersen, G., Mpofana, I.B., Khan, K., Sikhosana, Z.E.L., Moodley, S., Shen, Y.-J., Khoza, T., Mhlongo, N., Bucibo, Sana, Nyamande, K., Baisley, K.J., Cuadros, D., Tanser, F., Grant, A.D., Herbst, K., Seeley, J., Hanekom, W.A., Ndung'u, T., Siedner, M.J., Pillay, D., Suleman, M., Kalideen, J., Jackpersad, R., Moropane, K., Mfolo, B., Malomane, K., Khumalo, H., Buthelezi, N., Mbonambi, N., Ngubane, H., Simelane, T., Buthelezi, K., Ntuli, S., Zondi, N., Nene, S., Ndlovu, B., Ntimbane, T., Mbuyisa, M., Mkhize, X., Sibiya, M., Ntombela, N., Dlamini, M., Chonco, H., Dlamini, H., Mlambo, D., Mzimela, N., Buthelezi, Z., Mthembu, Z., Bhengu, T., Mtehmbu, S., Mthethwa, P., Mbatha, Z., Mthembu, W.P., Mkhwanazi, A., Sikhali, M., Mkhwanazi, P., Mkhwanazi, N., Myeni, R., Mfeka, F., Gumede, H., Mfeka, N., Zungu, A., Gumede, H., Mfekayi, N., Zulu, S., Buthelezi, M., Mkhwanazi, Senzeni, Dube, M., Matthews, P., Dlamini, S., Kambonde, H., Mthembu, L., Mchunu, S., Gumbi,

S., Madolo, T., Nkosi, T., Mkhwanazi, Sibusiso, Nsibande, Simbusio, Steto, M., Mhlongo, S., Vellem, V., Tshivase, P., Kwinda, J., Magwaza, B., Nsibande, Siyabonga, Mthombeni, S., Mthembu, S.C., Rapulana, A., Cousins, J., Zondi, T., Padayachi, N., Mabetlela, F., Ntshangase, S., Luthuli, N., Ngcobo, S., Brien, K., Ndlela, S., Ngema, N., Ntshakala, N., Singh, A., Singh, R., Pillay, L., Chetty, K., Govender, A., Ramkalawon, P., Mabaso, N., Perumal, K., Makhari, S., Khuluse, N., Zitha, N., Khati, H., Mofokeng, M., Majozi, N., Gqaleni, N., Keal, H., Ngcobo, P., Criticos, C., Zondo, R., Kalyan, D., Mavimbela, C., Ramnanan, A., and Harilall, S. (2021). Convergence of infectious and non-communicable disease epidemics in rural South Africa: A cross-sectional, population-based multimorbidity study. *The Lancet Global Health* 9(7): e967–e976. doi:10.1016/S2214-109X(21)00176-5.

# Appendices

## Appendix 1: Presentation of the InterVA algorithm

The InterVA model is based on Bayes's formula for conditional probabilities. It can be presented as follows:

Let $S = \{s_j\}$, with $j$ ranging from 1 to 245, the set of 245 indicators, the symptoms, and information reported by the final caregiver in the verbal autopsy. All indicators are binary variables. If the symptom or characteristic $j$ characterizes the deceased, $s_j = 1$; otherwise $s_j = 0$[13]. Let $C_i$, with $i$ ranging from 1 to 60, be the set of possible causes of death according to InterVA. Let J be the set of reported indicators – i.e., $j \in J \Leftrightarrow s_j = 1$. $P(C_i|S)$. The probability of cause $i$ given the set of information S is:

$$P(C_i|S) = \frac{P(C_i)\cdot P(S|C_i)}{P(S)} = \frac{P(C_i).\prod_{j\in J} P(s_j|C_i)}{\sum_{k=1}^{60} P(C_k)\prod_{j\in J} P(s_j|C_k)}. \quad \text{(A.1)}$$

This model is based on the simplifying assumption that all symptoms are independent. (See McCormick et al. 2016 for discussion and limitations of the model's assumptions.) Only positively reported indicators are used to determine the cause of death.

For a given death, InterVA calculates a probability for each cause of the predefined classification (Table A-1.1). If the probability of the most probable cause is less than 0.4, the death is classified as indeterminate with a probability of 1. Otherwise, the tool selects the most likely cause and then up to two additional causes if their probability is more than half the probability of the previous most probable cause (i.e., necessarily greater than 0.2). The sum of the probabilities of the selected causes rarely is equal to 1, and the difference is considered indeterminate to account for uncertainty. (For more details, see Byass et al. 2012; McCormick et al. 2016.)

