

Producing consistent long-term time-series of cause-of-death statistics with verbal autopsies Lessons from the harmonization of 25 years of data in the Africa Health Research Institute

Ariane Sessego^{1,2}, Siyabonga Nxumalo², Dickman Gareta², Alison Castle^{2,3}, Kobus Herbst²

¹French Institute of Demographic Studies, ²Africa Health Research Institute, ³Massachusetts General Hospital

Introduction – the challenges of producing consistent time-series to assess long-term trends with verbal autopsies

Producing consistent long-term time-series is a challenge shared by many fields working on long-term trends (economics, demography, quantitative history). Methodology of data production and classification is ever-changing, constantly adapting to reflect state-of-the-art practices and evolving societal and research interests. These adaptations are central to ensure the relevance and quality of the data collected, and as some argue in sociology of statistics, its persistence over time (Didier et al. 2009). However, they represent a challenge for the analysis of long-term trends, and necessitate the development of methodologies to insure comparability across time. This has been particularly true for the analysis of cause-of-death statistics, subject to the constant evolution of the International Classification of Disease (Pechholdová et al. 2017)

The development of such methods represents two challenges: (1) the development of data harmonization methods to ensure comparability across time (and space); (2) preserving transparency and traceability of the data transformations carried out, to ensure reproducibility and document potential limitations - a dimension that is often neglected. This is a particularly pressing question in lower- and middle-income countries, where statistics remain heavily reliant on rapidly evolving estimation methods and data collection practices in a context of limited data availability but growing development of data collection infrastructures. When unrecorded, these rapid changes of methodology can lead to false interpretation of long term trends, as Morten Jerven's shows in his analysis of the construction and uses of GDP statistics in Eastern Africa (Jerven 2013).

The verbal autopsy (VA) method is a prime example of these rapid evolutions and the need to create methods of harmonization. VAs were developed in the 1950s to estimate causes of death in settings where many deaths occur far from health facilities and death certificates are not available, through a detailed interview with the caregivers of the deceased about illness and symptoms leading to death (Biraud 1956). They constitute crucial sources of data to monitor long-term trends in cause-specific mortality in contexts and times where little to no data is available, and have been collected in certain sites since the 1980s (Niakhar in Senegal - Garenne and Fontaine, 1986, Matlab in Bangladesh - Zimicki 1990). However, the evaluation of trends across time and space using VAs faces two main challenges: (1) the evolution of data collection tools, particularly the changes in questionnaires determining the information collected the symptom leading to death (2) the evolution of the method of estimation of causes of death: from physician reviews to the use of algorithms – tools that are rapidly evolving to improve accuracy and adapt to new needs in a context of a rapid epidemiological shift (Chandramohan et al. 2021). Applying consistent rules to assign causes of death, VA algorithms seem to offer a promising tool to produce consistent time-series. However, built and validated for a given VA questionnaire, little is understood on how consistently they perform when applied together with other and often older VA standard with missing or partially corresponding questions.

This objective of this study is to contribute to this reflection by presenting Africa Health Research Center (AHRI) Health and Demographic Surveillance System's (HDSS) experience in harmonizing 25 years of VA data (2000-2025) using two VA algorithms (InSilicoVA and InterVA5). Prompted by the introduction in 2022 of the WHO 2022 standard VA questionnaire (WHO 2025) and the need to transform this new data into the format expected by InterVA and InSilicoVA, AHRI developed a new harmonization framework, implemented in an open source package, *VAConvert*. This paper aims to (1) present this new methodology designed to promote transparency and traceability of data transformation and support the construction of lasting

infrastructure for future harmonization and (2) assess the consistency of the resulting cause-of-death estimation despite the evolution of VA questionnaires.

1. Data - 25 years of VAs at AHRI's HDSS: 27,622 VAs, three questionnaires,

Since 2000, AHRI's HDSS monitors vital events (births, death, in- and out-migration) in the rural district of uMkhanyakude, KwaZulu-Natal, South Africa. Initially covering an area of 438 km² with a population of approximately 85 000 people, the HDSS expanded in 2017 to cover an area of 845 km², and includes today more than 147,000 participants (Gareta et al. 2021). From the start of the HDSS, VAs have constituted a core part of the surveillance methodology. When a death is recorded, a trained nurse conducts a VA with the caregiver of the deceased. However, the instrument used to collect information about the decedent have changed.

