DISCLAIMER – This document is a draft version and does not represent the final version. It will be updated to incorporate the latest version of the International Classification of Diseases 11th Revision (ICD-11). The draft is being released to avoid delaying the use of the document, with the final version to follow, including ICD-11 updates.

[DRAFT]

An approach to joint use of Verbal Autopsy and Medical Certificate of Cause of Death Data

Version 0.5

World Health Organization

Suggested citation. (Draft) An approach to joint use of Verbal Autopsy and Medical Certificate of Cause of Death Data; Geneva; World Health Organization; 2025; License: CC BY-ND 3.0 IGO.

ACKNOWLEDGMENTS

This draft document was authored by Erin Nichols and Brian Munkombwe from the U.S. Centers for Disease Control and Prevention (CDC); Don de Savigny and Daniel Cobos Muñoz from the Swiss Tropical and Public Health Institute, University of Basel; Magdalena Paczkowski and Farnaz Malik from Vital Strategies; and Jordana Leitao and Carla AbouZahr, independent contractors.

Statistical and analytic support was provided by Yulei He from the U.S. CDC; Marek Kiwatkowski from the Swiss Tropical and Public Health Institute, University of Basel; Sam Clark from the Ohio State University; and Tyler McCormick from the University of Washington.

Support for testing the approach was provided by Chabila Mapoma from the University of Zambia and Peter Moyo from the Zambia Department of National Registration Passport and Citizenship (DNRPC), which is the civil registration authority in Zambia.

This draft document benefited from the contributions of numerous organizations, institutions, and individuals who provided strategic guidance, technical expertise, and critical review. Special thanks are extended to Daniel Chandramohan from the London School of Hygiene & Tropical Medicine; Emily Cercone and Kristen Pettrone from the U.S. CDC; Martin Bratschi, Fatima Marinho, Robert Mswia, Romain Santon, and Philip Setel from Vital Strategies; Sonja Firth, former VARG member; William Msemburi from Institute for Disease Modelling; Peter Byass from Umeå University; Chalapati Rao from the Australian National University; and Colin Mathers, former WHO.

The production of this document was made possible by the review of the members of the WHO Verbal Autopsy Reference Group (VARG). In addition to those already acknowledged, the VARG includes Shams El Arifeen, International Centre for Diarrhoeal Disease Research; Aurelio Di Pasquale, Swiss Tropical and Public Health Institute; Henry Kalter, Johns Hopkins Bloomberg School of Public Health; Soewarta Kosen, National Institute of Health Research Indonesia; Arvind Pandey, National Institute of Medical Statistics, India; Riley Hazard, former member; Samuel Cheburet, Ministry of Health, Kenya; Edward Fottrell, University College London; Debbie Bradshaw and Pamela Groenewald, South African Medical Research Council.

From the World Health Organization (WHO), acknowledgement goes to Robert Jakob and Carine Alsokhn from the Classifications and Terminologies Unit; as well as Haidong Wang, Doris Ma Fat, and Bochen Cao from the Monitoring, Forecasting, and Inequalities Unit.

TABLE OF CONTENTS

INTRODUCTION

Multiple initiatives are promoting and supporting the improvement of civil registration and vital statistics (CRVS) systems around the globe, with the aim of increasing the availability of high-quality information on births and deaths for planning and evaluation purposes. An important aspect of mortality information is statistics on causes of death, disaggregated by key variables like age, sex, and location. While a medical certificate of cause of death is the ideal source of this information, there are many places where a physician is not able to attend every death, or where information on the cause of death is insufficient, as in "brought in dead" or "dead on arrival" cases. In such cases, a verbal autopsy (VA) can be used to establish a probable cause of death.

VA consists of a structured interview with persons close to the deceased to record information on signs and symptoms experienced by the deceased in the time before death. Recent developments in VA methods have increased the availability of this source of cause of death data to augment mortality statistics. Various efforts are underway to streamline and standardize the processes for analyzing VA interview data, including efforts to automate the cause of death assignment process. Also, there are numerous efforts to scale the application of VA to large, nationally representative populations. However, guidance is lacking on how to integrate this emerging routine CRVS data source into national and subnational routine vital statistics publications, particularly where VA is used instead of medical certificate of cause of death (MCCD) to provide the cause of death information, where MCCD is not possible.

A technical meeting on "Mortality Data Analysis with Verbal Autopsy" was therefore convened to generate a draft set of principles, considerations, and recommendations about how countries can: 1) yield high-quality statistics from VA data and 2) integrate VA results and other sources of mortality data into their mortality data processes. The meeting objectives were to outline key principles, considerations, and recommendations for:

- The use, analysis, and presentation of scaled VA data as a component of a national civil registration and vital statistics system, including the key statistical measures and information that can be derived from VA data; and
- Integrating VA data into mortality analysis, with a particular focus on how to analyze VA data in conjunction with cause of death data from other routine sources – particularly facility-based, medically certified and medico-legal sources.

Drawing from and expanding on the output of this meeting, the present document aims to guide how to aggregate mortality data from multiple sources and to integrate the results into

national statistical processes. While such guidance is expected to evolve as more countries gain additional experience in putting mortality data to use, this initial guidance aims to demonstrate the utility of VA data alongside MCCD data, moving countries towards using VA data now. Despite the inability for most VA data to be used as a legal product, this guidance includes practical steps for putting useful information derived from VA, in the hands of policy makers to support public health and policy development.

This draft document builds on the "Guidance for interpreting VA results" (D4H TWG, 2020) and "Integration of data from medical certification of cause of death and verbal autopsy" (Adair et al., 2020), which outline steps to interpreting and presenting VA data and include considerations for integrating VA and MCCD data. The present document offers additional practical step-by-step details for addressing various considerations, including selecting a common cause list and adjustments for incompleteness in data. Section 1 of this guidance document describes steps to understand the context of the data that are being analyzed. With this understanding, Section 2 outlines the prerequisites or minimum standards that the available mortality data should meet before further aggregation is considered. For data that are considered of sufficient quality, Section 3 outlines the steps for preparing and aggregating the datasets, such that a single mortality profile can be generated for the area of interest. Finally, Section 4 provides guidance on using and interpreting the resulting findings, including suggestions on how to describe the limitations of the aggregated data.

SECTION 1: UNDERSTANDING THE CONTEXT

Aggregating data from different sources and of different levels of quality requires a thorough understanding of the data sources, the context in which the data were collected, and the corresponding limitations and uncertainties associated with the use of data. Table 1 describes the various factors that should be considered when aggregating VA and MCCD data, including an indication of how each factor may be addressed. It is evident that while some factors may be readily addressed with existing data or knowledge, others will require further research to expand the understanding of the systematic and random processes that underlie and impact the collection, analysis, and use of mortality data in various contexts. Therefore, Table 1 may also be used to support a continuing operations research agenda.

It is recognized that there may be various VA applications and multiple sources of mortality data within countries (e.g., HDSS sites or various health projects) that are relevant to consider for inclusion in an aggregated mortality profile. While the present guidance can serve as a foundation from which methods to aggregate additional sources of mortality data can be further developed, it focuses on a standard case of aggregating centrally administered, routine VA applications with official MCCD data for the country.

Table 1: Factors to consider when aggregating VA and MCCD data

Factor	Definition and applicability	Action
Coverage	The extent to which the country is represented by the civil registration/ MCCD/VA data system and the corresponding impact on the availability of MCCD and VA data for analysis. Alignment is needed between the geographic area that the MCCD and VA data represent.	Ensure that MCCD and VA data for analysis represent the same geographic area.
Completeness and representativeness	Completeness refers to the extent to which deaths are registered (or, in this case, cause of death information is included in the appropriate dataset), among the target population or the population covered by the vital registration/ MCCD/VA system.[1] Representativeness describes the extent to which deaths in the available dataset(s) resemble the population from which the deaths were derived; relevant factors for comparison include age, sex, and location (e.g., health facility or community) of the death.	Estimate completeness of MCCD and VA datasets using death rate, total population size, and the hospital/community death ratio. Compare age and sex structure of the target population to the available MCCD and VA data; aggregation may not be appropriate if the age/sex distributions vary greatly. If data are aggregated, MCCD and VA frequencies could be weighted to adjust for non-representativeness by age and sex factors. Consider system design and other factors to estimate who is likely captured and missed in the system; describe interpretations as part of limitations in terms of ascertainment/ sampling and selection bias.

[1] https://www.who.int/healthinfo/statistics/mortcoverage/en/

		Confidence intervals may be used in advanced calculations to propagate the uncertainty due to lack of representativeness of the estimates.
Misclassification of cause of death	In MCCD and physician-coded VA (PCVA), misclassification is a systematic error due to intentional or unintentional bias of the physician in assigning a cause of death. Such bias may result from a lack of information, variations in physician knowledge and familiarity with epidemiologic circumstances, and variations in physician certification/coding practices. In VA where cause of death is assigned by an automated algorithm, misclassification is a systematic error due to incorrect logic or symptom-cause information of the algorithm. It is recognized that misclassification of cause of death is likely differential, where the potential for or extent of misclassification varies by cause (Polprasert et al., 2010).	Where resources are available, an ad hoc validation exercise could be undertaken to estimate the extent of misclassification or under-reporting (Rao et al., 2010; Bradshaw et al., 2020). The potential for misclassification by cause can also be taken into consideration when interpreting data (Polprasert et al., 2010).

VA classification uncertainty -- inherent variability in VA classification algorithms	Automated algorithms for classifying VA deaths apply different logic and probabilities associated with the symptom-cause relationship that result in variation in the cause assignment, even given the same symptoms in the same set of deaths. Furthermore, they rely on data that are characterized by variation and error sourced from potential inaccuracy in the VA responses and variability arising from individual variation in presentation of diseases. (Clark *et al.,* 2013)	The extent of uncertainty derived from these sources can be quantified by a confidence interval reported for individual cause assignments and population distributions. While work is ongoing in this space, currently only the InSilicoVA software reports such a confidence interval.
Different VA target cause of death lists	VA questionnaires include questions that are readily answerable by VA respondents and that help to characterize causes of death of public health importance. Accordingly, the questions are associated with a fixed list of potential causes of death to which automated VA cause-assignment software have been programmed.	If VA data are to be aggregated across datasets for which cause of death has been assigned using software associated with different cause lists (e.g., InterVA and SmartVA), causes will need to be mapped to a common list before the data can be aggregated or tabulated. While there have been efforts to map the lists (Cobos Muñoz and de Savigny, 2018) currently, there is no standard harmonized cause list. Accordingly, teams must decide how to map discrepant categories and document how discrepancies have been addressed to harmonize the cause assignments to the preferred target cause list.
Sampling variability	For VA, where a random sample has been applied, the observed sample of deaths may	If the VA sample has been selected at random, confidence intervals can be calculated around the aggregated,

	have different characteristics than the population the sample aims to represent.	weighted cause specific mortality fractions. Work is ongoing to determine a methodology for quantifying the sampling error to account for the variability from the VA sample, aggregated with MCCD set.

A solid understanding of the source population structure, data source, and quality of data is needed before any data analysis effort should be undertaken; this is particularly true when data on the same indicator are to be aggregated from different sources (e.g., data from underlying cause of death from medical certification sources and from VA sources). Given the universal value of mortality data, countries are striving to collect it routinely. However, while principles and recommendations for its routine collection are established,[2] the level of resources and political commitment available to support such collection varies widely around the world. This variation poses an important impact on the characteristics of the resulting mortality data that must be taken into consideration during analysis, use, and interpretation of the data. For example, a country with a significant portion of its population living in rural/remote areas with relatively few medically trained staff and therefore a low percentage of MCCD for decedents will likely rely on the use of VA to supplement its mortality information for the foreseeable future. Alternatively, a better-resourced country with widespread access to medically trained staff and commitment to improving quality of MCCD data may rely on the systematic inquiry process of VA only to complement the information available to improve the resulting MCCD. Each of these scenarios will require different considerations that must be factored into the analysis and interpretation of the data.

To better understand mortality data, an attempt should be made to describe the deaths captured in the various systems that record or register mortality information, including details on the percentage of deaths recorded in the various systems and the extent of overlap. For example, Figure 1 shows a hypothetical example of the deaths captured within civil registration, VA, and MCCD systems in a country. In this example, of all estimated deaths occurring in a given year, 49% of deaths are missed by all of the systems to record mortality (i.e., medical certification of cause of death in health facilities, civil registration, and/or the VA system) and are completely undocumented. This leaves 51% that are captured in at least one or more of the systems. Only 27% of deaths are registered in CRVS. If all deaths with an MCCD or a VA were passed onward for civil registration, CRVS completeness could be substantially improved. Similarly, the majority of registered deaths do not contribute to understanding cause-specific

[2] https://unstats.un.org/unsd/demographic/standmeth/principles/M19Rev3en.pdf

mortality fractions (CSMFs), as they have not been facilitated to obtain an MCCD or VA. These are the horizons for improving mortality documentation.

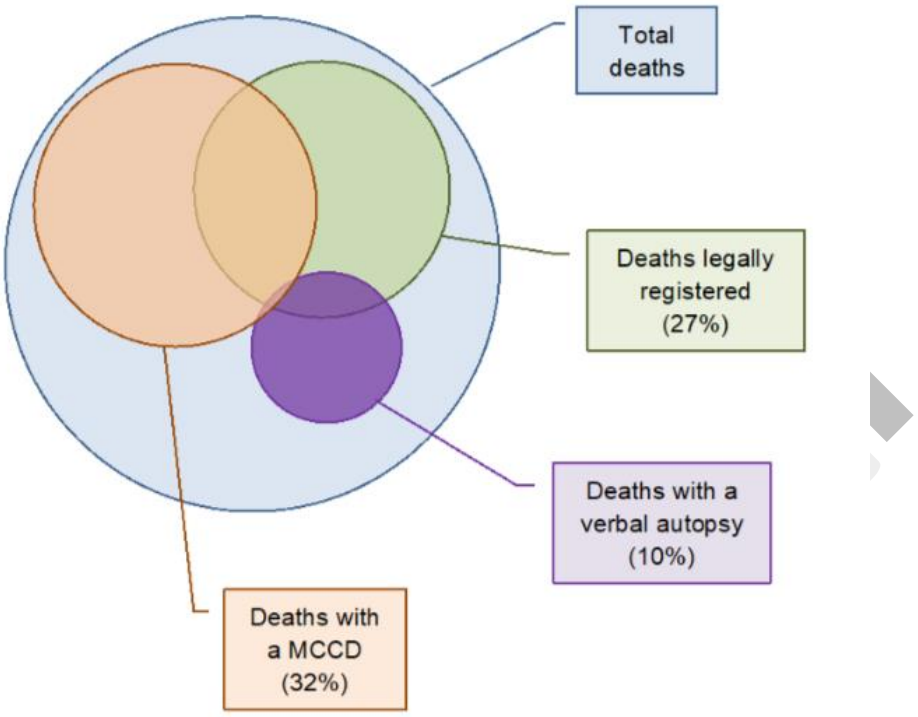


Figure 1: Summary of sources of mortality data in a hypothetical country

The blue space in this figure represents deaths that are not captured in any system-- that is, deaths for which we do not have any information. The extent of incompleteness from these missing deaths will impact the degree of potential bias in mortality estimates. Further understanding of the demographic characteristics, including age, sex, and location of the deaths that are captured versus not captured, can help analysts evaluate the impact of this bias. A better understanding of these dynamics can be achieved by considering how deaths are captured by the various systems. For example:

- What health facilities submit medical certificates of cause of death (e.g., all facilities, district-level facilities, public facilities, private facilities, university facilities, etc.)?
- Are any deaths likely omitted from the MCCD dataset (e.g., deaths referred for medical examination and/or police investigation)?
- Which deaths are targeted to receive a VA (e.g., rural/community deaths in selected geographic areas, deaths without an MCCD, ill-defined deaths from the MCCD dataset, dead on arrival, brought-in-dead)?
- Is verbal autopsy applied to all target deaths or a sample?

With this understanding of the source of MCCD and VA data, the next step is to better understand the quality of the available data.

SECTION 2: DATA PREREQUISITES (MINIMUM QUALITY STANDARDS)

To ensure that resulting data can be interpreted with a clear understanding of their strengths and limitations, including the potential cause of death misclassification and bias, the general quality of the available VA and MCCD data should be evaluated to ensure that they meet certain minimum standards before further analysis and aggregation is conducted. Various aspects on which the data should be evaluated are described below. It should be noted that these are suggestions based on known experience to date. As context varies, application of criteria may need to be adjusted accordingly. An example describing a system's context (Section 1) and data prerequisites (Section 2) is included in Appendix A for a dataset from Lusaka, Zambia.