In the InterVA algorithm, each cause is defined by a vector of probabilities, one for each symptom or characteristic (age, sex, treatments, and so on) potentially reported in the VA. In this vector, each of these indicators is associated to a probability of being reported given that cause of death. The matrix of conditional probabilities *probbase* = $(P(s_j|C_i))_{i,j}$ was elaborated by a committee of experts (Byass et al. 2012). They assigned letter grades to each probability, which were then translated into numerical values by the algorithm (see Table A-1.2). This probability matrix does not vary according to the time

[13] This means that characteristics covering more than one category correspond to several binary indicators. For example, six indicators correspond to the six age groups categorized by InterVA. All numeric indicators (mainly indicators of symptom duration) are dichotomized (e.g., "fever of any kind," "fever lasting more than two weeks or more," "fever lasting less than two weeks").

or geographical context of death, with the exception of two diseases that are subject to prior specification: malaria and AIDS. For these two diseases, the reference population must be characterized according to a higher or lower prevalence of infection, which changes the a priori probability of their occurrence $P(C_i)$. To calculate the theoretical similarity index, we used the default specifications of very low malaria and very low HIV. We performed a sensitivity analysis with a different specification and showed it had very limited impact on the results (Appendix 3). For the individual-based index, the specification appropriate to the Ouagadougou HDSS (high malaria, low HIV) was used.

**Table A-1.1: Classification of causes of death, InterVA-4**

| Causes of death (n=61) | ICD10 correspondence |
|---|---|
| Sepsis | A40-A41 |
| Acute respiratory infections (pneumonia) | J00-J22 |
| HIV/AIDS | B20-B24 |
| Diarrhoea | A00-A09 |
| Malaria | B50-54 |
| Measles | B05 |
| Meningitis, encephalitis | A39; G00-G05 |
| Tetanus | A33-A35 |
| Pulmonary tuberculosis | A15-A16 |
| Pertussis | A37 |
| Haemorrhagic fever | A90-A99 |
| Other infectious diseases | A17-A19; A20-A38; A42-A89; B00-B19; B25-49; B55-B99 |
| Oral neoplasms | C00-C06 |
| Digestive neoplasms | C15-C26 |
| Respiratory neoplasms | C30-C39 |
| Breast neoplasms | C50 |
| Reproductive neoplasms (male and female) | C51-C58; C60-C63 |
| Other and unspecified neoplasms | C07-C14; C40-C49; C60-D48 |
| Severe anaemia | D50-D64 |
| Severe malnutrition | E40-E46 |
| Diabetes | E10-E14 |
| Acute heart disease | I20-I25 |
| Sickle cell disease with crisis | D57 |
| Stroke | I60-I69 |
| Other cardiac diseases | 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 and unspecified non-communicable diseases | D55-D89; E00-E07; E15-E35; E50-E90; F00-F99; G06G09; GG10-G37; G50-G99; H00-H95; J30-J39; J47-J99; K00-K31; K35-K38; K40-K93; L00-L99; M00-M99; N00N16; N20-N99; R00-R09; R11-R94 |

**Table A-1.1: (Continued)**

| Causes of death (n=61) | ICD10 correspondence |
|---|---|
| Congenital malformation | Q00-Q99 |
| Prematurity | P05-P07 |
| Birth asphyxia | P20-P22 |
| Neonatal pneumonia | P23-P25 |
| Neonatal sepsis | P36 |
| Other neonatal causes | P00-P04; P08-P15; P26-P35; P37-P94; P96 |
| Fresh stillbirth | P95 |
| Macerated stillbirth | P95 |
| Road traffic accident | V01-V89 |
| Other transport accident | V90-V99 |
| Accidental fall | W00-W19 |
| Accidental drowning and submersion | W65-W74 |
| Accidental exposure to smoke, fire & flame | X00-X19 |
| Contact with venomous plant/animal | X20-X29 |
| Exposure to forces of nature | X30-X39 |
| Accidental poisoning and noxious substances | X40-X49 |
| Intentional self-harm | X60-X84 |
| Assault | X85-Y09 |
| Other external causes | S00-T99; W20-W64; W75-W99; X50-X59; Y10-Y98 |
| 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 |
| Pregnancy-related anaemia | O99.0 |
| Ruptured uterus | O71 |
| Other maternal causes | O01-O02; O20-O45; O47-O62; O68-O70; O73-O84; O86-O99 |
| Cause of death unknown | R95-R99 |