Since 2000, the HDSS has used three different VA questionnaire (see Table 1).

- The “Africa Center’s questionnaire” was a paper-based questionnaire used from 2000 to 2016, developed within the HDSS (Hosegood et al., 2004). It was organized around two components: (1) a core questionnaire (dependent on the age group) (2) and a symptom checklist, a summarised list of key signs and symptoms (partly redundant with the core questionnaire) used as a prompting tool for VA nurse during the interview and an aid to review the VAs for physicians. A first checklist, used from 2000 to 2012, was replaced in 2012 by a much shorter checklist, with less redundancy but the addition of a few new questions not present in the core questionnaire.
- In 2017, AHRI's HDSS adopted in 2017 the newly released WHO 2016 VA standard questionnaire, result of WHO's effort to standardise VA instruments in a context of growing use of the method (Chandramohan et al., 2021). This new questionnaire represented an important change in the content and structure data collection, with substantial change in types of questions and wording. It also marked the start of digital data collection for VAs (Gareta et al., 2021).
- In 2022, the HDSS adopted the updated version: the 2022 WHO VA standard, the result from a simplification of the 2016 WHO standard, based on qualitative and quantitative analyses (Nichols et al., 2022; WHO, 2025).

Overall, from 2000 to July 2025, 27,622 deaths have been documented through VAs. Of them, more than two thirds (70.4%, 19,434 deaths) were collected using the Africa Center’s questionnaire – 12, 965 with the first checklist, 5,214 with the second checklist –, 19.3% (5,335) using the WHO 2016 VA standard, 10.3% (2,853) using WHO 2022 VA Standard. As the original questionnaire was paper-based, the data was transcribed by data capturers into a digital format.

Table 1 – Number of deaths documented by questionnaire

| Dates of death | Questionnaire type | | Number of VAs N (%) | |
|-----------------|-------------------------------|--------------------------------|------------------------|-----------------|
| 2000–2016 | Africa Center’s questionnaire | Checklist 1 (VAC) 2000-2011 | 19,434 (70.4%) | 12,965 (46.9%)* |
| | | Checklist 2 (VCL) 2012-2016 | | 5,214 (18%)* |
| 2017-2021 | 2016 WHO VA standard | | 5,335 (19.3%) | |
| 2022 to present | 2022 WHO VA standard | | 2,853 (10.3%) | |
| Total | | | 27,622 (100%) | |

*These do not add up to the global number of Africa Center VAs, as some VAs do not have checklist, some VAs have both checklists, and very few only have a checklist and no core questionnaire.

2. Methods – data harmonization and assessment of the consistency of cause assignment

VAConvert package: a transparent infrastructure for VA data harmonization

To harmonize all historical VA data into the standard format expected by InterVA5 and InSilicoVA, we developed a structured data processing framework to promote transparency and traceability of data transformation, implemented in a Julia¹ package, *VAConvert* (soon to be made public). It builds on the structure of the open-source tool PyCrossVA², and extends them data pre-processing stages and quality control. Its key principle is that all data transformation is performed through correspondence tables : spreadsheets where each row represents a variable expected in the standard format and columns specify from which variables and how they are to be recoded. Easily accessible, legible and amendable by non-expert users, these correspondence or “mapping spreadsheet” make for a transparent framework of data transformation, that can be easily repurposed for future harmonizations (if a different input format is required) or by other sites collecting VAs.

While the WHO 2022 questionnaire is the latest international standard, there is yet no validated algorithm based on this questionnaire. We use InterVA5 and InSilicoVA’s input–based on WHO 2016 questionnaire - as our harmonization standard (both algorithms use the same input format and only differ in their statistical procedure - McCormick et al. 2016). All variables expected by the algorithm do not necessarily have an exact equivalence in all questionnaire. We assessed the quality of the mapping for each variable using three categories: (1) **exact**, when the harmonized and original question match perfectly (2) **approximate**, when the question mapped to the standard format only partially correspond to the information expected by the algorithm (3) **no matching question**, when there is no variable corresponding to the algorithm’s required input.