A. Size of datasets

For analysis, there are two primary considerations concerning the size of datasets. The first concerns tabulating data where low frequencies may result in random variation from year to year, increasing uncertainty and making it difficult to interpret trend data. The number of desired disaggregations (e.g., by sex, age groups, and/or location) inversely impacts the sample size for given estimates, and a balance must be achieved between the selected tabulations and the available data. Collectively, the MCCD and VA data must contain a sufficient number of deaths across the various cause categories and desired disaggregations for suitable interpretation.

A minimum cell frequency of 20 is recommended to avoid substantial fluctuations of the CSMFs due to low numbers (Heron, 2021). Assuming that countries would be interested in estimating the top 20 causes of death, this would mean that the cause of death (COD) ranked as the 20th should have at least 20 deaths after combining the MCCD and VA datasets. Table 2 shows the estimated number of deaths in the combined dataset given 2019 CSMF distributions from Sri Lanka, Tanzania, Zambia, Philippines, Ghana, and Brazil (IHME, 2019). From these examples, 1,500 deaths would meet the numbers needed for minimum cell frequency in most cases. Note that this number applies to any disaggregation desired, so if the CSMF distribution were needed for both males and females, this number would be needed in each group.

Table 2: Estimated number of deaths for each of the causes of deaths in the top 20 ranking, given a minimum cell frequency of 20 in the 20th cause (IHME, 2019)

CSMF Rank	Sri Lanka	Tanzania	Zambia	Philippines	Ghana	Brazil
1	457	263	334	288	193	242
2	200	208	176	181	162	141
3	194	125	160	164	143	105
4	122	107	127	76	137	91
5	99	96	106	76	136	85
6	90	90	74	65	79	78
7	85	84	69	57	61	74
8	81	56	62	50	47	64
9	76	55	57	40	46	45
10	54	50	54	39	44	39
11	37	49	41	34	42	32
12	31	48	35	32	41	27
13	31	43	35	27	36	26
14	26	39	31	27	36	25
15	26	36	30	26	27	25
16	23	31	25	22	26	24

CSMF Rank	Sri Lanka	Tanzania	Zambia	Philippines	Ghana	Brazil
17	23	30	22	22	26	22
18	21	29	20	22	23	22
19	21	23	20	21	23	20
20	20	20	20	20	20	20
TOTAL	1,717	1,481	1,497	1,288	1,348	1,207

The second consideration, which applies when a sample of deaths is being captured, relates to the uncertainty due to the sampling error; for example, if a sample of deaths is selected for VA. In most cases, it is anticipated that any random sampling in VA-MCCD work would most likely be applied to the VA deaths, with an attempt to capture all MCCD deaths in the target population. For such cases, additional work is needed to develop the formulae to estimate the uncertainty around the CMSFs produced when combining MCCD and VA data, considering sampling methods and completeness levels for each data source. In the interim, we can take the draft document from sample size estimates for VA samples.

As described in the "Sampling Strategies for Representative National CRVS Verbal Autopsy Planning" guidance document (VA sampling guide) (WHO and Bloomberg Philanthropies Data for Health Initiative, 2018), we are interested in comparing CSMFs over time and most probably sampling the same administrative units in a country (e.g., districts or wards). As a consequence, to estimate the uncertainty around the CSMFs produced from a VA dataset we would need to consider a cluster sampling frame with a matched design. The size of the cluster, the inter- and intra-cluster correlation, and the proportion of deaths with an MCCD are other factors influencing this estimate. Considering how these factors will vary depending on the local context, it is difficult to provide a "one size fits all" calculation of the minimum number of deaths recommended for analysis. The VA sampling guide provides all the relevant information and methods to calculate the appropriate number of VAs depending on the level of uncertainty that we expect to have for the different CSMFs of each disease. Box 1 describes some scenarios for sample size calculations in low- and middle-income counties (LMICs).

As mentioned above, in addition to the standard sample size calculation parameters, countries should consider whether they are interested in disaggregating by e.g., sex/gender, age or place of residence. If more data are available, further disaggregation could be considered (e.g., sub-national tabulations and/or more age groups). To meet these minimum sample size needs, annual tabulations, or possibly tabulations for multiple aggregated years, are recommended. Further guidance is available in Section 4 on managing small cell frequencies.

Box 1: Levels of CSMF certainty for various VA sampling scenarios

We assume for these calculations:

- Cluster design for which countries will select a number of districts to collect VAs to get a nationally representative sample.
- There is no need to disaggregate.
- The uncertainty is only based on the VA sample because MCCDs are not sampled.

A sample of 5,000 VAs will provide reasonable certainty about:

- First and second CODs in the ranking are within the top 3 causes.
- CODs 3-9 are within the top 20 causes.
- There will be too much uncertainty for COD 10th and above to make meaningful inference on the CODs distribution.

A sample of 10,000 VAs will provide reasonable certainty about:

- First and second CODs in the ranking are within the top 3 causes.
- CODs 3-7 are within the top 10 causes.
- COD 8-13 are within the top 20 causes.
- Too much uncertainty after COD number 14th to make meaningful inference on the CODs distribution.

A sample of 50,000 VAs will provide reasonable certainty about:

- First and second CODs in the ranking are most probable top 1 and 2 CODs.
- CODs 3-7 are within the top 7 causes.
- COD 8-13 are within the top 15 causes.
- COD 14-20 are within the top 20 causes.

B. Consistency in data collection and COD assignment methods

At minimum, MCCD and VA datasets should reflect consistent time periods, e.g., deaths occurring in a given year. It is recognized that countries may have multiple sources of mortality data that they might consider including in an aggregated mortality profile (e.g., from different VA applications and/or from Health and Demographic Surveillance Sites or other mortality survey or surveillance efforts). While the present document can serve as a foundation from which methods to aggregate additional sources of mortality data can be further developed, it focuses on a standard case of aggregating centrally administered, routine, and sample-based VA data with official MCCD data for the country. While data from different sources (i.e., VA or MCCD for the purpose of this document) may vary in the methods in which they were collected and in which the cause of death was assigned, the extent of variation in methods and potential implications on interpretation should be considered. Factors to consider include the purpose of data collection (e.g., surveillance for a particular cause versus routine tracking of all-cause mortality); data collection methods (e.g., questionnaire used); and COD assignment method used, considering the associated cause list and probability matrix for automated methods (see Table 1). The ability to compare trends over time also requires consideration of the consistency in data collection and COD assignment methods used in the datasets for comparison.

C. Quality of data

Using high quality data for analysis will minimize measurement error and its resulting impact on misclassification in the results. Select quality indicators are described below for both VA and MCCD data. Use of electronic tools for analysing mortality rates and cause of death data, is recommended to assess the plausibility of the data, as an indicator of the accuracy of the information. Currently available tools include WHO's "Analysing mortality level and cause-of-death data" or ANACoD tool (WHO, 2021) and the related "Analysis of Causes of National Deaths for Action" or ANACONDA mortality data quality assessment tool for MCCD data.[3] Related guidance in the VA context is available in "Guidance for interpreting VA results" (D4H TWG, 2020), with the corresponding VIPER tool.[4] With the age and sex structure of the source population and MCCD- or VA-derived cause of death data by sex and standard age groups, these tools automatically perform a variety of calculations as part of a comprehensive data quality review and can be used to evaluate the indicators described below.

[3] https://www.who.int/standards/classifications/classification-of-diseases/services/analysing-mortality-levels-and-causes-of-death
https://bmcmedicine.biomedcentral.com/articles/10.1186/s12916-020-01521-0

[4] https://data4healthlibrary.org/resources/viper; the alpha version of this tool is compatible with SmartVA; additional work is underway to extend its compatibility to output from other VA algorithms.

C.1. Coverage, completeness, and representativeness

This draft document uses the concepts of coverage and completeness as described by WHO, where coverage refers to the extent to which the country is represented by the vital registration/MCCD/VA data system, while completeness refers to the extent to which deaths are captured - in this case, cause of death information for target deaths is included in the appropriate dataset - among the target population or the population covered by the vital registration/MCCD/VA system.[5] Insufficiencies in completeness may result from various issues, including, but not limited to, deaths not being notified or registered, cause of death not being assigned to a death, or cause of death not being recorded in the available dataset (e.g., MCCD records are lost due to systematic errors within a hospital information system, or findings from police investigation or medical examination are not provided back to the registration authorities). To clarify, where the target deaths for analysis are from a sample, as is often the case for VA data, completeness is applied to the proportion of target deaths captured in the sample, rather than the proportion of the whole population. Also, to note, insufficiencies in completeness may vary by age and sex. This document provides an overview of issues related to completeness. Additional details can be found elsewhere (Adair and Lopez, 2018; D4H TWG, 2020; Rao *et al.*, 2020).

With an aim to optimize the usefulness of available data, datasets should be reasonably complete (e.g., ideally above 80%, but at least above 50% (D4H TWG, 2020) though they may not represent full coverage of the country. Where coverage is lacking, but available datasets are of sufficient completeness, care should be taken to ensure that the VA and MCCD datasets targeted for aggregation represent the same broad geographic areas. Information on the source population structure, including age and sex distribution, is needed to evaluate the completeness of the datasets. For example, populations that have a high proportion of the population in age groups where mortality rates are highest (i.e., below 4 years and above 60 years) can be expected to have a higher crude death rate (CDR). If the observed mortality data suggest a low CDR, under-reporting and incompleteness of cause of death information is expected. ANACoD, ANACONDA, and VIPER can be used to evaluate the completeness of mortality data.

A related concept to completeness is representativeness, which applies when a sample of deaths has been taken to represent a broader population. Representativeness describes the extent to which deaths in the available dataset(s) resemble the population from which the deaths were derived. A variety of factors may contribute to a lack of representativeness,

[5] https://www.who.int/data/gho/data/indicators/indicator-details/GHO/completeness-of-cause-of-death-data-(-)

including high refusal by VA respondents, insufficient information in the VA interview to complete a cause of death, or deaths not being reported and therefore a VA interview not being completed. For aggregation purposes, age and sex are two practical and informative factors on which representativeness can be assessed. If the age and/or sex distributions of the available sample dataset differs greatly (e.g., more than 20%) from that of the source or target population, the data may not be appropriate for aggregation. It is recognized that other population and system dynamics are likely to contribute to the representativeness of the data system, and where available, further information can be used for assessment.

C.2. Accuracy of information (measurement error)

Accurate information is needed from the VA interview and medical certificate of cause of death in order to assign causes with minimal bias and/or misclassification. A variety of measurement errors may contribute to a lack of accuracy, including reporting bias, or selective revealing or suppression of information (e.g., about past medical history, smoking, or other risk factors) by respondents; recall bias, or differences in accuracy or completeness of recall to memory of past events or experiences related to the death; cognitive error, where respondents do not comprehend a given question as intended (particularly an issue for VA interviews); and erroneous data entry or recording of responses. Where physicians are involved in assigning cause of death, accuracy of information may also be compromised by a lack of knowledge/skill or by physician diagnostic error, due to diagnostic suspicion bias (i.e., when knowledge of the decedent's exposure to a putative factor influences the outcome of the diagnostic process).

ANACoD and ANACONDA offer guidance on the quality and plausibility of mortality data. Indicators that may reflect issues in the accuracy of information from which MCCD causes are derived include the following:

- Distribution of death by age and sex: deviations from expected patterns of age and sex distributions may indicate errors in age or sex information or selective bias in age-specific death reporting;
- Deaths labelled with codes not valid for underlying cause of death;
- Implausible sex/cause combinations, implausible disease/age combinations, or deaths due to diseases unlikely to cause death;
- Distribution of death by cause, age patterns of broad groups of causes, leading causes, and ratio of non-communicable to communicable causes: unexplained deviations from expected patterns may indicate errors in the accuracy of information; and
- Ill-defined causes by age/sex category: deaths classified to ill-defined causes are not useful for public health purposes; while a greater proportion may be expected (and

acceptable) in older age categories, high quality MCCD practices typically generate fewer than 10% ill-defined causes.

As previously noted, the above steps have been adapted for the VA context in the "Guidance for interpreting VA results" (D4H TWG, 2020), with the corresponding VIPER tool.[6] Additional indicators that may reflect issues in the accuracy of information from which VA causes are derived include the following:

- Average length of interview: The length of the VA interview will vary based on the age and sex of the decedent, reported symptoms, and the disposition of the respondent, among other factors. However, the majority of interviews should fall between 20-50 minutes[7] (Di Pasquale *et al.*, 2019). Accuracy of information should be questioned if the average interview length falls outside of this time frame.
- Period between date of death and date of interview: Recall bias has been shown to increase significantly after one year from the death. Accuracy of information should be questioned if the death to interview interval exceeds 1 year (Hussain-Alkhateeb *et al., 2016;* Serina *et al.,* 2016).
- Consistency between narrative and symptoms reported: Where narratives are collected, a sample of VAs can be reviewed (e.g., 10%) to check if key symptoms reported as present in the narrative are also recorded in the closed section of the VA interview. Significant inconsistency likely indicates issues with the accuracy of information reported.
- Item response patterns: The VA data can be assessed for item response patterns that may be indicative of inaccurate or uninformative data. Examples include: if a large percentage of responses are recorded as "don't know", there are inconsistencies between items where consistent responses would be expected, or there is an implausible lack of variability in yes/no responses. An analysis of more than 20,000 VAs using the 2016 WHO VA questionnaire found that 90% of items had fewer than 13% "don't know" responses, suggesting that VA items would ideally have fewer than 10-15% "don't know" responses (openVA Team, 2020). Further guidance on such assessment is currently under development.
- Percentage of undetermined causes of death: Related to deaths classified to ill-defined causes for MCCD, deaths classified as undetermined for VA are not useful for public health purposes. As high fractions of undetermined CODs can impact COD patterns,

[6] https://data4healthlibrary.org/resources/viper; the alpha version of this tool is compatible with SmartVA; additional work is underway to extend its compatibility to output from other VA algorithms.

[7] Estimates based on use of the 2016 WHO VA standard instrument.

datasets would ideally have fewer than 10-20% of deaths classified as undetermined (WHO, 2014; D4H TWG, 2020; Adair *et al.,* 2020).

- For any observed data quality issues, an attempt should be made to explain the possible reasons for the issues. If significant flaws are identified, and a lack of quality in the data is the suspected explanation, the data are likely not appropriate for aggregation. If the level of identified error is considered tolerable, the error levels can be considered again when interpreting the results (see Section 4), as a means to identify the likely directionality of potential bias. The development of VA data management tools is underway, and these tools can assist in monitoring data quality.[8]

If based on the above listed criteria the available datasets are considered of sufficient quality, the data can be prepared and aggregated following steps described in Section 3.

[8] VMan: https://www.ihi.or.tz/our-projects/project/126/details/; https://41.188.151.156; VA Explorer: https://github.com/VA-Explorer/va_explorer

SECTION 3: PREPARING THE DATASETS AND AGGREGATING DATA

The following steps describe the key data management preparations and decisions that need to be made to aggregate VA and MCCD data. A worked example of the process is described throughout this section for demonstration purposes.

1. Prepare the datasets

Standardize the variables

The VA and MCCD datasets should both contain the following variables: unique identifier, exact age (to allow for calculation of early-, late-, and post-neonatal periods during analysis), sex, ICD-10 (3 or 4 digits), ICD-11,[9] or VA code (e.g., code from WHO VA cause list or mapped ICD code from SmartVA cause list), and source of data/cause list (e.g., MCCD, VA-WHO, or VA-SmartVA). Make sure that these variables are standard across the two datasets. A geographic variable (e.g., place of usual residence) may be of interest for analysis purposes, but it is not required for combining purposes.

Address overlap between the VA and MCCD populations

While there may be cases where individuals have both a VA and MCCD record, when aggregating datasets, each decedent should be represented in only one dataset. Drawing from information compiled about the system design, information flow, and representativeness of the data in Sections 1 and 2, assess if there is potential for overlap in the datasets. For example, if the system is designed such that VAs are conducted for dead-on-arrival cases at a hospital, but there is also potential that a physician has completed a medical certificate of cause of death in such cases, there are likely decedents represented in both datasets.