## Table A-1.2: Correspondence between letter grades and numerical values for InterVA-4 (from Byass 2012)

| Interpretation | Letter | Value |
|---|---|---|
| Always | 1 | 1.0 |
| Almost always | A+ | 0.8 |
| Common | A | 0.5 |
| | A- | 0.2 |
| Often | B+ | 0.1 |
| | B | 0.05 |
| | B- | 0.02 |
| Unusual | C+ | 0.01 |
| | c | 0.005 |
| | c- | 0.002 |
| Rare | D+ | 0.001 |
| | D | 0.0005 |
| | D- | 0.0001 |
| Hardly ever | E | 0.00001 |
| Never | N | 0 |

## Table A-1.3: Distribution of probability of the most likely cause, as determined by InterVA-4

| Probability | Frequency | % | Cumulated % |
|---|---|---|---|
| [0.0:0.4[ or indeterminate | 3,403 | 4.70 | 4.23 |
| [0.4:0.6[ | 8,285 | 11.45 | 14.54 |
| [0.6:0.8[ | 9,700 | 13.41 | 26.60 |
| [0.8:1 .0[ | 24,431 | 33.78 | 56.98 |
| 1 | 26,511 | 36.65 | 89.95 |
| Total | 80,409 | 100.00 | 100.00 |
| | **Median = 0.97** | | |

*Scope*: 72,330 adult deaths in 22 HDSS sites (1992–2012).
*Source*: INDEPTH Network 2014.

## Appendix 2: Deaths attributed more than one cause by site, age and sex

### Table A-2: Deaths with more than one probable cause attributed by InterVA-4, by site, sex, and age

| | Frequency | % |
|---|---|---|
| **By site** | | |
| Burkina Faso, Nouna | 318 | 10.3 |
| Burkina Faso, Ouagadougou | 51 | 11.5 |
| Cöte d'Ivoire, Taabo | 51 | 13.6 |
| Ethiopia, Kilite Awlaelo | 32 | 9.8 |
| Ghana, Dodowa | 290 | 11.4 |
| Ghana, Navrongo | 1,201 | 14.4 |
| The Gambia, Farafenni | 191 | 11.9 |
| India, Ballabgarh | 200 | 11.9 |
| India, Vadu | 70 | 13.0 |
| Indonesia, Purworejo | 77 | 9.9 |
| Kenya, Kilif | 331 | 10.0 |
| Kenya, Kisumu | 1,398 | 11.7 |
| Kenya, Nairobi | 266 | 11.2 |
| Malawi, Karonga | 127 | 9.3 |
| Senegal, Bandafassi | 165 | 15.9 |
| South Africa, Africa Centre | 509 | 5.6 |
| South Africa, Agincourt | 994 | 10.8 |
| Vietnam, Filabavi | 77 | 11.1 |
| Bangladesh, AMK | 254 | 9.6 |
| Bangladesh, Bandarban | 28 | 12.1 |
| Bangladesh, Chakaria | 88 | 9.6 |
| Bangladesh, Matlab | 1,016 | 10.4 |
| **By demographic characteristics** | | |
| **Sex** | | |
| Female | 3,908 | 11.1 |
| Male | 3,826 | 10.3 |
| **Age group** | | |
| 15–49 years | 2,515 | 9.0 |
| 50–64 years | 1,598 | 11.1 |
| 65+ years | 3,621 | 12.1 |

*Scope*: 72,330 VAs of adults in 22 sites of the INDEPTH Network, 1992–2012.
*Source*: INDEPTH Network 2014.

## Appendix 3: The theoretical similarity indices: distributions and correlations[14]

We defined and tested four different similarity indices. Formally, let *A* and *B* be the vectors of probabilities associated with cause *a* and cause *b*, respectively. We define the following similarity indices between *a* and *b*, ranging from 0 (exactly similar) to 1 (as different as possible):

- Norm indices: $I_{a,b} = I_{b,a} = \frac{\|A-B\|}{\|A+B\|}$, using both the Euclidean norm and the absolute norm
- Uniform normalization: Let *N* be the number of indicators used to construct the index:
  $I_{a,b} = I_{b,a} = \frac{\|A-B\|}{N}$, using the Euclidean norm
- The normalized scalar product, based on the Cauchy-Schwartz inequality:[15]
  $I_{a,b} = I_{b,a} = 1 - \frac{\langle A,B\rangle}{\|A\|\|B\|}$ , using the Euclidean norm