Approximation are mainly due to two configuration. First, (i) when distinctions were not made in the original questionnaire but are made by the algorithm. For example, the algorithm’s input differentiates between the following three indicators used for determining stillbirths and birth asphyxia: “*Did the baby ever breathed?*”, “*Did the baby ever cry?*”, “*Did the baby ever moved?*”. However, in Africa Center’s questionnaire (2000-2016), these are all combined into one question: “*Was the child alive (breathed, cried or moved) at birth?*”, leading to three approximate mapping. Conversely, (ii) when only part of the symptom expected by the algorithm is documented in the questionnaire. For example, one of the algorithm’s indicator is “*Did (s)he have swollen legs or feet?*”, but in Africa Center’s questionnaire the only information about the swelling of the ankles was collected “*Did s/he have swelling around ankles?*”, this is considered an approximate mapping.

Assessing cause assignment consistency

Building on the harmonized datasets, the next step will assess the impact of questionnaire evolution, particularly missing indicators on cause-of-death attribution. We will use the 5,335 VAs (2017-2021) collected using the WHO 2016 questionnaire, for which all indicators are exactly mapped, to test the consistency of cause-of-death attribution. Removing the indicators missing in the other questionnaires, we will compare the causes of death attributed when all indicators are present and when some are missing. We will assess the consistency of cause-of-death (CoD) assignment both at the population level and at the individual level with InSilicoVA and InterVA separately, using descriptive statistics (eg. confusion matrices) and summary measures (eg. cause-specific mortality fraction accuracy). We will assess this consistency by age groups, and at different level of cause-of-death aggregation. We aim to provide guidelines to analyze robust trends over time, such as appropriate grouping of causes.

¹ Julia is a data management coding language and software particularly efficient to handle large datasets, <https://julialang.org/>

² <https://github.com/verbal-autopsy-software/pyCrossVA>, created by the OpenVA team (<https://openva.net/>)

3. Preliminary results and expected development

Table 2 summarizes the quality of data harmonization across the three questionnaires. InterVA5 and InSilicoVA were developed using the 2016 WHO questionnaire as a framework, this questionnaire was the easiest one to map. Of the 353 indicators required by the algorithms, 352 could be mapped exactly to the algorithm's standard and only 1 was mapped approximately. The 2022 WHO questionnaire represents a simplification of the 2016 questionnaire: overall 84% of indicators could still be mapped exactly, but 7% had to be mapped approximately and 10% (35 indicators) could not be mapped at all as there were no corresponding question in the collected data³. The oldest questionnaire - spanning from 2000 to 2016 – was the hardest to harmonize: a third of indicators (33%, 117 indicators) could not be mapped, and 8% (30 indicators) of the mapping is approximate.

Table 2– Quality of the mapping of each questionnaire to the format expected by the algorithms

| Quality of the mapping | Africa Center's questionnaire (2000-2016) | | WHO 2016 (2017-2021) | | WHO 2022 (2022 to present) | |
|------------------------|---|-------|----------------------|-------|----------------------------|-------|
| | n | % | n | % | n | % |
| Exact | 206 | 58.4 | 352 | 99.7 | 295 | 83.6 |
| Approximate | 30 | 8.5 | 1 | 0.3 | 23 | 6.5 |
| No matching question | 117 | 33.1 | | | 35 | 9.9 |
| Total | 353 | 100.0 | 353 | 100.0 | 353 | 100.0 |

Evaluating the consistency of cause-of-death assignment, we expect to find minimal discrepancy for WHO 2022 questionnaire, as the number of missing question is low and these questions were strategically to minimise information loss (Nichols et al., 2022; WHO, 2025). However, we expect to find higher discrepancies for the older questionnaire (2000-2016) as many indicators are missing, particularly for certain age groups (neonates and children) and cause categories (maternal deaths, cardiovascular diseases) where questions have considerably evolved. We also hypothesize that InSilicoVA will produce more stable and internally consistent CoD assignments than InterVA, as it distinguishes between “no” and a missing information compared to InterVA (McCormick et al. 2016). We hope to provide guidelines for future data users, such as appropriate grouping of causes, to be able to analyze trends over time.