Using a common unique identifier in the VA and MCCD datasets, search the two datasets for duplicate records. For any records present in both VA and MCCD datasets, the MCCD record should be used, unless the MCCD cause is undetermined or ill-defined/unknown (R00-R99). In that case, the VA record may be used. Document the extent of duplication and how duplicate records were resolved.

[9] Icd.who.int

2. Select a common cause list.

To aggregate VA and MCCD data, a decision must be made as to how the causes available from VA will be aligned to the ICD-coded causes from the MCCD data. Both datasets must have the same ICD structure and detail. Select cause lists commonly used in reporting cause of death statistics are included in Appendix B. These lists include a mapping of categories of causes to and from the relevant ICD codes. Three possible methods for alignment are presented below (lists are not in any order of preference).

The Global Burden of Disease Cause List -- Level 1, Broad Groups

The Global Burden of Disease (GBD) Cause list presents an internationally-comparable grouping across broad disease categories:

- Group I:[10] Communicable diseases (e.g., TB, pneumonia, diarrhoea, malaria, measles), maternal and perinatal causes (e.g. maternal haemorrhage, birth trauma) and nutritional conditions (e.g., protein-energy malnutrition)
- Note: an expanded version of this list can list communicable diseases and nutritional conditions[11], maternal[12], and perinatal causes[13] separately.
- Group II:[14] Non-communicable diseases (e.g., cancers, diabetes, heart disease, stroke)
- Group III:[15] External causes of mortality (e.g., accidents, homicide, suicide)

The distribution of deaths across these groupings can provide insight on the status of a country in relation to the "health transition." It can also be used to assess the plausibility of data by comparing observed patterns to the expected patterns according to life expectancy (WHO, 2021). Lists may be color-coded to indicate the causes that fall into these broad groups (see an example in the arrow diagrams in Figures 3 and 4). Note that these groupings do not include causes that are considered "ill-defined," or insufficiently detailed to be of value for public health purposes (WHO, 2021).

The VA Cause List

The most conceptually straight-forward way to aggregate VA and MCCD deaths is to aggregate the deaths according to the ICD-categories that comprise the VA list. With this approach to aggregation, each death in the MCCD dataset is assigned to the VA cause according to the corresponding ICD categorization. The final tabulation from this approach will reflect the VA cause list, with ranges of codes associated with each cause. The VA cause lists associated with the available methods for automatically assigning cause of death are included in Appendix B—the 2022 WHO Verbal Autopsy Cause list is represented in Appendix B-1, and Appendix B-2 includes the SmartVA cause list (see Adair *et al.,* 2020 for more details

[10] ICD-10:A00-B99, G00-G04, N70-N73, J00-J06, J10-J18, J20-J22, H65-H66, O00-O99, P00-P96, E00-E02, E40-E46, E50, D50-D53, D64.9, E51-64

[11] ICD-10:A00-B99, G00-G04, N70-N73, J00-J06, J10-J18, J20-J22, H65-H66, E00-E02, E40-E46, E50, D50-D53, D64.9, E51-64

[12] ICD-10: O00-O99

[13] ICD-10: P00-P96

[14] ICD-10:C00-C97, D00-D48, D55-D64 (except D64.9) D65-D89, E03-E07, E10-E16, E20-E34, E65-E88, F01-F99, G06-G98, H00-H61, H68-H93, I00—I99, J30—J98, K00-K92, N00-N64, N75-N98, L00-L98, M00-M99, Q00-Q99

[15] ICD-10: V01-Y89

on use of the SmartVA cause list). As this approach maps the "higher resolution" MCCD cause list to the "lower resolution" VA cause list, it maintains consistency with the notion that VA is a blunt instrument for determining probable cause of death at the population level. Using the VA cause list to aggregate VA and MCCD data will avoid the potential to extrapolate more detailed information from the VA data than the VA instrument is designed to provide. However, the level of detail in the cause list will be limited to that of the VA instrument, and additional cause information gained from the MCCD process will be lost.

The MCCD Tabulation Lists

Aggregating VA and MCCD data across MCCD causes maintains the detail provided by MCCD cause assignment, optimizing the information available in the MCCD dataset. However, given that many VA cause categories really represent more than a single cause/ICD code, standard MCCD tabulations need to be modified in some way to address these inherent incompatibilities. An example tabulation aligning the 2022 WHO VA cause list with the WHO GHE cause list is provided in Appendix B-1 to demonstrate a base tabulation to which each VA cause can be mapped; Appendix B-3 includes additional details of the GHE cause list for reference. Recommended modifications to accommodate VA causes are also described. Modifications include distributing VA causes across the range of ICD causes that they represent or collapsing a range of ICD causes to a broader cause category that is consistent with the VA cause. Other commonly used ICD tabulation lists can be similarly adapted. Updated cause of death lists, including WHO-recommended modifications to aggregate VA and MCCD data across MCCD causes, will be made available via the WHO and GHE websites as they are developed.[16] It should be noted that dengue fever is readily identifiable by VA, though it is not typically listed as a discrete cause category in standard tabulations. Where dengue fever is of particular interest, countries can consider adapting their tabulation list to include this cause.

Code the deaths with the selected cause list category

Once a common cause list has been selected, deaths from each method of collection (i.e., VA and MCCD) should be coded into the selected cause list categories within their respective datasets. This step will require loading a mapping file with the range of ICD codes included in each category, and then allocating each death to one of the categories in the selected tabulation list. Annotated statistical code is provided in Appendix C, and an electronic calculation tool is available to map the deaths to the GBD, GHE, or VA tabulation lists. If an infant/child-specific cause list is to be used (e.g., versus a general mortality list), a variable may

[16] Verbal Autopsy Standards: https://www.who.int/standards/classifications/other-classifications/verbal-autopsy-standards-ascertaining-and-attributing-causes-of-death-tool
Global Health Estimates: https://www.who.int/data/global-health-estimates

need to be added in each dataset to designate the records to which the child-specific cause categories should be applied.

3. Weight the de-duplicated VA and MCCD data to adjust for the non-representativeness of the datasets

Inverse probability weighting can be used to adjust the VA and MCCD datasets for missingness, accounting for place of death (i.e., community vs. hospital/facility), the age and sex structure of the population. Such adjustments will require additional inputs as described in Table 3.

Table 3: Inputs required to adjust for non-representativeness in data

#	Input	Suggested source
1	Standard population structure (i.e., mid-year population for 5-year age groups by sex) representative of the area(s) where the datasets were generated, as estimated from census counts and projections[17]	National statistics office or UN's World Population Prospects[18]
2	Current (e.g., annual, as available) age- and sex-specific death rate (ASDR) for the respective population	National statistics office or UN's World Population Prospects[5] for national estimates; official or other surveillance site estimates for subnational estimates, as available
3	Estimate of the ratio (or probability) of deaths occurring in the community (for the VA dataset) versus in hospitals/facilities (for the MCCD dataset)	Health officials

Note that to the extent possible, these inputs should reflect the area from which the datasets were generated. That is, if the datasets represent a subnational area, the inputs should reflect the same subnational area. Given that these analytic methods are being applied to strengthen mortality data where they are otherwise known to be weak, it is recognized that these inputs

[17] For flexibility in analysis options, it is recommended to include calculations for age groups for 0 and 1-4 years; age groups can be combined in subsequent steps as desired.

[18] https://population.un.org/wpp/

will likely be estimates. In selecting the best input data source, consider the various strengths and limitations that characterize each source of data, and document the known limitations of the selected source and describe the potential resulting bias. As an example, in the steps below, we reference an example using 2019 data from Lusaka, Zambia. The mid-year population and ASDR estimates were provided by the Zambia Statistics Agency. Approximately 30% of the deaths occur at home and about 70% occur in health facilities. Limitations of these input data include the following:

- The mid-year population and ASDR estimates are projected from Zambia's 2010 census, and the 9-year-old projections may not accurately reflect the actual population dynamics at the time the VA and MCCD data were collected; and
- The lowest age category of the ASDR is "under 1 year," so while a "neonates" (under 28 days) category is possible for the VA and MCCD data, the lowest category for key tabulations of CSMFs will be "under 1 year."

The example above offers a practical method for utilizing readily available statistical information to adjust for non-representativeness of data. When selecting factors and inputs to adjust for non-representativeness, caution should be taken to ensure that weighting does not result in over-representation of the data and that adjustment factors represent independent, non-overlapping characteristics of the population. Further resources for estimating completeness are available (Rao *et al.*, 2020).

Calculate the estimated number of deaths, probability of inclusion, and inverse probability weights

The population structure data (mid-year population for 5-year age groups by sex), along with the ADSR by sex and ratio of community to hospital/facility deaths, will be used to calculate the estimated VA and MCCD deaths in each age-/sex-specific category. Together with the observed deaths, you can then calculate completeness and the inverse probability weights to adjust for incompleteness. To do so, enter the population structure and ASDR into a dataset, as shown in Columns A-C in Table 4.

Table 4: Example of dataset with population structure to calculate expected deaths and weights for adjustment

Calculating Inverse Probability Weights for Female VA and MCCD Deaths by Age-Group--Lusaka, Zambia, 2019											
Age Group	Female Mide-Yr Pop	Female ASDR	Total Female Expected Deaths	Expected * (0.289) Female VA Deaths	Expected * (0.711) Female MCCD Deaths	VA Observed Female Deaths	MCCD Observed Female Deaths	VA Completeness	MCCD Completeness	VA Weights	MCCD Weights
A	B	C	D=BxC	E =D x 0.289	F=D x 0.711	G	H	J=G/E	K=H/F	L=1/J	M=1/K
0	49258	0.067	3295	952	2343	172	983	18%	42%	5.54	2.38
1-4	154164	0.012	1804	521	1282	74	318	14%	25%	7.04	4.03
5-9	163360	0.003	441	127	314	20	75	16%	24%	6.37	4.18
10-14	141305	0.002	311	90	221	18	81	20%	37%	4.99	2.73
15-19	135307	0.004	501	145	356	31	125	21%	35%	4.67	2.85
20-24	136214	0.006	831	240	591	40	196	17%	33%	6.00	3.01
25-29	130854	0.009	1112	321	791	66	294	21%	37%	4.87	2.69
30-34	115306	0.012	1372	397	976	56	318	14%	33%	7.08	3.07
35-39	87865	0.013	1133	328	806	82	451	25%	56%	3.99	1.79
40-44	59106	0.016	916	265	651	68	359	26%	55%	3.89	1.81
45-49	37213	0.015	566	163	402	56	317	34%	79%	2.92	1.27
50-54	23143	0.017	391	113	278	43	249	38%	90%	2.63	1.12
55-59	18196	0.020	364	105	259	51	266	48%	103%	2.06	0.97
60-64	11781	0.027	322	93	229	41	237	44%	104%	2.27	0.96
65-69	7137	0.036	253	73	180	56	236	76%	131%	1.31	0.76
70-74	4742	0.048	225	65	160	61	253	94%	158%	1.07	0.63
75-79	2930	0.064	189	55	134	85	248	156%	185%	0.64	0.54
80+	4175	0.089	372	108	265	143	307	133%	116%	0.75	0.86
Total	**1277882**		**14399**	**4161**	**10237**	**1163**	**5313**	**28%**	**52%**	**3.58**	**1.93**
Missing						1	86				

Source: Zambia Statistical Agency, 2019

Table 4 shows an example of the inverse probability weight calculations for females from Lusaka, Zambia.[19] Each row represents a separate age-specific category (for females) as defined in Column A of Table 4. The corresponding mid-year population and age- (and sex-) specific death rates (ADSR) for each of these categories are listed in Columns B and C, respectively. The expected number of total deaths by sex is calculated by multiplying the mid-year population by the ADSR for each age/sex category (Column D). The expected number of VA and MCCD deaths is then calculated by multiplying the total deaths by the factor that represents the percentage of community deaths (Column E) and health facility deaths (Column F). The observed VA and MCCD deaths are listed in Columns G and H, respectively, and probabilities of inclusion for both VA and MCCD are then calculated in Columns J and K by dividing their observed deaths by their expected deaths (WHO, 2010).[20]

[19] Appendix A provides Lusaka system's context and data prerequisites used for this referenced example.

[20] Alternative methods to calculate completeness may be used, including: 1) U.S. Census RUP Software, cohort component projection for mortality to estimate the number of total deaths

If the average probability of inclusion is greater than 95%, no adjustment needs to be made, and therefore, no weights need to be calculated (D4H TWG, 2020; Adair *et al.,* 2020). However, if the average is less than 95%, the weights that will be used to adjust for VA missingness across age categories by sex are calculated in Column H by taking the inverse of the completeness.

Calculate the adjusted CSMFs

The weights are then applied to the observed VA and MCCD frequencies to get the weighted frequency of deaths due to a given cause in the MCCD and VA areas separately. For each age/sex category, multiply the number of deaths for each relevant cause category by the weights for VA and MCCD, respectively, as shown for female infant deaths in Table 5, Columns C and E. In this example, the VA and MCCD weights for female infant deaths are 5.54 and 2.38, respectively.

Add the VA and MCCD weighted frequencies together (rounded to the nearest whole number) for each cause in the selected cause list (Column F). Calculate the total weighted CSMF for each cause by dividing the total weighted frequency by the sum of the weighted frequencies across all causes (Column G). The calculations can then be summed across all age groups for each sex to get an overall CSMF by sex, or across all age/sex groups for a CSMF representing the total population.

(https://www.census.gov/data/software/rup.html) or 2) Adair & Lopez 2018 *Estimating the Completeness of Death Registration: An Empirical Method.*

Table 5: Example table for calculating weighted frequencies

VA and MCCD Cause of Death Frequendes, Weighted Frequencies, and Weighted CSMFs for Female Infants (Age 0) Decedents-- Lusaka, Zambia 2019						
Cause	VA Frequency	VA Weighted Frequency	MCCD Frequency	MCCD Weighted Frequency	Total Weighted Frequency	Total Weighted CSMF
A	B	C = B x	D	E = D x	F = C+E	G = F/ sum of total weighted frequency
Neonatal tetanus	8	44	164	391	435	13.2%
Birth asphyxia	6	33	147	350	384	11.6%
Other and unspecified perinatal cause of deaths	0	0	159	379	379	11.5%
Congenital malformation	12	66	88	210	276	8.4%
Prematurity	13	72	40	95	167	5.1%
Unspecified infectious disease	23	127	9	21	149	4.5%
Severe malnutrition	4	22	49	117	139	4.2%
Diarrheal diseases	19	105	12	29	134	4.1%
Sepsis	0	0	33	79	79	2.4%
Acute Resiratory infection, including Pneuminia	0	0	30	72	72	2.2%
Meningitis and encephalitis	8	44	8	19	63	1.9%
Neonatal pneumonia	7	39	9	21	60	1.8%
Other and unspecified disease	2	12	17	41	51	1.5%
Other and unspecified NCD	0	0	20	48	48	1.4%
Epilepsy	7	39	1	2	41	1.2%
Macerated stillbirth	7	39	0	0	39	1.2%
HIV/AIDS related death	2	11	11	26	37	1.1%
Severe anaemia	0	0	9	21	21	0.7%
Accidental exposure to smoke, fire and flames	0	0	5	12	12	0.4%
Road traffic accident	1	6	1	2	8	0.2%
Malaria	0	0	2	5	5	0.1%
Severe anaemia	0	0	2	5	5	0.1%
Cause of death unknown	53	293	167	398	692	21.0%
Total	**172**	**952**	**983**	**2343**	**3295**	

Notes / Assumptions: This is a simple approach to the aggregation of the MCCD and VA dataset. The example provided above reflects practical adjustments that can be made using data that are likely to be readily available. In this document, we recommend the use of age-/sex-specific death rates to determine the expected number of deaths across the age categories of interest, by sex, given the interest in cause-specific mortality information and considering that causes vary by age and sex. In the absence of age- / sex-specific death rates, a single crude death rate could be considered for use, though results would likely be biased to an unknown extent.

This approach assumes that age, sex, and location of death factors are independent and that the distributions of causes of death are the same for the deaths that are covered and the deaths that are not covered. Other factors affecting completeness of the data can be adjusted for if they are independent from each other and if sufficient data are available to calculate the needed weight.