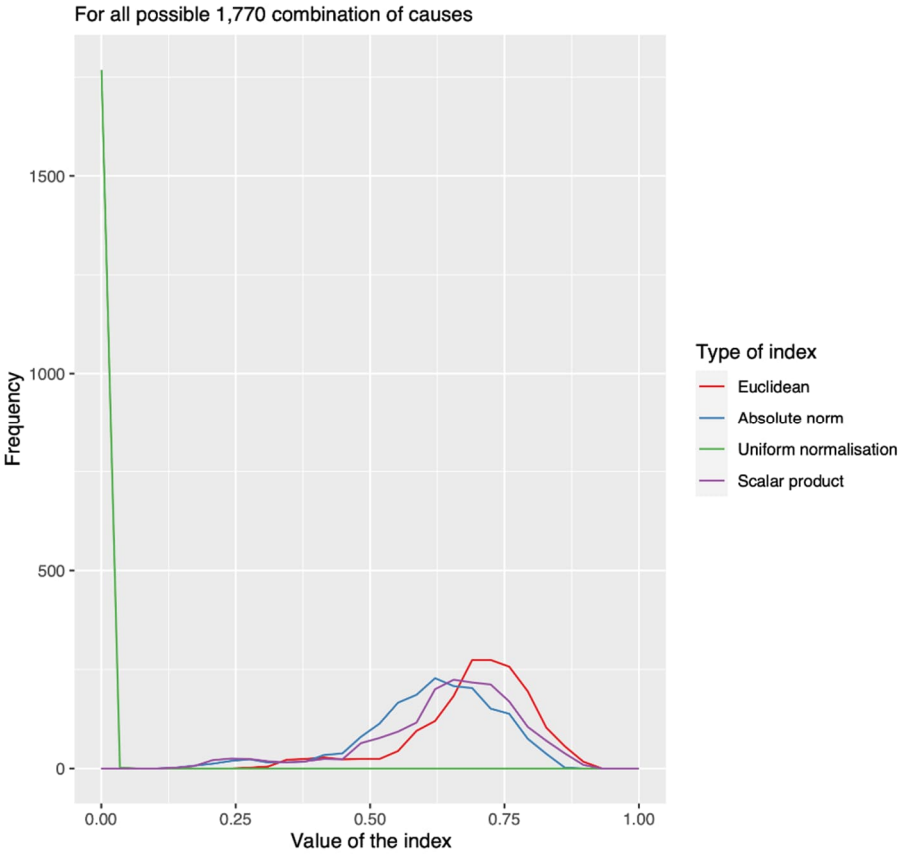
An interactive visualization of these different theoretical similarity indices is available as supplementary material.[16]

[14] The code to replicate those results is available here: https://www.demographic-research.org/volumes/vol52/8/files/52-8%20VACauseSimilarity_Index.zip .

[15] It states that for all *A* and *B* vectors of real numbers, $|\langle A,B\rangle| \leqslant \|A\|\|B\|$, where < . > is the scalar product and ||.|| is the norm associated with that scalar product (here the Euclidean norm).

[16] See VACauseSimilarity Shiny app: https://www.demographic-research.org/volumes/vol52/8/interactive/Figure%201%20%E2%80%93%20Similarity%20heatmap/default.htm.