Preliminary conclusion and scope

At a time when many VA data collectors are switching from WHO 2016 to WHO 2022 standard, this work hopes to contribute to the reflection around building transparent and robust harmonization processes for VAs and beyond. Our aim is to raise the question of continuity in cause-of-death assignment despite evolving data collection methods, a recurring problem when using VA algorithm that - to our knowledge – has not yet been addressed in the literature. We hope to provide ways to overcome this limitation.

³ While the WHO 2022 questionnaire is the latest international standard, there is yet no validated algorithm based on this questionnaire.

References

- Biraud, D.Y. (1956). *Méthode pour l'enregistrement par les non-médecins des causes élémentaires de décès dans des zones sous-développées*. Genève: World health organization diffusion/distribution.
- Chandramohan, D., Fottrell, E., Leitao, J., Nichols, E., Clark, S.J., Alsokhn, C., Cobos Munoz, D., AbouZahr, C., Di Pasquale, A., Mswia, R., Choi, E., Baiden, F., Thomas, J., Lyatuu, I., Li, Z., Larbi-Debrah, P., Chu, Y., Cheburet, S., Sankoh, O., Mohamed Badr, A., Fat, D.M., Setel, P., Jakob, 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.
- Didier, E., Névanen, S., Zauberman, R., and Robert, P. (2009). La solidité des institutions. Les statistiques de « victimation » de l'Insee (1996-2006). *Genèses* n° 74(1):128. doi:10.3917/gen.074.0128.
- Garenne, M. and Fontaine, O. (1986). Assessing probable causes of deaths using a standardized questionnaire : a study in rural Senegal. *Seminar on comparative studies of mortality and morbidity : old and new approaches to measurement and analysis*:22.
- Gareta, D., Baisley, K., Mngomezulu, T., Smit, T., Khoza, T., Nxumalo, S., Dreyer, J., Dube, S., Majozi, N., Ordning-Jespersion, G., Ehlers, E., Harling, G., Shahmanesh, M., Siedner, M., Hanekom, W., and Herbst, K. (2021). Cohort Profile Update: Africa Centre Demographic Information System (ACDIS) and population-based HIV survey. *International Journal of Epidemiology* 50(1):33–34. doi:10.1093/ije/dyaa264.
- Jerven, M. (2013). *Poor Numbers: How We Are Misled by African Development Statistics and What to Do About It*. Cornell University Press. Ithaca.
- McCormick, T.H., Li ,Zehang Richard, Calvert ,Clara, Crampin ,Amelia C., Kahn ,Kathleen, and 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.
- Pechholdová, M., Camarda, C.-G., Meslé, F., and Vallin, J. (2017). Reconstructing Long-Term Coherent Cause-of-Death Series, a Necessary Step for Analyzing Trends. *European Journal of Population = Revue Européenne De Demographie* 33(5):629–650. doi:10.1007/s10680-017-9453-1.
- WHO (2025). *Verbal Autopsy Standards: The 2022 WHO Verbal Autopsy Instrument*. Geneva: World Health Organization. License: CC BY-ND 3.0 IGO. https://cdn.who.int/media/docs/default-source/classification/other-classifications/autopsy/2022-va-instrument/verbal-autopsy-standards_2022-who-verbal-autopsy-instrument-designed.pdf?sfvrsn=5633badb_1&download=true.
- Zimicki, S. (1990). Approches to Assessment of the Cause Structure of Mortality : A Case-Study from Bangladesh. In: Vallin, J., D'Souza, S. and Palloni, A. (eds). *Measurement and Analysis of Mortality : New Approaches*. Oxford New York: Clarendon Press Oxford University Press. https://campus-condorcet.primo.exlibrisgroup.com/discovery/fulldisplay?docid=alma991006560439705786&context=L&vid=33CCP_INST:CCP&lang=fr&search_scope=ALL&adaptor=Local%20Search%20Engine&tab=ALL&query=sub,exact,MESURE%20DE%20LA%20MORTALITE%20%2F%20MORTALITY%20MEASUREMENT,AND&query=creator,contains,Vallin,AND&mode=advanced&offset=0.