SECTION 4: DATA USE AND INTERPRETATION

The general aim of this draft document is to equip users with the necessary tools to put VA data in the hands of policy makers in order to support public health decision making and policy development. Specifically, this document provides an approach on the use of VA data alongside existing MCCD data. In LMICs where many deaths occur outside health facilities, VA is essential in determining causes of death, as it is currently the only alternative to medically certified cause of death (MCCD). Thus, though an imperfect tool, the use of VA data should still be a priority for decision-making, planning, and improving such health systems. Analyzing and interpreting VA data can help identify gaps in systems and increase the awareness of VA's value, thereby driving demand for scale-up. Health system strengthening using VA data can be achieved through buy-in from and active engagement of the government throughout the VA implementation process.

Box 2

The value of VA data in health policy and planning

As VA has become gradually standardized and more frequently applied in LMICs, some countries have discovered its value to inform health policy and planning. For example, in Ghana, approximately 70% of deaths occur outside of health facilities, and many of these deaths are not registered with the Births and Deaths Registry nor have a cause of death. The data on cause of death come mainly from health facility deaths, which provides an incomplete picture for essential health policy and planning, as these tend to be in more urban settings (and are reflective of causes of death more common in urban than in rural populations). As a result, VA was launched and is currently being rolled out in Ghana's Volta Region. Recent analyses on cause-specific mortality fractions using the pre-test VA results showed that road traffic accidents ranked 4th among the top 20 leading causes of death in Ghana. However, a recent ANACONDA analysis using national MCCD data from DHIS2 did not yield any road traffic deaths, marking a clear gap in mortality data. According to the World Health Organization, Ghana ranks 31 as one of the countries with the highest rates of road

traffic deaths globally, highlighting VA's importance to identifying gaps and health priorities in-country.

When deaths occur in medical settings and causes of death are reported according to international standards, it is possible to calculate age, sex and cause-specific mortality rates if appropriate population denominators are available (and hospital deaths can be classified by place of residence). However, this is unlikely to be the case in settings where deaths in hospitals are the exception rather than the rule. Therefore, hospital-based mortality data in such settings are often reported as leading causes of hospital deaths or in terms of institutional case fatality rates rather than in terms of cause-specific mortality rates. Mortality data from VA are reported as CSMFs. The common practice is to report on the 10 or 20 leading causes of death in the population as a percentage of total deaths. With these considerations, the steps below provide an approach on tabulating aggregated VA and MCCD data, interpreting findings, and other suggestions for using this valuable mortality data source.

A. National reporting practices

Many countries that are beginning to incorporate VA results into their national vital statistics processes are also beginning to publish national vital statistics reports for the first time. While mortality data may be incomplete, it is recommended that what is available be published to increase attention on data quality, demand for improvement, and to promote the value of mortality data for public health purposes. Guidance on developing a national vital statistics report is available in the "Production of a Vital Statistics Report: Guide" (Vital Strategies, 2020). When VA and MCCD data are both available, data should first be presented separately by source (e.g., MCCD vs VA 20 leading causes) to review the plausibility of the results. Interpretation of differences should account for known systemic differences that are likely to impact the cause distributions (e.g., injury-related deaths where medical care is likely to be sought are more likely to appear in the MCCD data). If they are of sufficient quality, the data can then be aggregated as described above for a single mortality profile. Where VA and MCCD data are only available at a subnational level-- as in the case of the Lusaka District example described in Section 3-- the VA and MCCD data could be presented in a separate chapter of the broader national report or in an additional special report.

B. Description of core tabulations

Consult section 5 in Volume 2 of the International Statistical Classification of Diseases and Related Health Problems, 10th Revision[21] for guidance on regulations regarding statistics for international comparisons and on data presentation in national and subnational statistical tables. While the present document focuses on how to aggregate MCCD and VA data, the MCCD and VA data should also be tabulated and presented separately, to facilitate a more thorough understanding of the dynamics impacting mortality in a given country.

Only three-character rubrics of the ICD should be used when aggregating VA and MCCD data. If the "VA cause list" is selected for aggregation, the core tabulation will reflect the VA cause list associated with the VA cause of death assignment method. If an "MCCD tabulation list" is selected for aggregation, the team will need to select the appropriate level of detail. For example, the GHE cause list comprises four levels of detail (see Appendix B-2). Level 2, which includes 23 cause categories would be suitable for most aggregated datasets. Levels 3 or 4 might be considered where larger datasets are available, perhaps that include data pooled over several years or nationally representative data with high completeness. The "selected" special tabulation lists for mortality, as described in ICD-10 Volumes 1[22] and 2[13], ICD-11 browser[23], or Global Burden of Disease[24] cause lists may also be suitable, provided that the necessary adaptations are made to accommodate the VA cause categories (see GHE example adaptations in Appendix B-2).

Countries may elect to adapt these tabulation lists by omitting certain cells or rows of the tabulations where no cases occur, or many rows are empty. As recommended in ICD-11 reference guide and ICD-10 Volume 2, "when only the occasional case of a disease occurs in a country, the line can be regularly omitted from the published table and a footnote added to indicate either that there were no cases or, when sporadic cases do occur, in which cell the case would have appeared." Furthermore, "for cells with very low frequencies, especially those relating to diseases that would not be expected to occur, it is important to establish that the cases existed and did not result from a coding or processing error. This should be carried out as part of the general quality control of the data."[5] In countries and areas with small population numbers, data may be aggregated over a three- to five-year period and averages calculated to overcome unstable fluctuations that are likely to result from the small numbers.

[21] https://icd.who.int/browse10/Content/statichtml/ICD10Volume2_en_2016.pdf
[22] https://apps.who.int/iris/bitstream/handle/10665/246208/9789241549165-V**1**-eng.pdf
[23] https://icd.who.int/browse11/l-m/en
[24] http://ghdx.healthdata.org/record/ihme-data/gbd-2019-cause-icd-code-mappings

C. Interpreting findings

There are a number of bodies within the government and partner institutions that are part of the CRVS governance structures and would be interested to participate in the interpretation of findings. The structure, tasks, and operating procedures of such bodies or committees will vary from country to country, and who to engage will be highly context specific. National CRVS coordination committees and mortality coordination committees are among these bodies and would likely play an essential role in overseeing the interpretation of the findings and unpacking the results of the analysis (de Savigny, et al., 2017). Detailed guidance on vital statistics reporting and interpreting MCCD and VA data are beyond the scope of this document and can be referenced elsewhere (Vital Strategies, 2020; D4H TWG, 2020). However, some important considerations and limitations particularly relevant to aggregating these two data sources are described below.

C.1. What to do with odd, unexpected, or discrepant findings

An attempt should be made to explain odd, unexpected, or discrepant findings, including differences in cause distributions between MCCD and VA deaths. Such findings may be explained by factors related to the context of the specific MCCD and VA reporting systems (as described in Section 1). For example, deaths related to motor vehicle accidents (MVAs) may have been seen at the hospital but were referred to the police or medical examiner for investigation, and the results were not captured in the MCCD. In such systems, VA data are likely to have a higher fraction of deaths due to MVAs than the MCCD data. Such findings may also be attributed to observed data quality issues (as described in Section 2). The impact of any observed data quality issues, including the directionality of potential bias in results (e.g., a high percentage of “don’t know” or “refused to answer” for a key symptom in the VA interview would likely result in under-reporting/under-representation of its associated cause in the overall CSMF). Major events or programmatic changes (e.g., extreme weather events or introduction of a new vaccine) that are likely to impact mortality patterns may also result in odd, unexpected, or discrepant findings. Finally, changes to the data collection, analytical, or reporting methods, including processes used to assign and tabulate cause of death, are also likely to impact findings. Such changes should always be documented in technical notes, and potential implications described. The explanations for odd, unexpected, or discrepant findings should be reviewed for opportunities for corrective action.

C.2. Limitations

A description of the known and/or potential limitations of the data collection and analysis procedures should be included in any reporting, to aid the reader in interpreting the results correctly. A few factors that are likely to affect MCCD and VA data are described below.

Limits of the reduced VA cause list

While aggregating MCCD and VA data into a single mortality profile may be helpful for observing trends and policymaking at higher levels, disease-specific programs requiring more detailed cause-specific mortality information should consider consulting additional data sources. Where the aggregated cause list is based on the reduced list of VA causes, residual categories (e.g., other cancers, other non-communicable disease) are not likely to provide sufficient programmatic detail. Data at this level of analysis are not designed to provide the needed level of specificity. At the same time, as already mentioned, there is also potential for misclassification if VA causes are assigned to a single ICD code, where VA causes really represent a range of ICD codes. The following sources of data may provide further cause-specific detail:

- Morbidity data from hospitals that provide information on the diseases presenting at hospitals; specific mortality surveillance and program data such as from maternal/perinatal death notifications, and registries for cancers, malaria, HIV/AIDS, diabetes and tuberculosis;
- Cause-of-death distributions from Health and Demographic Surveillance (HDSS) or other surveillance sites;
- Periodic household surveys such as Demographic and Health Surveys (DHSs) or maternal mortality surveys; and/or
- GBD Compare / GHDx website for GBD data (https://vizhub.healthdata.org/gbd-compare/).
- Non-representativeness of deaths included in the datasets.

Sections 2 and 3 include steps to review the expected population structure and the population observed in the available datasets. These steps should be used to estimate and describe who is likely captured and missed in the mortality reporting systems. Observed discrepancies should be noted in the limitations with comment on the extent of the resulting systematic selection bias (i.e., distortions in the findings that result from factors related to whether or not records are included in the dataset for analysis). The level of confidence in the available population data that have been used for comparison (e.g., time since census), should be considered when describing the extent of and sources of non-representativeness of the data. See Table 1 for more information about factors that may contribute to non-representativeness.

Potential for misclassification of causes of death

In MCCD and PCVA, misclassification is a systematic error due to intentional or unintentional bias of the physician in assigning a cause of death. Where resources are available, an ad hoc validation exercise could be undertaken to estimate the extent of misclassification or under-reporting (Rao *et al.,* 2010; Bradshaw *et al.,* 2020). In VA where cause of death is assigned by an automated algorithm, misclassification is a systematic error due to incorrect logic or symptom-cause information of the algorithm. It is recognized that misclassification of cause of death is likely differential, where the potential for or extent of misclassification varies by cause (Polprasert *et al.,* 2010). For example, causes associated with specific symptom patterns are likely to be more accurately assigned than those associated with non-specific patterns, as is commonly seen among co-morbidities in the elderly. In another case, some causes are likely to be captured differently by MCCD compared to VA, as with injury-related deaths, which are likely to be assigned an immediate (e.g., bleeding) or intermediate cause (e.g., type of trauma) in the MCCD system, while VA assigns the probable underlying cause of the injury (e.g., road traffic, fall, burn, drowning); this may result in a higher number of injury-related deaths appearing in the VA data compared to the MCCD data. While work is underway to better describe such patterns, knowledge of the mortality reporting systems and associated weaknesses, and familiarity with clinical and epidemiologic patterns of various causes may be drawn on to attempt to describe the extent and directionality of misclassification. See Table 1 for more information about factors that may contribute to misclassification.

C.3. Confidence intervals

The analytic procedure outlined in Section 3 yields adjusted CMSFs. It does not, however, quantify the uncertainty of the CSMF estimates. It is also unable to take into account expert knowledge that does not fit into the adjustment framework, such as intentional or unintentional COD misclassification. Addressing these shortcomings would require development of a reference Bayesian model of MCCD and VA data. The main advantages of a Bayesian approach include the following:

- Confidence intervals produced for all estimates;
- Integration of expert knowledge, such as known survey biases or known patterns of COD misclassification, in the form of priors;
- The ability to express the uncertainty around key parameters such as the population structure, and to propagate this uncertainty all the way to the CSMF estimates; and
- The possibility to improve the estimation using secondary data relating to these key parameters inside the same model.

However, unlike our adjustment procedure above, such a model will require a user-friendly interface to be useful for the practitioners. Further work in this area is recommended.

D. Other uses of data

Many opportunities for using and applying VA data to improve population health will emerge as countries continue to develop and implement data collection. VA data can be applied alongside MCCD data strategically to establish priorities, target resources, develop legal and regulatory initiatives, and plan programs to improve public health (Thomas, *et al.,* 2018). Frequently, however, data collected by public health agencies and ministries of health are in danger of being unused; existing policymaking processes often unfold without the benefit of data and evidence, where opportunities for rigorous assessment of potential impacts and costs of various options are not realized. Given the current efforts in implementing and expanding VA data collection, efforts should be undertaken to ensure these data are used to enhance public health policymaking and decision-making, alongside other mortality data sources where appropriate. Further details are provided below on how aggregated VA and MCCD data can be incorporated into policy briefs, trend comparisons, international development reporting mechanisms (e.g., Sustainable Development Goals (SDGs)), and national health information systems (e.g., DHIS2).

D.1. Policy brief

Policymaking can be enhanced by leveraging VA data alongside other mortality data analysis to develop policies with high-yield and cost-effective policy recommendations. VA data can be used to formulate policy questions and identify policy options, in conducting health impact and cost-effectiveness analyses, and in visualizing data. VA data can also be used to inform government actions to inform health-related laws, such as those mandating seatbelt use. Policies aimed at reducing deaths due to road traffic crashes can especially benefit from utilizing VA data because these deaths are usually not well captured by traditional mortality reporting/surveillance processes. Additionally, VA data can be used to support regulations set by government agencies, including those aimed at reducing maternal and child mortality. Lastly, VA data could be critical for resource allocation such as increasing funding to reduce mortality from high-burden infectious or non-communicable diseases, as well as to develop new, needed public health programs (Thomas, *et al.,* 2018).

Impactful policymaking requires not only a conducive political environment and advocacy but an appropriate document that can be shared with key stakeholders and policymakers. The best practice is to summarize data and analyses in evidence-based, data-driven policy briefs that tell compelling stories for why change is needed, providing support for specific, high-impact

intervention strategies. A 3- to 4-page policy brief that states the health problem, discusses the costed policy options, and provides feasible recommendations on how to address the health problem is one of the most appropriate documents used to target policymakers. Additional details on developing policy briefs with mortality data can be found in the "Guidance for interpreting VA results" (D4H TWG, 2020). Staff responsible for collecting and analyzing VA and MCCD data could work closely with government epidemiologists and policy analysts to create policy briefs using the aggregated mortality data on an ongoing basis to address public health priorities. Health programs or government public health agencies could then share briefs with decision makers, who could in turn utilize them as supporting documents to address the specific health problem.

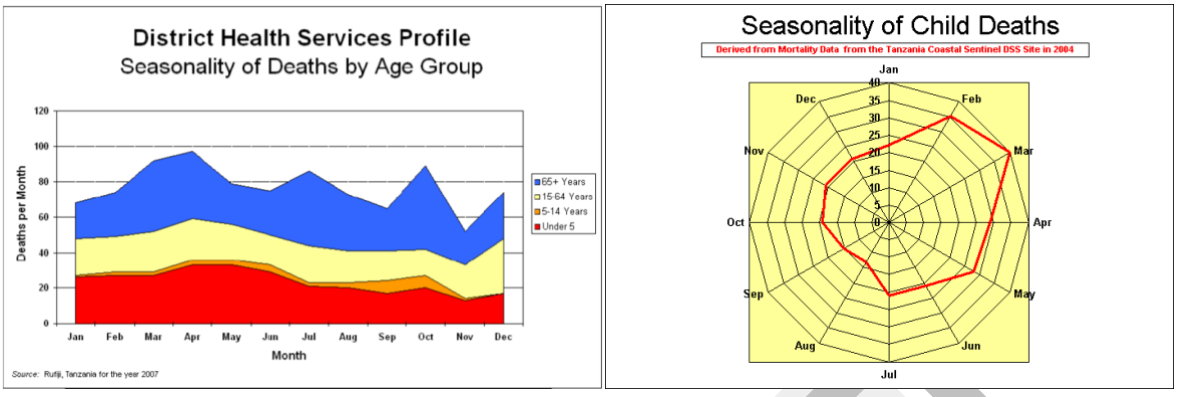
D.2. Trend comparisons

Trend comparisons are one of the most informative analyses for decision makers for health planning and priority setting, as they show change in a health outcome over time. For example, observing changes in CSMFs for the most common causes of death can help determine which health interventions to prioritize. While cause-specific mortality rates cannot be derived from VA data for trend comparisons where information on the population denominators is not available, trends in mortality numbers and relative cause fractions, disaggregated by age, sex, and geographic location, can be visualized in a variety of ways to support further interpretation of the data.