## Figure A-3.1: Distribution of the similarity indices



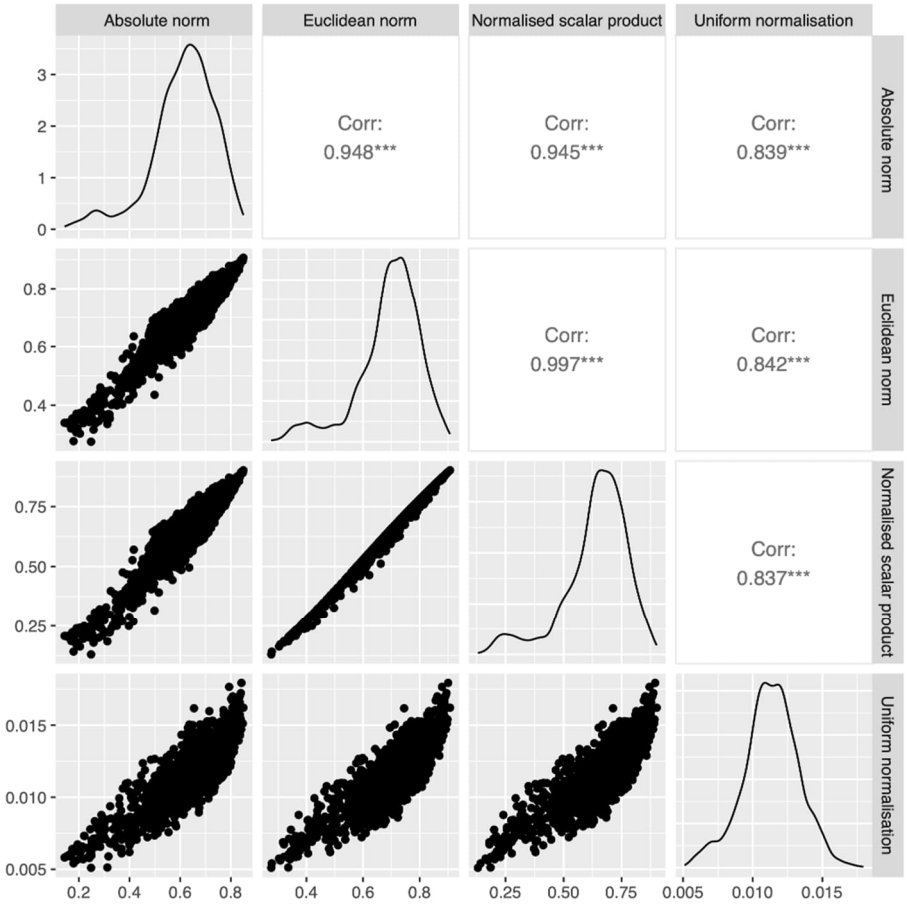

*Note*: Computed from the default probability matrix of InterVA (probbase; malaria: VL; HIV: VL).

## Table A-3.1: Distribution statistics of the indices

| Index | Min | Mean | Q3 | Max | Standard deviation | Coefficient of variation (SD/mean) |
|---|---|---|---|---|---|---|
| Euclidean | 0.28 | 0.69 | 0.65 | 0.91 | 0.11 | 0.16 |
| Absolute norm | 0.14 | 0.61 | 0.55 | 0.85 | 0.13 | 0.21 |
| Uniformly normalised | 0.01 | 0.06 | 0.05 | 0.1 1 | 0.02 | 0.29 |
| Normalised scalar product | 0.13 | 0.64 | 0.58 | 0.90 | 0.14 | 0.22 |

*Source*: Computed from the default probability matrix of InterVA-4 (probbase; malaria: VL; HIV: VL).

**Figure A-3.2: Correlation between similarity indices**



We explored possible refinements of the index using the individual information available in the INDEPTH dataset: HIV and malaria prevalence, age, and sex. For adults, the differences are minimal. For simplicity, we used the general index.

**Specific indices for the prevalence of HIV and malaria**

As seen in Appendix 1, InterVA-4 requires that an a priori level of malaria and HIV prevalence be specified for cause-of-death estimation. This changes the a priori prevalence probability of the cause in the probability matrix accordingly. In this paper, we used the default setting of very low malaria and HIV prevalence (VL-VL) to compute the theoretical indices.

We conducted a sensitivity analysis to investigate the impact of other prevalence levels of HIV and malaria on the indices, focusing on associations with HIV-related deaths or malaria, which are the only indices affected by this change. The difference is extremely small: Pearson's correlation coefficient between the indices with the most disparate setting (high malaria and HIV prevalence) and the default setting (VL-VL) is 0.9999847.

For the individual index, we adjusted malaria and HIV prevalence in keeping with the settings appropriate for Ouagadougou (high malaria, low HIV).

**Age- and sex-specific indices**[17]

Certain indicators given in the VA are age- or sex-specific. The conditional probability matrix of InterVA-4 specifies which indicators are not asked for certain age or sex groups,[18] so that the index can be calculated using only the indicators asked for these characteristics. We can also exclude causes that are considered impossible for a given age or sex (with an associated probability of 0) by setting the indicator to 1 by default.

We define age- and sex-specific indices as follows: Given $P(s_j|a)$, the a priori probability of presenting the indicator *j* for a given cause is *a*. Let J be the set of possibly reported indicators for a specific age group at death and sex *k*; let $s_{sex}$ be the indicator associated with sex; and let $s_{age}$ be the indicator associated with age at death.

The age- and sex-specific index between cause *a* and cause *b* is:

[17] These indices can also be visualized with the interactive online supplementary material (VACauseSimilarity Shiny app): https://www.demographic-research.org/volumes/vol52/8/interactive/Figure%201%20%E2%80%93%20Similarity%20heatmap/default.htm.