Without conducting deeper analysis, mortality data over time can be visualized through various types of graphs, including scatter plots, line graphs, and arrow diagrams, to determine if the trend is increasing, decreasing or fluctuating. Mortality numbers can be observed by various age, sex, seasonality, or other stratifications to better understand the trends in the data. For example, Figure 2 shows monthly patterns in mortality by age group, using VA data from Demographic Surveillance System sites in Tanzania. Such a figure can be used to identify if the mortality patterns in some age groups are affected by seasonality — that is, are there larger numbers of deaths due to a specific cause during certain months of the year? IHME's Vizhub[25] is an excellent resource that provides examples of these three types of visualizations over time.

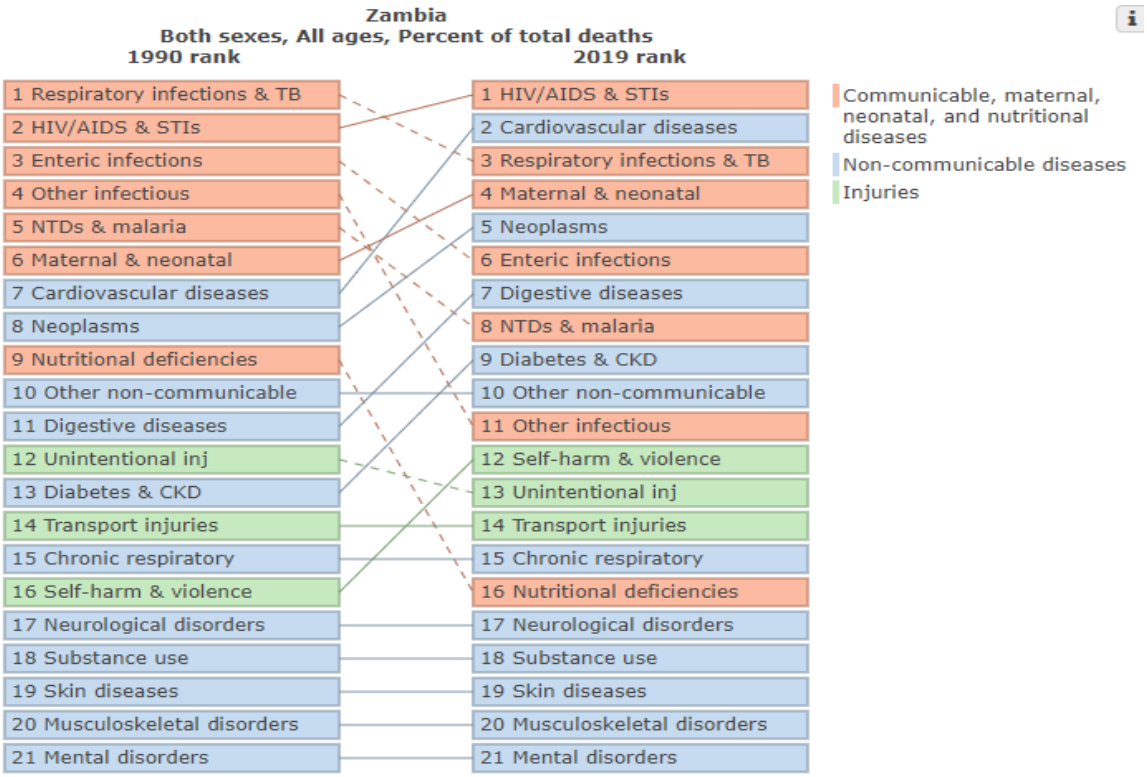
[25] Institute for Health Metrics and Evaluation (IHME). GBD Compare. Seattle, WA: IHME, University of Washington, 2015. Available from http://vizhub.healthdata.org/gbd-compare. Accessed: August 7 2019.

Figure 2. Seasonality in mortality data based on DSS sites in Tanzania



An arrow diagram is a useful visualization for displaying cause of death data over time, particularly CSMFs. In an arrow diagram, causes of death are ranked from the highest to the lowest for one year or a given time period and then compared to rankings from another year or given time period, usually 10 or more years apart. An arrow diagram comparing rankings of deaths by cause for all ages from 1990 to 2019 in Zambia is shown in Figures 3, as modeled within the Global Burden of Disease. Note the additional use of color to denote whether the causes are communicable, maternal, neonatal, and nutritional; non-communicable; or injuries.

Figure 3. Arrow diagram comparing the CSMF ranking of deaths in 1990 to 2019 by cause for all ages, in Zambia



Arrow diagrams are best used to show changes over longer time periods and can highlight progress in decreasing preventable deaths as well as visualizing shifts from communicable to non-communicable causes of deaths.

Visualizing mortality data over time is more meaningful when several years of data are available. Arrow diagrams can be used to visualize MCCD data or VA data or an integration of the two. Countries that are still in the process of collecting VA data or only have a few years of VA data could focus on visualizing their MCCD data instead. Any changes in data quality that have occurred over time, especially in completeness and the quality of cause of death codes, should be noted and taken into account when interpreting any changes in mortality over time. In situations where the number of deaths are small for a particular cause (i.e., 20 or less), analysts should consider combining single years into broader time periods, for example two- or three- year time periods, to reduce noise and facilitate better interpretation. Additional guidance on trend comparisons that can be made with mortality data is provided in Appendix D.

D.3. Reporting for international development

International demand for improved mortality and cause of death reporting is increasing, due in large measure to regional and global goals, targets and monitoring indicators, including the SDGs (Mills, *et al., 2017).* Eight of the SDG targets require information on deaths and causes of death (Table 6).

Table 6: Sustainable Development Goal targets that require mortality and cause-of-death data

Goal / Target	Definition
3.1	By 2030, reduce the global **maternal mortality ratio** to less than 70 per 100,000 live births.
3.2	By 2030, end preventable **deaths of newborns and children under 5 years** of age, with all countries aiming to reduce neonatal mortality to at least as low as 12 per 1,000 live births and under-5 mortality to at least as low as 25 per 1,000 live births.
3.3	By 2030, end the epidemics of **AIDS, tuberculosis, malaria, and neglected tropical diseases** and combat **hepatitis**, waterborne diseases, and other communicable diseases.
3.4	By 2030, reduce premature mortality from **noncommunicable diseases** by one-third through prevention and treatment and promote mental health and well-being.
3.6	By 2020, halve the number of global deaths and injuries from **road traffic accidents**.
3.9	By 2030, substantially reduce the number of deaths and illnesses from **hazardous chemicals and air, water, and soil pollution and contamination.**
11.5	By 2030, significantly reduce the number of deaths and the number of people affected and substantially decrease the direct economic losses relative to global gross domestic product caused by **disasters**, including water-related disasters, with a focus on protecting the poor and people in vulnerable situations.
16.1	Significantly reduce all forms of **violence** and related death rates everywhere.

Work is underway to provide guidance for calculating these indicators in countries with incomplete death registration and poor-quality COD data (Adair, *et al.*, 2021*).* SDG indicators 3.1, 3.2, 3.3, 3.4 and 16.1 are defined in terms of population ratios or rates; numerators are numbers of deaths due to each specific cause; denominators are total population at risk at a defined moment in time (usually the mid-year population). For infant, child and maternal mortality indicators, denominators are live births in the given year. By contrast, indicators for SDG 3.6, 3.9 and SDG 11.5 are couched in terms of numbers of deaths.

In view of differences in the outputs from VA compared to medical certification, they have traditionally been reported separately. In this document we describe approaches and considerations for combining the outputs of these two sources. Given that VA data are reported as CSMFs, analysis of hospital data combined with VA will only produce CSMFs rather than population cause-specific mortality rates. If VA has been conducted on a nationally representative sample of all deaths, or in localized areas where VA has been conducted on all deaths (e.g., in health and demographic surveillance sites), it would in principle be possible to transform the CSMFs into absolute numbers of deaths and then to population-based mortality rates. However, this would not be possible in settings where VA is implemented in selected sites that are not nationally representative. Furthermore, for reporting on maternal and child mortality rates, information would be needed on total live births in the population at a specified time period.

It may be possible to use VA data, or VA data in combination with medical certificate data, for reporting SDG indicators 3.6, 3.9 and 11.5 which refer to numbers of deaths due to specified causes in settings where VA is conducted on a nationally representative sample. However, for reporting on SDG indicators 3.1, 3.2, 3.3, 3.4 and 16.1, different ways of reporting on progress will need to be devised for countries reliant on cause of death derived from VA.

D.4.Monitoring the implementation of methods that generate cause-of-death data

The SDGs and other national and international goals and targets are generally focused on outcome measures. By contrast, the United Nations Economic and Social Commission for Asia and the Pacific (UNESCAP), in its Regional Action Plan monitors indicators of process, that is, the proportion of total deaths the causes of which are determined through medical certification or VA. The Action Plan states that:

Every death should have a medically certified cause associated with it. For statistical purposes, special measures, such as VA, may be needed to ensure that all deaths are associated with a defined cause of death, especially in settings where many deaths occur outside of health facilities and without attention from a medical practitioner. (UNESCAP, 2017)

The UNESCAP Goal 3 is that: "Accurate, complete and timely vital statistics (including on causes of death) are produced based on registration records and are disseminated." The targets associated with the goal are intended to be set by countries in accordance with their capacities. Target 3E refers specifically to the cause of death determined by way of VA (see Table 7).

Table 7: UNESCAP CRVS Regional Action Plan Goal 3 targets related to birth and death registration and causes of death

Target #	Definition
3A	By ...(year), annual nationally representative statistics on births – disaggregated by age of mother, sex of child, geographic area and administrative subdivision – are produced from registration records or other valid administrative data sources.
3B	By ...(year), annual nationally representative statistics on deaths – disaggregated by age, sex, cause of death defined by ICD (latest version as appropriate), geographic area and administrative subdivision – are produced from registration records or other valid administrative data sources.
3C	By 2024, at least … percent of deaths occurring in health facilities or with the attention of a medical practitioner have an underlying cause of death code derived from the medical certificate according to the standards defined by ICD (latest version as appropriate).
3D	By 2024, the proportion of deaths coded to ill-defined codes will have been reduced by …percent compared with the baseline year.
3E	**By 2024, at least … percent of deaths taking place outside of a health facility and without the attention of a medical practitioner have their underlying cause of death code determined through VA in line with international standards.**

3F	By … (year), key summary tabulations of vital statistics on births and deaths using registration records as the primary source, are made available in the public domain in electronic format annually, and within one calendar year.
3G	By … (year), key summary tabulations of vital statistics on causes of death using registration records as the primary source, are made available in the public domain in electronic format annually, and within two calendar years.
3H	By … (year), an accurate, complete and timely vital statistics report for the previous two years, using registration records as the primary source, is made available in the public domain.

WHO and the Health Data Collaborative have developed the *SCORE for Health Data* Technical Package designed to assist Member States in strengthening countries' health data systems and capacities to respond to the monitoring requirements of the health-related Sustainable Development Goals and other regional, national and subnational priorities (WHO-SCORE, 2018). In this initiative, the focus is on the ability of countries' statistical and/or health information systems to report on births and deaths registered and on causes of death from medical certification or from VA.

The overall target is that "Countries should have the capacity to report leading causes of death that account for large proportions and numbers of deaths in the total population, and within specified population groups, for recent time periods. Statistics on causes of death are best generated from the medical certification of cause of death according to the standards set out in the ICD. **Where this is not possible, VA can be used to estimate cause of death distributions in the population**."

The main outcome indicators for the 'C' (COUNT births, deaths and causes of death) component are completeness of birth and death registration and certification and reporting of cause of death. Input/process indicators relate to "*Core attributes of a functional system to generate cause-of-death statistics*" of which one of the components is VA. The *SCORE* indicators and monitoring framework also asks that countries report on the availability and quality of health services data under the heading 'O' (OPTIMIZE health service data**).** Two indicators relate to cause-specific mortality*, hospital deaths by major diagnostic category (ICD)* and *institutional maternal mortality ratios*. The indicators are assessed on the basis of availability at national and subnational level and the availability of disaggregations by age and sex. Ability to report the SDG cause-specific indicators is not mentioned.

CONCLUSION

To support the increased use of valuable mortality data, this document has offered an approach on how to aggregate mortality data from two sources-- MCCD and VA. The document emphasizes key factors to consider regarding the system context and minimum data standards, and it provides a step-by-step description of how to prepare and aggregate the different datasets. This description offers options for addressing common challenges in mapping mortality data to a common cause list and adjusting data for incompleteness. Finally, the document concludes with suggestions and considerations for reporting findings from these data sources, to support better use of the information. The information provided in this document is complementary to several other guidance documents that have been cited throughout this document. Such approaches will continue to evolve, as mortality data and its use continue to improve.

REFERENCES

Adair T, AD Lopez. "Estimating the completeness of death registration: An empirical method." PLoS One 13(5): e0197047 (2018).

Adair T, Firth S, Phyo TPP, Bo KS, Lopez AD. Monitoring progress with national and subnational health goals by integrating verbal autopsy and medically certified cause of death data. BMJ Glob Health. May;6(5):e005387 (2021). https://doi.org/10.1136/bmjgh-2021-005387

Adair T, Firth SM, Lopez AD. Integration of data from medical certification of cause of death and verbal autopsy. CRVS best-practice and advocacy. Melbourne, Australia: Bloomberg Philanthropies Data for Health Initiative, Civil Registration and Vital Statistics Improvement, University of Melbourne (2020). https://data4healthlibrary.org/sites/default/files/resources/131_UMelbourne_RFQ04251-BD4H-Integration_medical%20certification%20of%20cause%20of%20death%20and%20verbal%20autopsy.pdf

Bradshaw D, Joubert JD, Maqungo M, Nannan N, Funani N, Laubscher R, Cheyip M, Zinyakatira N, Awotiwon O, Nojilana B, Bezuidenhout F, Martin L, Dempers J, Kahn K, Price J, Lombard C, Morof D, Nichols E, Rao C, Groenewald P. South African National Cause-of-Death Validation Project: Methodology and Description of a National Sample of Verbal Autopsies. Cape Town: South African Medical Research Council (2020). https://www.samrc.ac.za/sites/default/files/attachments/2022-08/NationalCause-of-deathValidationReport.pdf

Cobos Muñoz D, de Savigny D. Rosetta stone for medical certification and verbal autopsy tabulations. 2.0 ed. Basel: Swiss Tropical and Public Health Institute (2018).

Clark SJ, McCormick T, Li Z, Wakefield J. InSilicoVA: A method to automate cause of death assignment for verbal autopsy. Working paper no. 133. University of Washington, Center for Statistics and the Social Sciences. 24 Aug. (2013). https://csss.uw.edu/files/working-papers/2013/wp133.pdf

D4H Technical Working Group (D4H TWG). Guidelines for interpreting verbal autopsy data. CRVS resources and tools. Melbourne, Australia: Bloomberg Philanthropies Data for Health Initiative, Civil Registration and Vital Statistics Improvement, The University of Melbourne (2020). https://getinthepicture.org/resource/guidelines-interpreting-verbal-autopsy-data

De Savigny D, Riley I, Chandramohan D, Odhiambo F, Nichols E, Notzon S, AbouZahr C, Mitra R, Cobos Munoz D, Firth S, et al. Integrating community-based verbal autopsy into civil registration and vital statistics (CRVS): system-level considerations. Glob Health Action. 10(1):1272882 (2017).

Di Pasquale A., Leitao J., Nichols E. Electronic data capturing for the WHO verbal autopsy instrument to facilitate routine cause of death monitoring. [PowerPoint presentation]. 11th European Congress on Tropical Medicine and International Health. Liverpool, UK (2019).
Institute for Health Metrics and Evaluation (IHME). GBD Compare. Seattle, WA: IHME, University of Washington, 2019. Available from http://vizhub.healthdata.org/gbd-compare. Accessed: July 21, 2021.

Mills SL, AbouZahr C, Kim JH, Rassekh BM, Sarpong D. Civil registration and vital statistics (CRVS) for monitoring the Sustainable development goals (SDGS). Washington, D.C., USA: World Bank Group (2017). http://documents.worldbank.org/curated/en/979321495190619598/Civil-registration-and-vital-statistics-CRVS-for-monitoring-the-Sustainable-development-goals-SDGS

openVA Team. Quantitative item response pattern results over full dataset. [Presentation]. WHO Verbal Autopsy Reference Group Virtual Workshop: Response Pattern Analysis for the Revision of the 2016 WHO Verbal Autopsy Instrument. Virtual Workshop (2020).
Polprasert W., Rao C., Adair T. et al. Cause-of-death ascertainment for deaths that occur outside hospitals in Thailand: application of verbal autopsy methods. Popul Health Metr. 8, 13 (2010). https://doi.org/10.1186/1478-7954-8-13

Rao C, Mswia R, Bratschi M, Setel P. Estimating Completeness of Birth and Death Registration: Methods and Options for Estimating Completeness of Civil Registration. Vital Strategies, NY (2020). https://www.vitalstrategies.org/wp-content/uploads/Estimating-Completeness-of-Birth-and-Death-Registration.pdf

Rao C, Porapakkham Y, Pattaraarchachai J, Polprasert W, Swampunyalert N, Lopez AD. Verifying causes of death in Thailand: rationale and methods for empirical investigation. Popul Health Metr. 8, 11 (2010). https://doi.org/10.1186/1478-7954-8-11

Serina P, Riley I, Hernandez B, Flaxman AD, Praveen D, Tallo V, Joshi R, Sanvictores D, Stewart A, Mooney MD, Murray CJ, Lopez AD. What is the optimal recall period for verbal autopsies? Validation study based on repeat interviews in three populations. Popul Health Metr. 14, 40 (2016). https://doi.org/10.1186/s12963-016-0105-1

Heron M. Deaths: Leading causes for 2018. Technical Notes. National Vital Statistics Reports; vol 70 no 4. Hyattsville, MD: National Center for Health Statistics. (2021). https://doi.org/10.15620/cdc:104186

Hussain-Alkhateeb, L., Petzold, M., Collinson, M. et al. Effects of recall time on cause-of-death findings using verbal autopsy: empirical evidence from rural South Africa. Emerg Themes Epidemiol 13, 10 (2016). https://doi.org/10.1186/s12982-016-0051-1

Thomas LM, D'Ambruoso L, Balabanova D. Verbal autopsy in health policy and systems: a literature review. BMJ Glob Health. 3(2):e000639 (2018). https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5935163/

United Nations Economic and Social Commission fo rAsia and the Pacific (UNESCAP). Asian and Pacific Civil Registration and Vital Statistics Decade 2015-2024. Bangkok, Thailand: UNESCAP (2015). https://www.unescap.org/sites/default/files/Asian_and_Pacific_Civil_Registration_and_Vital_Statistics_Decade2015-2024_Booklet.pdf

World Health Organization (WHO). ANACoD3: analysing mortality and cause-of-death. Geneva, Switzerland: WHO (2021). icd.who.int/anacod

World Health Organization (WHO). SCORE for health data technical package: essential interventions. Geneva, Switzerland: WHO (2020). https://apps.who.int/iris/handle/10665/334005.

World Health Organization (WHO) and Bloomberg Philanthropies Data for Health Initiative. Sampling Strategies for Representative National CRVS Verbal Autopsy Planning: A Guidance Document and Sample Size Calculator Tool. Geneva, Switzerland: WHO (2018). https://cdn.who.int/media/docs/default-source/classification/other-classifications/autopsy/crvs-va-national-sampling-strategies-for-representative-va-implementation.-v2.4.pdf?sfvrsn=2f67c8c6_3

World Health Organization (WHO) and the University of Queensland, School of Population Health. Improving the quality and use of birth, death and cause-of-death information: Guidance for a standards-based review of country practices. Geneva, Switzerland: WHO (2010). https://apps.who.int/iris/bitstream/handle/10665/44274/9789241547970_eng.pdf;sequence=1

Vital Strategies, United Nations Economic Commission for Africa, United Nations Economic and Social Commission for Asia and the Pacific, and Statistics Norway. Production of a Vital Statistics Report: Guide. New York, USA: Vital Strategies (2020). https://getinthepicture.org/sites/default/files/resources/Production%20of%20a%20Vital%20Statistics%20Report%20-%20Guide.pdf

APPENDIX A: Example Description of the System Context (Section 1) and data prerequisites (Section 2) for a VA and MCCD dataset from Lusaka, Zambia

PURPOSE: To demonstrate how the steps for aggregating VA and MCCD data can be applied in a context with both MCCD data for hospital deaths and VA data for community deaths, in this case "brought-in-dead" cases.

A. Understanding the context

In Lusaka District the civil registration system coverage is 100%; the whole district is covered by the system, and all deaths (both facility and non-facility) are expected to be captured by the system. Facility deaths are expected to be captured with an MCCD form completed for legal and statistical purposes. Non-facility deaths are expected to be captured by the police and through the burial office when a burial permit is sought. A small (select) proportion of the non-facility deaths go through medico-legal death investigation (MLDI) or post-mortem and cause of death completed on the MCCD form. With the recent implementation of VA for "BID" cases, cause of death is now available for statistical purposes for all non-facility deaths that don't have postmortem. By design, overlap in registration/capture of the facility and community systems is unlikely.

Completeness estimates for Lusaka District suggest that 95.3% of deaths were captured in the death registration system in 2018 (DNRPC, 2018). Approximately 30% of deaths took place in the community and 70% in health facilities. Both MCCD and VA (using WHO 2016 VA questionnaire) data are presently available for a period of three months, from May to July 2018.

B. Data prerequisites (minimum standards)

Based on the data quality criteria stated in Section 3, the Lusaka District data were considered of sufficient quality and were therefore considered suitable for further analysis and aggregation following the steps described in Section 3.

B.1. Size of datasets

A total of 3,829 deaths were available for analysis in the three-month period for which data were available. These deaths included 2,823 MCCDs and 1,006 VA interviews with the following age and sex distributions:

Distribution of Facility Deaths with MCCD in Lusaka District from May-July 2018

	No of Deaths		
Age Group	**Males**	**Females**	**Total**
0	253	198	451
1 _4	77	55	132
5_9	22	15	37
10_14	24	18	42
15-19	27	28	55
20-24	54	41	95
25-29	102	81	183
30-34	124	98	222
35-39	159	119	278
40-44	155	108	263
45-49	133	89	222
50-54	96	78	174
55-59	64	65	129

60-64	71	52	123
65-69	48	43	91
70-74	61	51	112
75-79	32	38	70
80+	50	94	144
Total	1,552	1,271	2,823

Distribution of VA Deaths in Lusaka District from May-July 2018

VA Deaths by Age Group		**VA Deaths by Sex**	
Age Group	**Frequency**	**Gender**	**Frequency**
Adult	793	**Female**	411
Child	167	**Male**	595
Neonate	46		
Total	1,006	Total	1,006

B.2. Consistency in data collection and COD assignment methods

All MCCD data were generated from facility deaths that were certified by doctors and coded following international ICD standards. All VA data were generated from VA interviews for community deaths as BIDs at hospital mortuaries in Lusaka; all VA data included in the analysis were collected using the WHO 2016 VA questionnaire and InterVA to assign cause of death.

B.3. Quality of data

B.3.1. Coverage, completeness, and representativeness

The WHO's ANACoD tool was used to assess the MCCD data while the VA data was assessed by the results from the InterVA (Open VA) outputs.

As noted above, death registration **coverage** in Lusaka District was **100%,** while **completeness** was **95.3%,** as calculated from estimates from the DNRPC, the Central Statistics Office (CSO), and the UN Population Division (UNPD) as follows:

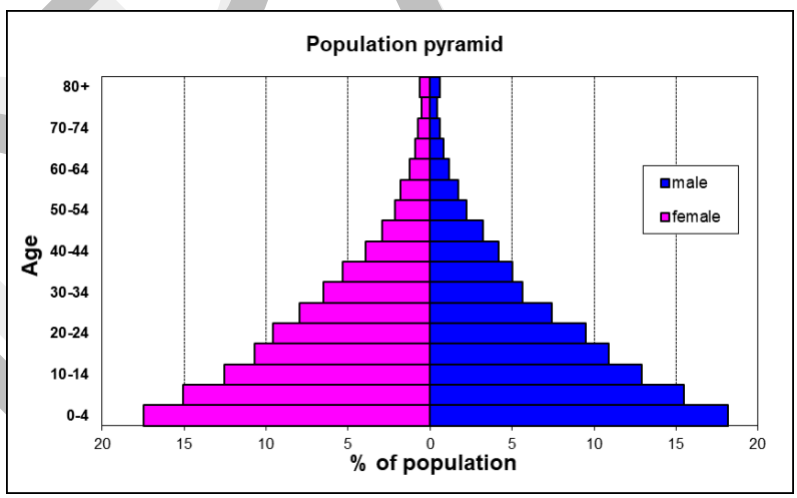
Registered deaths = 16,882 (from DNRPC)

Expected deaths = CDR X Total Population

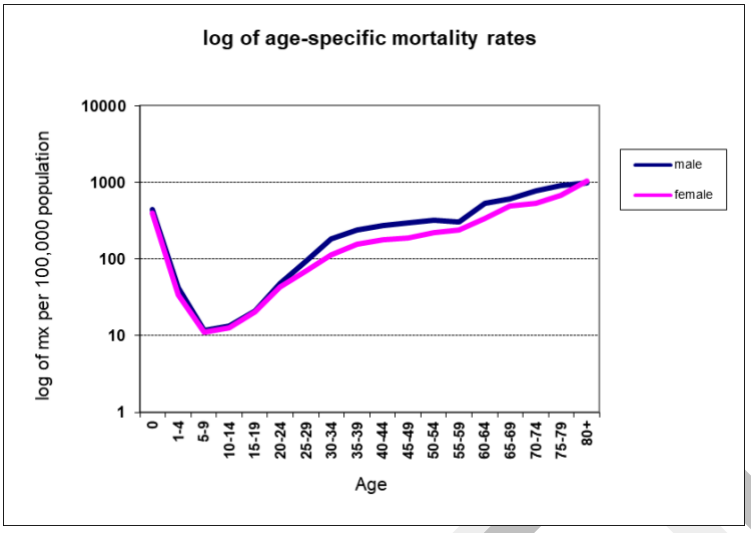
Expected deaths = 0.0073 (from CSO, UNPD) X 2,426,898 (from CSO) = 17,716

Completeness = (16,882/17,716) *100 = **95.3%**

A review of the population structure and death distributions patterns for Lusaka (from ANACoD) suggested that there was no major expected under-reporting or incompleteness in death registration.



As expected for low-income countries with a very young population and high mortality rates, the population structure for Lusaka shows a high proportion of those below 4 years, but low above 60 years.



Furthermore, the age (sex)-specific mortality pattern is also as expected, with high mortality in infancy, which reduces in the childhood ages, increases in the early adulthood ages and the log of the death rate increases linearly from age 35-44 on. The population CDR of 7.3 deaths per 1,000 population is consistent with other (low income) countries with similar population structures and death patterns (ANACoD).

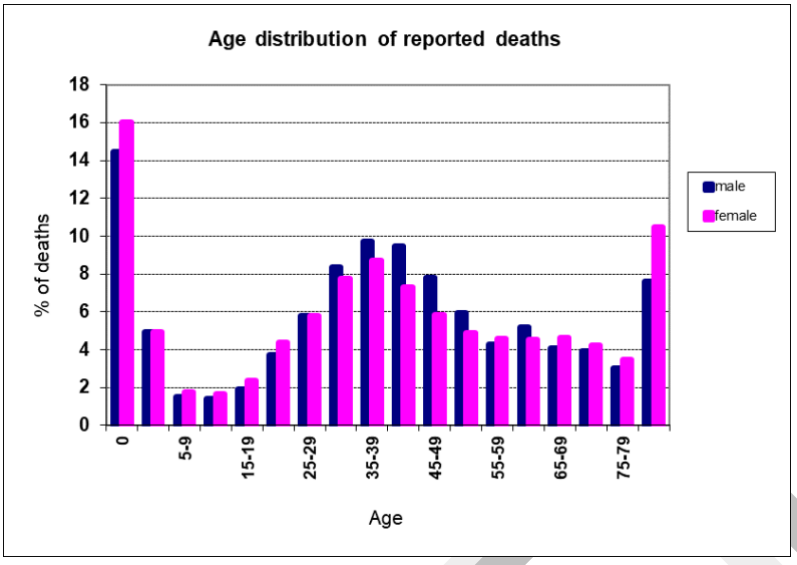
The completeness rate of over 95% for the death registration in Lusaka is sufficient to ensure that the MCCD and VA datasets targeted for aggregation are **representative** of all the deaths in Lusaka District.

B.3.2. Accuracy of information (measurement error)

Accuracy of MCCD data (per ANACoD)

Based on the findings described below, the MCCD data for Lusaka was of sufficient quality to be used for further aggregation with VA data.

- **Distribution of death by age and sex:** The age-sex mortality patterns shown in the graph below, together with the log of age-specific mortality rates shown above, are consistent with what is expected in a low-income country, especially with high HIV burden that affects mostly young adults in the reproductive age groups. These expected mortality patterns suggest minimal errors in the data.

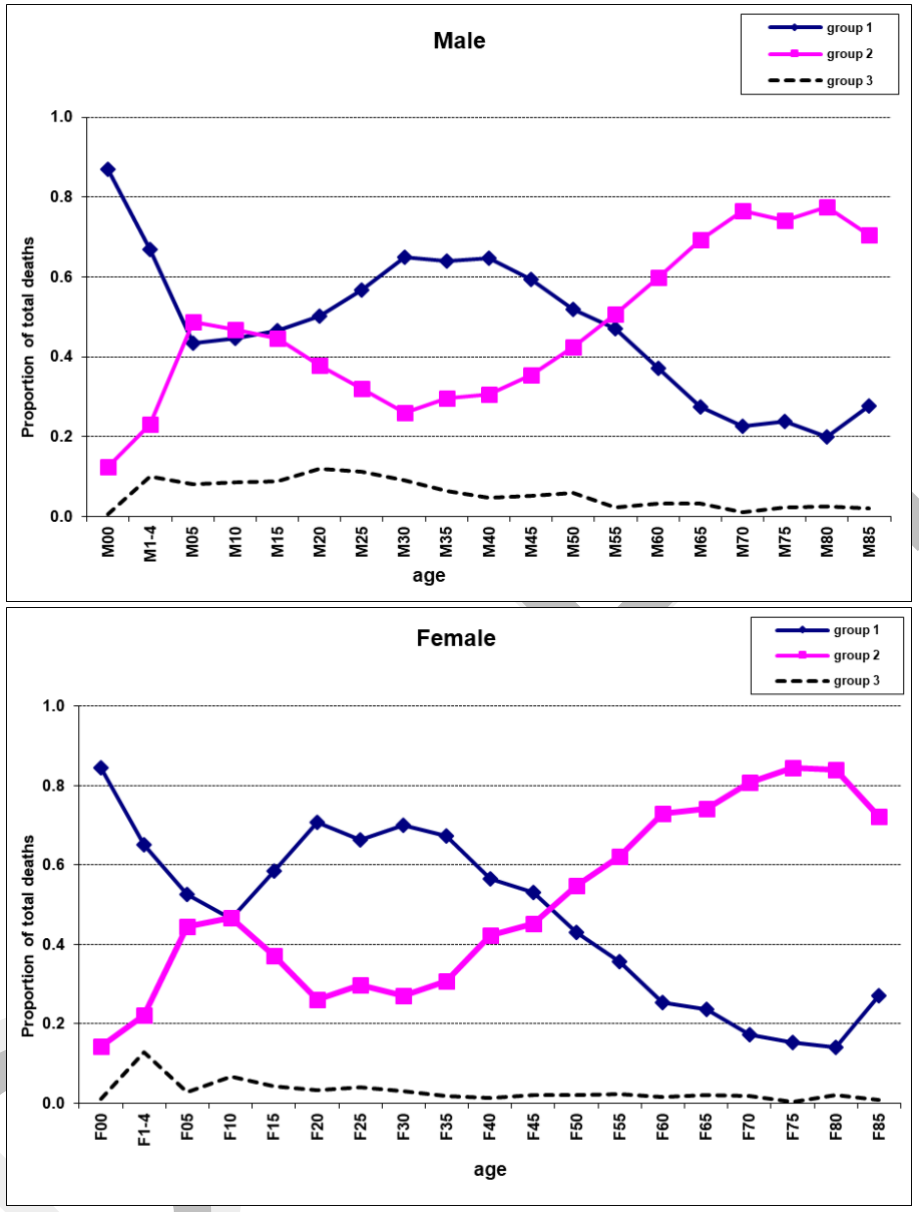


- **Deaths labelled with codes not valid for underlying cause of death (ANACoD Step 1):** In examining the ANACoD output of the Lusaka data, we found no invalid codes for underlying causes of death.