[18] To access the detailed matrix, see the material shared on Github.

$$I_{a,b|k} = I_{b,a|k} = \begin{cases} 1 \ if \ P(s_{sex}|a) \ or \ P(s_{age}|a) \ or \ P(s_{sex}|b) \ or \ P(s_{age}|b) = 0 \\ else \ \frac{\sqrt{\sum_{i \in J}(P(s_i|a) - P(s_i|b))^2}}{\sqrt{\sum_{i \in J}(P(s_i|a) + P(s_i|b))^2}} \end{cases} . \quad (2)$$

Interactive heat maps of these indices can be accessed in the supplementary materials.

Despite the important reduction in the number of relevant indicators used to calculate the age- and sex-specific indices (140 indicators on average compared to 245), the correlation between these indicators and the general Euclidean index is very high, especially for adults (0.95 or more for males and females aged 15+). This correlation is lower for children, especially neonates (under the age of 11 months), probably due to the specificity of the indicators and causes of death in these age groups.

Considering the very high correlation for the age groups concerned by this study (15+ years), we concluded that these refinements are not necessary in the context of this article but could be interesting variants for further developments.

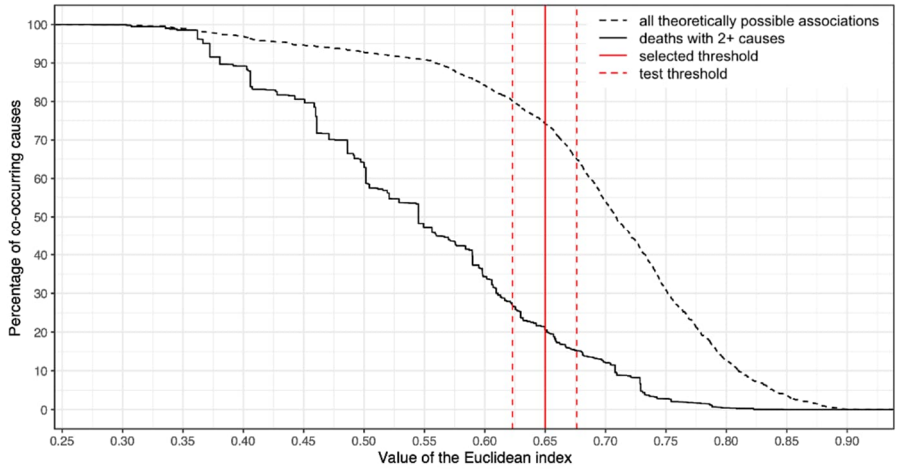
**Table A-3.2: Number of relevant indicators per sex and age group and correlation of their specific index with the theoretical index**

| Sex | Age group | Number of indicators | Correlation* |
|---|---|---|---|
| Female | 65+ years | **141** | **0.947** |
| | 50–64 years | **142** | **0.964** |
| | 15–49 years | **181** | **0.946** |
| | 5–15 years | **174** | **0.965** |
| | 1–4 years | **134** | **0.92** |
| | 1–11 months | **128** | **0.899** |
| | < 1 month | **123** | **0.707** |
| Male | 65+ years | **138** | **0.959** |
| | 50–64 years | **137** | **0.952** |
| | 15–49 years | **138** | **0.936** |
| | 5–15 years | **134** | **0.927** |
| | 1–4 years | **134** | **0.924** |
| | 1–11 months | **128** | **0.902** |
| | < 1 month | **123** | **0.741** |

* Correlation between the theoretical and age- and sex-specific indices is calculated excluding associations with causes considered impossible for the given age group and sex, as they could not be attributed by InterVA.
*Source*: Our calculations from the InterVA-4 probability matrix (VL-VL).

# Appendix 4: Threshold sensitivity analysis

**Figure A-4.1: Percentage of co-occurring causes according to a selected threshold**




*Note:* Interpretation: 75% of all possible cause associations have an associated theoretical similarity index greater than 0.65, while 20% of deaths attributed more than one cause are associated with an index greater than 0.65.
*Source*: INDEPTH Network (2014) and our calculation of the theoretical similarity index using the InterVA-4 probability matrix (malaria: VL; HIV: VL).

The selected threshold value corresponds to the 25th percentile ($p_{25} \approx 0.65$) of the distribution of all possible associations. To test its sensitivity, we present the results with a slightly less stringent threshold (20th percentile; $p_{20} \approx 0.62$) and a more restrictive threshold (35th percentile; $p_{35} \approx 0.68$) (Table A-4.1).