- **Implausible sex/cause combinations, implausible disease/age combinations, or deaths due to diseases unlikely to cause death (ANACoD Step 1):** We found no implausible sex/cause combinations, nor implausible disease/age combinations, nor deaths due to diseases unlikely to cause death in the data.

- **Distribution of deaths by cause; age patterns of broad groups of causes; leading causes; and ratio of non-communicable to communicable causes (ANACoD Steps 6-9):**
    - The following graphs of age and sex patterns of broad cause groups are consistent with expected disease patterns in low-income countries, where 55% of the deaths were due to communicable causes, 41% non-communicable causes, and 4% external causes.

Age and Sex Distribution of Cause of Deaths Broad Groups



- The ratio of non-communicable to communicable causes was 0.7 for Lusaka, which is consistent and comparable to 0.6 expected for low-income countries (ANACoD step 9).
- The proportion of the ill-defined causes (consisting of signs and symptoms) as underlying causes of deaths was **19.1%;** The majority of these ill-defined causes were in the very young age 1-4 years (at 30.0% and 30.8% for male and females, respectively) and the very old 85+ years (at 21.3% and 31.7% for males and females, respectively). While this proportion may be high, it is expected for a

country that is working on improving its quality of MCCD data through doctor MCCD training and improving its ICD coding practices.

B.3.3. Accuracy of the VA data

The indicators described below reflect an assessment of the accuracy of the VA data, which collectively suggests that the VA data for Lusaka was of sufficient quality to be used for further aggregation with the MCCD data.

- **Average length of interview:** The average length of the VA interviews using the WHO VA questionnaire in Lusaka was **45 minutes**, which falls within the acceptable range. Variations based on the age and sex of the decedent and reported symptoms were observed.
- **Median period between date of death and date of interview:** An analysis of this indicator found that **85% of all VAs were conducted on the same day of the death** of the decedent. While it is acknowledged that WHO recommends a culturally appropriate mourning period before VAs are conducted, in the routine VA implementation in Lusaka, the VA interviews were conducted immediately when the deceased, who died in the community, were brought to the mortuary, which is usually the same day of their death.
- **Consistency between narrative and symptoms reported:** This analysis is yet to be conducted.
- ***Item response patterns*:** This analysis is yet to be conducted.
- **Percentage of undetermined causes of death:** The InterVA outputs of the VA data found that overall only **5.5%** of the VA results were undetermined, which suggests reasonably good quality data. Below is the distribution of the undetermined VAs by age. However, some higher percentages for neonates and children may suggest a need to investigate further before attempting to merge the data. Consideration to only merge adult data may also be ideal.

VA Questionnaire	Percent undetermined
All ages	5.5%
Adults	1.5%
Children	21.0%
Neonates	17.4%

Based on the above overall listed quality criteria for both the MCCD and VA datasets, the Lusaka data are considered of sufficient quality for further analysis and aggregation following steps described in Section 3.

APPENDIX B-1: WHO VA cause list with corresponding International Classification of Diseases (ICD) and Global Health Estimate (GHE) codes for use with Inter-VA and InSilico-VA automated cause of death assignment methods

VA code	Verbal autopsy title	ICD-10 codes (from ICD)	ICD-11 codes (from ICD)	GHE code	Comments
VAs-01.01	**Sepsis**	A40-A41	1G40-1G41	0370	
VAs-01.02	**Acute respiratory infection, including pneumonia**	J00-J22 J851	CA00-CA07.1; CA40-CA43; CA45; CA4Z; 1E30-1E32	0390	Assumed all are low respiratory infections
VAs-01.03	**HIV/AIDS related death**	B20-B24	1C60-1C62	0100	
VAs-01.04	**Diarrheal diseases**	A00-A09	1A00-1A40.Z	0110	Assumed all are diarrhoeal diseases
VAs-01.05	**Malaria**	B50-B54	1F40-1F4Z	0210	
VAs-01.06	**Measles**	B05	1F03	0120	
VAs-01.07	**Meningitis and encephalitis**	A39 G00-G03 G04-G05	1B53-1B54; 1C1C; 1C80-1C8F; 1D00-1D02; 8B41	0170 0180	Distributed according to MCCD CSMFs for meningitis (A39; G00-G03) and encephalitis (G04-G05)

VA code	Verbal autopsy title	ICD-10 codes (from ICD)	ICD-11 codes (from ICD)	GHE code	Comments
VAs-01.08	**Tetanus**	A33-A35	1C13-1C14	0120	
VAs-01.09	**Pulmonary tuberculosis**	A15-A16	1B10	0030	
VAs-01.10	**Pertussis**	A37	1C12	0120	
VAs-01.11	**Hemorrhagic fever**	A92-A96 A98-A99	1D40-1D4Z; 1D6Z; 1D60-1D6Z	0370	Potential to add a category since the VA can provide more granularity that is policy relevant
VAs-01.12	**Dengue fever**	A97	1D20 - 1D2Z	0300	
VAs-01.13	**Coronavirus disease (COVID-19)**	U07.1; U07.2	RA01.0; RA01.1	0395	
VAs-01.99	**Unspecified infectious disease**	A17-A19 A20-A32 A36 A38 A42-A89 B00-B04 B06-B19 B25-B49 B55-B99	1A60-1A9Z; 1B11-1B51; 1B5Y-1B9Z; 1C10-1C11.Y; 1C16-1C1B; 1C1D-1C62; 1C8Y-1C8Z; 1D03-1D0Z; 1D80-1E1Z; 1E50-1E91.Z;	0370	

VA code	Verbal autopsy title	ICD-10 codes (from ICD)	ICD-11 codes (from ICD)	GHE code	Comments
			1F00-1F02; 1F04-1F2Z; 1F50-1G2Z; 1G60-1H0Z; AA00-AA0Z; AA3Y-AA3Z; DB90; EA00-EA6Y; EE12; EG61; FA90-FA91; FB30; GA00-GA02; GA05; GA07; GB02; GC08		
VAs-02.01	**Oral neoplasms**	C00-C06	2B60-2B66	0620	
VAs-02.02	**Digestive neoplasms**	C15 C16 C18-C21 C22 C25 C24	2B56.3; 2B70-2B72; 2B80-2B81; 2B90-2B9Y; 2C00-2C1Z	0630 0640 0650 0660 0670 0752	Distributed according to MCCD CSMFs for Oesophagus cancer (C15), Stomach cancer (C16), Colon and rectum cancers (C18-C21), Liver cancer (C22), Gallbladder and biliary tract cancer

VA code	Verbal autopsy title	ICD-10 codes (from ICD)	ICD-11 codes (from ICD)	GHE code	Comments
		C17, C23, C26		0780	(C24), Pancreas cancer (C25) and other malignant neoplasms.
VAs-02.03	**Respiratory neoplasms**	C33-C34 C32	2C20-2C2Z	0680 0753	Distributed according to MCCD CSMFs for Trachea, bronchus, lung cancers (C33-C34), Larynx cancer (C32) and other malignant neoplasm
		C30-C31 C35-C39		0780	
VAs-02.04	**Breast neoplasms**	C50	2C60-2C6Z	0700	
VAs-02.05	**Female reproductive neoplasms**	C53 C54-C55 C56 C51-C52 C57-C58	2C70-2C7Z	0710 0720 0730 0780	Distributed according to MCCD CSMFs for Cervix uteri cancer (C53), Corpus uteri cancer (C54-C55), Ovary cancer (C56) and other malignant neoplasms
VAs-02.06	**Male reproductive neoplasms**	C61 C62 C60, C63	2C80-2C8Z	0740 0742 0780	Distributed according to MCCD CSMFs for Prostate cancer (C61), Testicular cancer (C62) and other malignant neoplasms
VAs-02.99	**Other and unspecified neoplasms**	C07-C14 C40-C49 C64-D48 C91-C95	2A00-2A0Z; 2A20-2A90; 2B00-2B56.2; 2B56.Y-2B5Z; 2B67-2B6Y; 2C30-	0780	

VA code	Verbal autopsy title	ICD-10 codes (from ICD)	ICD-11 codes (from ICD)	GHE code	Comments
			2C5Z; 2C90-2E6Z; 2E80-2F9Z		
VAs-03.01	**Severe anaemia**	D50-D64	3A00-3A4.Z; 3A61-3A9Z	0780	
VAs-03.02	**Severe malnutrition**	E40-E46	5B50-5B54; 5B71-5B7Z	0550	
VAs-03.03	**Diabetes mellitus**	E10-E14	5A10-5A14	0800	Level 2 GHE: Diabetes Mellitus
VAs-04.01	**Acute cardiac disease**	I20-I25	BA01; BA40-BA6Z; BB00; BD11; MC82;	1130	
VAs-04.02	**Stroke**	I60-I69	8B00-8B23; 8B25-8B2Z	1140	
VAs-04.03	**Sickle cell with crisis**	D57	3A51	0812	
VAs-04.99	**Other and unspecified cardiac disease**	I00-I15 I26-I52 I70-I99	BA00; BA02-BA2Z; BA50-BA5Z; BA81-BA8Z; BB01-BC91; BC9Y-BC9Z;	1160	

VA code	Verbal autopsy title	ICD-10 codes (from ICD)	ICD-11 codes (from ICD)	GHE code	Comments
			BD10; BD12-BE2Z; 1B40-1B42		
VAs-05.01	**Chronic obstructive pulmonary disease (COPD)**	J40-J44	CA20-CA22	1180	
VAs-05.02	**Asthma**	J45-J46	CA23	1190	
VAs-06.01	**Acute abdomen**	R10	MD81	N/A	See comment below in VAs-99
VAs-06.02	**Liver cirrhosis**	K702 K703 K717 K74	DB93; DB94.2; DB94.3, DB95.5	1230	
VAs-07.01	**Renal failure**	N17-N19	GB60-GB6Z	1270	
VAs-08.01	**Epilepsy**	G40-G41	8A60-8A6Z	0970	
VAs-09.01	**Ectopic pregnancy**	O00	JA01	0420	Potential to add a category since the VA can provide more granularity that is policy relevant
VAs-09.02	**Abortion-related death**	O03-O08	JA00; JA05-JA0Z	0420	Potential to add a category since the VA can provide more granularity that is policy relevant
VAs-09.03	**Pregnancy-induced hypertension**	O10-O16	JA20-JA2Z	0420	Potential to add a category since the VA can provide more granularity that is policy relevant

VA code	Verbal autopsy title	ICD-10 codes (from ICD)	ICD-11 codes (from ICD)	GHE code	Comments
VAs-09.04	**Obstetric haemorrhage**	O46 O67 O72	JA40-JA4Z	0420 0420	Potential to add a category since the VA can provide more granularity that is policy relevant
VAs-09.05	**Obstructed labour**	O63-O66	JB03-JB06	0420	Potential to add a category since the VA can provide more granularity that is policy relevant . O63 ICD10 code might be included in this category
VAs-09.06	**Pregnancy-related sepsis**	O753 O85	JB0D.2; JB40	0420 0420	Potential to add a category since the VA can provide more granularity that is policy relevant . O75.3 ICD10 code might be included in this category
VAs-09.07	**Anaemia of pregnancy**	O990	JB64.0	0420	Potential to add a category since the VA can provide more granularity that is policy relevant
VAs-09.08	**Ruptured uterus**	O710-O711	JB0A.0; JB0A.1	0420	Potential to add a category since the VA can provide more granularity that is policy relevant
VAs-09.99	**Other and unspecified maternal cause**	O01-O02	JA02-JA04; JA60-JA6Z; JA80-JA8Z;	0420	Potential to add a category since the VA can provide more granularity that is policy relevant

VA code	Verbal autopsy title	ICD-10 codes (from ICD)	ICD-11 codes (from ICD)	GHE code	Comments
		O20-O45 O47-O62 O68-O70 O713-O719 O73-O84 O86-O99	JB00-JB02; JB07-JB09; JB0A.2-JB0D.1; JB0D.3-JB0Z; JB20-JB2Z; JB41-JB4Z; JB60-JB63; JB64.1-JB6Z		
VAs-10.01	**Prematurity or low birth weight**	P07	KA20-KA21	0500	
VAs-10.02	**Birth asphyxia**	P20-P22	KB20-KB23; KD30.0; KD30.1	0510	P22 is included in the GHE as Low birth weight
VAs-10.03	**Neonatal pneumonia**	P23 P24	KB24; KB26	0510 0520	Distributed according to MCCD CSMFs for Congenital pneumonia (P23) and Neonatal aspiration syndromes (P24)
VAs-10.04	**Neonatal sepsis**	P36	KA60	0520	
VAs-10.05	**Neonatal tetanus**	A33	1C15	0120	
VAs-10.06	**Congenital malformation**	Q00-Q99	9A00; EC10-EC7Y; GB81-GB82;	1460	

VA code	Verbal autopsy title	ICD-10 codes (from ICD)	ICD-11 codes (from ICD)	GHE code	Comments
			GB8Z; LA00-LD9Z		
VAs-10.99	**Other and unspecified perinatal cause of death**	P00-P05 P08-P15 P26-P35 P37-P94 P96 R95	EH10-EH3Y; KA00-KA0Z; KA22-KA4Z; KA61-KA8Z; KB00-KB0Z; KB25; KB27-KB8Z; KC00-KC9Z; KD10-KD1Z; KD30.2-KD5Z; MH11	0530	
VAs-11.01	**Fresh stillbirth**	P95	KD3B.1	0530	
VAs-11.02	**Macerated stillbirth**	P95	KD3B.0	0530	
VAs-12.01	**Road traffic accident**	V011 V021 V031 V041 V051 V061 V092	PA00-PA5Z	1530	

VA code	Verbal autopsy title	ICD-10 codes (from ICD)	ICD-11 codes (from ICD)	GHE code	Comments
		V093			
		V114-V119			
		V124-V129			
		V134-V139			
		V144-V149			
		V154-V159			
		V164-V169			
		V174-V179			
		V184-V189			
		V194-V199			
		V204-V209			
		V214-V219			
		V224-V229			
		V234-V239			
		V244-V249			
		V254-V259			
		V264-V269			
		V274-V279			
		V284-V289			
		V294-V299			
		V305-V309			
		V315-V319			

VA code	Verbal autopsy title	ICD-10 codes (from ICD)	ICD-11 codes (from ICD)	GHE code	Comments
		V325-V329			
		V335-V339			
		V345-V349			
		V355-V359			
		V365-V369			
		V375-V379			
		V385-V389			
		V394-V399			
		V405-V409			
		V415-V419			
		V425-V429			
		V435-V439			
		V445-V449			
		V455-V459			
		V465-V469			
		V475-V479			
		V485-V489			
		V494-V499			
		V505-V509			
		V515-V519			
		V525-V529			
		V535-V359			

VA code	Verbal autopsy title	ICD-10 codes (from ICD)	ICD-11 codes (from ICD)	GHE code	Comments
		V545-V549			
		V555-V559			
		V565-V569			
		V575-V579			
		V585-V589			
		V594-V599			
		V605-V609			
		V615-V619			
		V625-V629			
		V635-V639			
		V645-V649			
		V655-V659			
		V665-V669			
		V675-V679			
		V685-V689			
		V694-V699			
		V705-V709			
		V715-V719			
		V725-V729			
		V735-V739			
		V745-V749			
		V755-V759			

VA code	Verbal autopsy title	ICD-10 codes (from ICD)	ICD-11 codes (from ICD)	GHE code	Comments
		V765-V769 V775-V779 V785-V789 V794-V799 V800-V809 V811-V819 V821-V829 V830-V833 V840-V843 V850-V853 V860-V863 V870-V879 V892 V893 Y850			
VAs-12.02	**Other transport accident**	V90-V99 Y859			
VAs-12.03	**Accidental fall**	W00-W19	PA60-PA6Z	1550	
VAs-12.04	**Accidental drowning and submersion**	W65-W74	PA90-PA9Z	1570	