The proportion of causes considered to be co-occurring varies substantially depending on the threshold chosen, ranging from 27% to 16% of all associated causes. However, the distribution of co-occurring causes by group is relatively stable. The proportion of associations between non-communicable diseases and communicable, maternal, and neonatal causes remains the most frequent (56% for $p_{15,}$ 61% for 0.65, and 65% for $p_{35}$). This is consistent with the overall result of the study, which suggests a possible double burden of disease at the individual level.

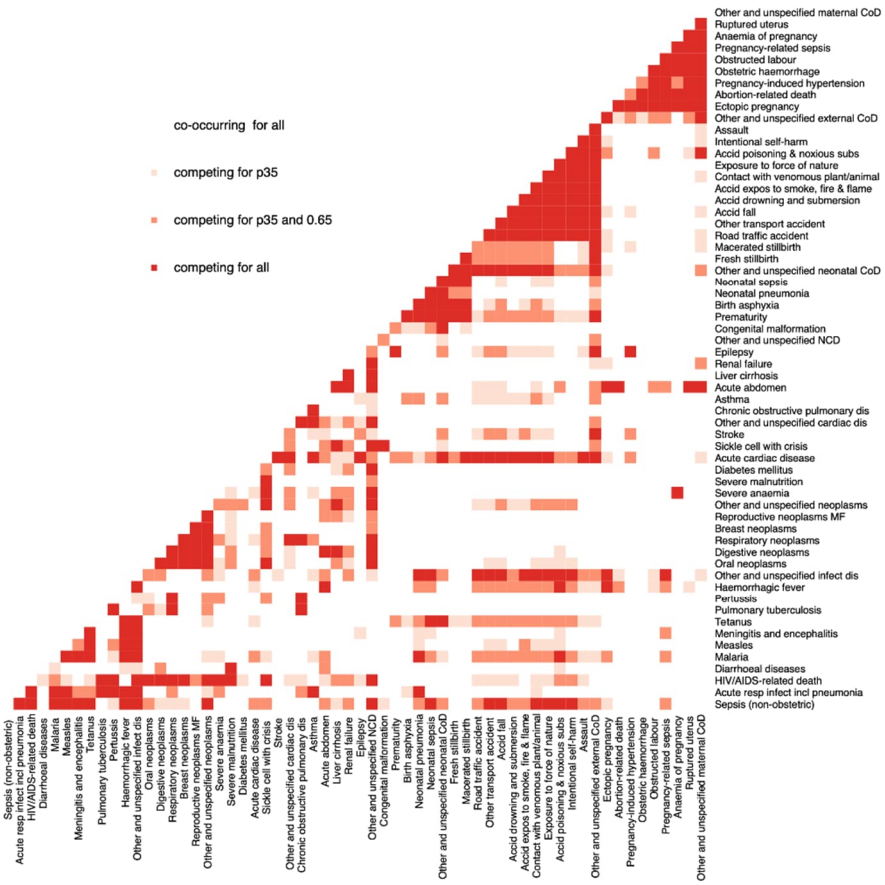
## Table A-4.1: Co-occurring causes according to selected threshold

| | | | | Threshold | | |
|---|---|---|---|---|---|---|
| | | | | **$p_{20}$** | **0.65** | **$p_{35}$** |
| Associations considered co-occurring – % (n) | | | | 26.7 (2,064) | 20.6 (1,591) | 16.9 (1,305) |
| **Most frequent associations of causes (n ≥ 30 )*** | | | | **Co-occurring associations** | | |
| # | Cause B | Cause A | n | | | |
| 1 | Diabetes mellitus | Stroke | 148 | X | X | X |
| 2 | Acute respiratory infection including pneumonia | Other and unspecified cardiac diseases | 141 | X | X | X |
| 3 | Pulmonary tuberculosis | Other and unspecified cardiac diseases | 123 | X | X | X |
| 4 | HIV/AIDS-related death | Intentional self-harm | 117 | X | | |
| 5 | Diabetes mellitus | Other and unspecified cardiac diseases | 57 | X | | |
| 6 | Acute respiratory infection including pneumonia | Respiratory neoplasms | 55 | X | | |
| 7 | Acute respiratory infection including pneumonia | Chronic obstructive pulmonary disease | 52 | X | | |
| 8 | Acute respiratory infection including pneumonia | Acute cardiac disease | 49 | X | X | X |
| 9 | Stroke | Acute abdomen | 48 | X | X | |
| 10 | Other and unspecified cardiac diseases | Acute abdomen | 48 | X | | |
| 11 | Acute respiratory infection including pneumonia | Stroke | 44 | X | X | X |
| 12 | Diabetes mellitus | HIV/AIDS-related death | 43 | X | X | |
| 13 | Other and unspecified cardiac diseases | HIV/AIDS-related death | 43 | X | X | |
| 14 | Acute abdomen | HIV/AIDS-related death | 42 | X | X | |
| 15 | Stroke | Digestive neoplasms | 42 | X | X | X |
| 16 | Acute respiratory infection including pneumonia | Acute abdomen | 40 | X | X | X |
| 17 | Stroke | Other and unspecified NCD | 31 | X | X | |
| **Distribution of probable co-occurring causes by group – % (n)** | | | | | | |
| Group A | | Group B | | | | |
| Non-communicable diseases | | Infectious, maternal, and neonatal causes | | 55.5 (1146) | 61.3 (976) | 64.9 (763) |
| Non-communicable diseases | | Non-communicable diseases | | 31.7 (654) | 30.5 (486) | 27.8 (327) |
| Infectious, maternal, and neonatal causes | | Infectious, maternal, and neonatal causes | | 5.4 (112) | 6.1 (97) | 6.1 (72) |
| Infectious, maternal, and neonatal causes | | Injuries and violent deaths | | 5.8 (120) | 0.2 (3) | 0.1 (1) |
| Non-communicable diseases | | Injuries and violent deaths | | 1.6 (32) | 1.8 (29) | 1.1 (13) |

* Considered at least co-occurring by one of the thresholds.
*Source*: INDEPTH Network (2014) and our calculation of the theoretical similarity index using the InterVA-4 probability matrix (malaria: VL; HIV: VL).

We considered 0.65 to be the most appropriate threshold based on our knowledge of possible confusions and the sensitivity of the results. At $p_{20}$, many additional associations considered as co-occurring appear as possible confusions: specifically, associations of diseases located in the same organs (acute respiratory infection associated with respiratory neoplasm [no. 6] or chronic obstructive pulmonary disease [no. 7]) and the association that have similar declarative symptoms (in particular the association of diabetes mellitus and other unspecified cardiac). On the other hand, although $p_{35}$ is significantly more restrictive (10% more possible associations are considered to be competing), it has a limited impact on the results (exclusion of 18% of associations compared to 0.65 and only associations with a frequency of less than 50). However, it does exclude a few associations that we considered sufficiently dissimilar to be regarded as multimorbidity: HIV/AIDs-related deaths with diabetes mellitus (no. 12) and other cardiac diseases (no. 13) – associations that are likely in older people living with HIV, especially if they are on treatment. In addition, these associations have no clear similarities in terms of symptoms. At the same time, we recognize that the selection of a threshold remains to some extent arbitrary and would require systematic physician review or a gold standard dataset of multiple causes to be validated.

## Figure A-4.2: Heat map of competing causes according to selected threshold




*Source*: InterVA-4's probability matrix (malaria: VL; HIV: VL) and our calculations of the theoretical similarity index.

## Appendix 5: Co-occurring vs. competing causes

### Figure A-5: Association of causes selected as co-occurring vs. resulting from uncertainty


*Source*: InterVA-4's probability matrix (malaria: VL; HIV: VL) using the theoretical similarity index with a threshold of 0.65.

## Appendix 6: Co-occurring causes by age and sex

### Table A-6: Probable co-occurring causes by age group and sex

| | All | | | Female | | | Male | | |
|---|---|---|---|---|---|---|---|---|---|
| | % of all death | % of more than one cause | n | % of all death | % of more than one cause | n | % of all death | % of more than one cause | n |
| 15–49 years | 1.3 | 14.9 | 374 | 1.5 | 16.0 | 211 | 1.2 | 13.6 | 163 |
| 50–64 years | 2.3 | 20.9 | 334 | 2.6 | 22.0 | 162 | 2.1 | 20.0 | 172 |
| 65+ years | 3.0 | 24.4 | 883 | 3.1 | 25.6 | 475 | 2.8 | 23.1 | 408 |
| All | 2.2 | 20.6 | 1591 | 2.4 | 21.7 | 848 | 2.0 | 19.4 | 743 |

*Notes:* Reading: Among the deceased aged 15 to 49, across both sexes, 1.34% of deaths were identified as resulting from co-occurring causes, which represents 374 deaths and 14.86% of the deaths among the deceased aged 15 to 49, across both sexes, attributed more than one cause by InterVA-4.
*Source:* VAs of 72,330 adults in 22 sites of the INDEPTH Network; 1992–2012 data.