VA code	Verbal autopsy title	ICD-10 codes (from ICD)	ICD-11 codes (from ICD)	GHE code	Comments
VAs-12.05	**Accidental exposure to smoke, fire and flames**	X00-X19	PB10-PB15; PB1Y-PB1Z; PB55	1560	
VAs-12.06	**Contact with venomous animals and plants**	X20-X29	PA78; PA79	1590	
VAs-12.07	**Accidental poisoning and exposure to noxious substance**	X40-X49	PB20-PB36	1540	Moved to Poisoning
VAs-12.08	**Intentional self-harm**	X60-X84 Y870	PB80-PD3Z	1610	
VAs-12.09	**Assault**	X85-Y09 Y871	PD50-PF2Z; PJ20-PJ2Z	1620	
VAs-12.10	**Exposure to force of nature**	X30-X39	PJ00-PJ0Z	1580	Moved to natural disaster
VAs-12.99	**Other and unspecified external cause of death**	S00-T99 W20-W64 W75-W99 X10-X19 X50-X59 Y10-Y84 Y86 Y872 Y88-Y89	EL51-EL54; NA00-NF2Z; PA70-PA77; PA7Y-PA8Z; PB00-PB0Z; PB16; PB50-PB54; PB56-PB6Z; PF40-PH8Z; PJ20-PJ2Z; PJ40-PL2Z	1590	

VA code	Verbal autopsy title	ICD-10 codes (from ICD)	ICD-11 codes (from ICD)	GHE code	Comments
VAs-98	**Other and unspecified non-communicable disease**	D65-D89 E00-E07 E15-E35 E50-E90 F00-F99 G06-G09 G10-G37 G50-G99 H00-H95 J30-J39 J47-J99 K00-K31 K35-K38 K40-K69 K77-K93 L00-L99 M00-M99 N00-N16 N20-N99		N/A	Potential to add a category since the VA can provide more granularity that is policy relevant
VAs-99	**Unknown and ill-defined cause of death**	R00-R09 R11-R94	MA00-MB42;	N/A	

VA code	Verbal autopsy title	ICD-10 codes (from ICD)	ICD-11 codes (from ICD)	GHE code	Comments
		R96-R99	MB44-MB4D; MB60-MB9Y; MC21; MC80-MD80; MD82-ME81; ME83; ME86-ME92; ME9Y-MF39; MF3Y-MF53; MF55; MF57-MF7Z; MF90-MH10; MH12-MH2Y		

APPENDIX B-2: VA cause list with corresponding International Classification of Diseases (ICD) codes for use with Tariff / SmartVA automated cause of death assignment method

Cause of death list for SmartVA with corresponding ICD-10 codes (adult)

Text for Smart VA cause (ADULT)	ICD-10 code (to ICD)	ICD-10 codes (from ICD-10)
Diarrhoea/dysentery	A09	A00–A09
Tuberculosis	A16	A15–A19
AIDS	B24	B20–B24
Malaria	B54	B50–B54
Other infectious diseases	B99	A10–A14, A20–B19, B25–B49, B55–B99
Esophageal cancer	C15	C15
Stomach cancer	C16	C16
Colorectal cancer	C18	C18–C21
Lung cancer	C34	C34
Breast cancer	C50	C50
Cervical cancer	C53	C53
Prostate cancer	C61	C61
Leukemia/lymphoma	C96	C81–C85; C91–C96
Other cancers	C76	C00–C14, C17, C22–C33, C35–C49, C51–C52, C54–C60, C62–C80, C86–C90, C97–D48
Diabetes	E14	E10–E14
Other cardiovascular diseases	I99	I00–I19 I26–I59, I70–I99
Ischemic heart diseases	I24	I20–I25
Stroke	I64	I60–I69
Pneumonia	J22	J10–J22, J85
Chronic respiratory diseases	J44	J40–J46
Cirrhosis	K74	K70–K76
Chronic kidney disease	N19	N17–N19
Maternal	O95	O00–O99
Undetermined	R99	R00–R99
Road traffic	V89	V01–V89

Falls	W19	W00–W19
Drowning	W74	W65–W74
Fires	X09	X00–X19
Bite of venomous animal	X27	X20–X29
Poisonings (accidental)	X49	X40–X49
Suicide (intentional self-harm)	X84	X60–X84
Homicide (assault)	Y09	X85–Y09
Other injuries	X58	S00–T98, V90–V99, W20–W64, W75–W99, X30–X39, X50–X59, Y10–Y98
Other non-communicable diseases	UU1*	All other ICD-10 codes NCDs#

Notes: Column 1 lists the Smart VA cause text; column 2 lists the ICD-10 codes that would be used if the condition labelled by column 1 were coded to ICD-10; column 3 lists the ICD-10 categories that need to be grouped to match the content of the relevant VA entity.
This code is specific to SmartVA.
* This other non-communicable diseases group covers all non-communicable conditions/diseases that could not be assigned to a specific non-communicable disease.

Cause of death list for SmartVA with corresponding ICD-10 codes (child)

Text for SmartVA cause (CHILD)	ICD-10 code (to ICD)	ICD-10 Codes (from ICD-10)
Diarrhoea/dysentery	A09	A00–A09
Sepsis	A41	A40–A41
Haemorrhagic fever	A99	A92–A99
Measles	B05	B05
AIDS	B24	B20–B24
Malaria	B54	B50–54
Other infectious diseases	B99	A10–A39, A42–A91, B00– B04, B06–B49, B55–B99
Cancers	C76	C00–D48
Meningitis	G03	G00–G03, A39,A87
Encephalitis	G04	G04, A83–A86
Cardiovascular diseases	I99	I00–I99
Pneumonia	J22	J10–J22, J85
Digestive diseases	K92	K00–K93
Undetermined	R99	R00–R99
Road traffic	V89	V01–V89
Falls	W19	W00–W19

Drowning	W74	W65–W74
Fires	X09	X00–X19
Bite of venomous animal	X27	X20–X29
Poisonings	X49	X40–X49
Homicide	X09	X85–Y09
Other defined causes of child deaths	UU2*	All other ICD-10 codes[#]

Cause of death list for SmartVA with corresponding ICD-10 codes (neonate)

Text for SmartVA (NEONATE)	ICD-10 code (to ICD)	ICD-10 code (to ICD)
Preterm delivery	P07	P05–P07
Birth asphyxia	P21	P20–P22
Pneumonia	P23	P23–P25,J10–J22
Meningitis/sepsis	P36	P36, G00–G04, A39, A87
Stillbirth	P95	P95
Congenital malformation	Q89	Q00–Q99
Undetermined	R99	All other ICD-10 codes

APPENDIX B-3: Global Health Estimate (GHE) cause list codes and levels

GHE code	Description	Level
0000	All causes	Level 0
0010	I. Communicable, maternal, perinatal and nutritional conditions	Level 1
0020	A. Infectious and parasitic diseases	Level 2
0030	1. Tuberculosis	Level 3
0040	2. STDs excluding HIV	Level 3
0050	a. Syphilis	Level 4
0060	b. Chlamydia	Level 4
0070	c. Gonorrhoea	Level 4
0080	d. Other STDs	Level 4
0090	3. HIV/AIDS	Level 3
0100	4. Diarrhoeal diseases	Level 3
0110	5. Childhood-cluster diseases	Level 3
0120	a. Pertussis	Level 4
0130	b. Poliomyelitis	Level 4
0140	c. Diphtheria	Level 4
0150	d. Measles	Level 4
0160	e. Tetanus	Level 4
0170	6. Meningitis	Level 3
0180	7. Hepatitis B	Level 3
0190	Hepatitis C	Level 3
0200	8. Malaria	Level 3
0210	9. Tropical-cluster diseases	Level 3
0220	a. Trypanosomiasis	Level 4
0230	b. Chagas disease	Level 4
0240	c. Schistosomiasis	Level 4
0250	d. Leishmaniasis	Level 4
0260	e. lymphatic filariasis	Level 4
0270	f. Onchocerciasis	Level 4
0280	10. Leprosy	Level 3
0290	11. Dengue	Level 3
0300	12. Japanese encephalitis	Level 3
0310	13. Trachoma	Level 3
0320	14. Intestinal nematode infections	Level 3
0330	a. Ascariasis	Level 4
0340	b. Trichuriasis	Level 4
0350	c. Hookworm disease	Level 4
0360	Other intestinal infections	Level 3

0370	Other infectious diseases	Level 3
0380	B. Respiratory infections	Level 2
0390	1. Lower respiratory infections	Level 3
0400	2. Upper respiratory infections	Level 3
0410	3. Otitis media	Level 3
0420	C. Maternal conditions	Level 2
0430	1. Maternal haemorrhage	Level 3
0440	2. Maternal sepsis	Level 3
0450	3. Hypertensive disorders	Level 3
0460	4. Obstructed labour	Level 3
0470	5. Abortion	Level 3
0480	Other maternal conditions	Level 3
0490	D. Perinatal conditions	Level 2
0500	1. Low birth weight	Level 3
0510	2. Birth asphyxia and birth trauma	Level 3
0520	Other perinatal conditions	Level 3
0530	E. Nutritional deficiencies	Level 2
0540	1. Protein-energy malnutrition	Level 3
0550	2. Iodine deficiency	Level 3
0560	3. Vitamin A deficiency	Level 3
0570	4. Iron-deficiency anaemia	Level 3
0580	Other nutritional disorders	Level 3
0590	II. Noncommunicable diseases	Level 1
0600	A. Malignant neoplasms	Level 2
0610	1. Mouth and oropharynx cancers	Level 3
0611	a. Lip and oral cavity	Level 4
0612	b. Nasopharynx	Level 4
0613	c. Other pharynx	Level 4
0620	2. Oesophagus cancer	Level 3
0630	3. Stomach cancer	Level 3
0640	4. Colon and rectum cancers	Level 3
0650	5. Liver cancer	Level 3
0660	6. Pancreas cancer	Level 3
0670	7. Trachea, bronchus, lung cancers	Level 3
0680	8. Melanoma and other skin cancers	Level 3
0681	a. Malignant skin melanoma	Level 4
0682	b. Non-melanoma skin cancer	Level 4
0690	9. Breast cancer	Level 3
0700	10. Cervix uteri cancer	Level 3
0710	11. Corpus uteri cancer	Level 3
0720	12. Ovary cancer	Level 3
0730	13. Prostate cancer	Level 3

0731	14. Testicular cancer	Level 3
0732	15. Kidney and ureter cancer	Level 3
0740	16. Bladder cancer	Level 3
0741	17. Brain and nervous system cancers	Level 3
0742	18. Gallbladder and biliary tract cancer	Level 3
0743	19. Larynx cancer	Level 3
0744	20. Thyroid cancer	Level 3
0745	21. Mesothelioma	Level 3
0750	22. Lymphomas, multiple myeloma	Level 3
0751	a. Hodgkin lymphoma	Level 4
0752	b. Non-Hodgkin lymphoma	Level 4
0753	c. Multiple myeloma	Level 4
0760	23. Leukaemia	Level 3
0770	24. Other malignant neoplasms	Level 3
0780	B. Other neoplasms	Level 2
0790	C. Diabetes mellitus	Level 2
0800	D. Endocrine disorders	Level 2
0810	E. Neuropsychiatric conditions	Level 2
0820	1. Unipolar depressive disorders	Level 3
0830	2. Bipolar disorder	Level 3
0840	3. Schizophrenia	Level 3
0850	4. Epilepsy	Level 3
0860	5. Alcohol use disorders	Level 3
0870	6. Alzheimer and other dementias	Level 3
0880	7. Parkinson disease	Level 3
0890	8. Multiple sclerosis	Level 3
0900	9. Drug use disorders	Level 3
0910	10. Post-traumatic stress disorder	Level 3
0920	11. Obsessive-compulsive disorder	Level 3
0930	12. Panic disorder	Level 3
0940	13. Insomnia (primary)	Level 3
0950	14. Migraine	Level 3
0960	15. Mental Retardation	Level 3
0970	Other neuropsychiatric disorders	Level 3
0980	F. Sense organ diseases	Level 2
0990	1. Glaucoma	Level 3
1000	2. Cataracts	Level 3
1010	3. Vision disorders, age-related	Level 3
1020	4. Hearing loss, adult onset	Level 3
1030	Other sense organ disorders	Level 3
1040	G. Cardiovascular diseases	Level 2
1050	1. Rheumatic heart disease	Level 3
1060	2. Hypertensive heart disease	Level 3

1070	3. Ischaemic heart disease	Level 3
1080	4. Cerebrovascular disease	Level 3
1090	5. Inflammatory heart diseases	Level 3
1100	Other cardiovascular diseases	Level 3
1110	H. Respiratory diseases	Level 2
1120	1. Chronic obstructive pulmonary disease	Level 3
1130	2. Asthma	Level 3
1140	Other respiratory diseases	Level 3
1150	I. Digestive diseases	Level 2
1160	1. Peptic ulcer disease	Level 3
1170	2. Cirrhosis of the liver	Level 3
1180	3. Appendicitis	Level 3
1181	4. Gastritis and duodenitis	Level 3
1182	5. Paralytic ileus and intestinal obstruction	Level 3
1183	6. Inflammatory bowel disease	Level 3
1184	7. Gallbladder and biliary diseases	Level 3
1185	8. Pancreatitis	Level 3
1190	9. Other digestive diseases	Level 3
1200	J. Genitourinary diseases	Level 2
1210	1. Nephritis and nephrosis	Level 3
1220	2. Benign prostatic hypertrophy	Level 3
1230	Other genitourinary system diseases	Level 3
1240	K. Skin diseases	Level 2
1250	L. Musculoskeletal diseases	Level 2
1260	1. Rheumatoid arthritis	Level 3
1270	2. Osteoarthritis	Level 3
1280	3. Gout	Level 3
1290	4. Back pain	Level 3
1300	Other musculoskeletal disorders	Level 3
1310	M. Congenital anomalies	Level 2
1320	1. Abdominal wall defect	Level 3
1330	2. Anencephaly	Level 3
1340	3. Anorectal atresia	Level 3
1350	4. Cleft lip	Level 3
1360	5. Cleft palate	Level 3
1370	6. Oesophageal atresia	Level 3
1380	7. Renal agenesis	Level 3
1390	8. Down syndrome	Level 3
1400	9. Congenital heart anomalies	Level 3
1410	10. Spina bifida	Level 3
1420	Other Congenital anomalies	Level 3
1430	N. Oral conditions	Level 2
1440	1. Dental caries	Level 3

1450	2. Periodontal disease	Level 3
1460	3. Edentulism	Level 3
1470	Other oral diseases	Level 3
1475	O. Sudden infant death syndrome	Level 2
1480	III. Injuries	Level 1
1490	A. Unintentional injuries	Level 2
1500	1. Road traffic accidents	Level 3
1510	2. Poisonings	Level 3
1520	3. Falls	Level 3
1530	4. Fires	Level 3
1540	5. Drownings	Level 3
1541	6. Exposure to mechanical forces	Level 3
1542	7. Natural disasters	Level 3
1550	8. Other unintentional injuries	Level 3
1560	B. Intentional injuries	Level 2
1570	1. Self-inflicted injuries	Level 3
1580	2. Violence	Level 3
1590	3. War	Level 4
1600	Other intentional injuries	Level 3
1610	Ill-defined diseases	Level 1
1620	Ill-defined injuries/accidents	Level 2

APPENDIX C: Annotated SAS code and select output for preparing datasets and aggregating data

SAS Code to Map VA InterVA outputs to corresponding (3-digit) ICD-10 Codes and export the results by sex to MS Excel

```
libname VA_ICD 'C:\Users\VA_MCCD_Lusaka';
data VA_ICD.VA2019;
 set VA_ICD.InterVA19;
 if sex = 'male' then sex1 = 1; if sex = 'female' then sex1 = 2;

If interva5 = 'Abortion-related death' then ICD = 'O06'; If interva5= 'Accid drowning and submersion' then  ICD = 'W74';
If interva5= 'Accid expos to smoke fire & flame' then ICD = 'X09'; If interva5= 'Accid fall' then ICD = 'W19';
If interva5= 'Accid poisoning & noxious subs' then ICD = 'X49'; If interva5 = 'Acute abdomen' then ICD = 'R10';
If interva5= 'Acute cardiac disease' then ICD = 'I24'; If interva5= 'Acute resp infect incl pneumonia' then ICD = 'J22';
If interva5= 'Anaemia of pregnancy' then ICD = 'O99'; If interva5= 'Assault' then ICD = 'Y09'; If interva5= 'Asthma' then ICD = 'J45';
If interva5= 'Birth asphyxia' then ICD = 'P21'; If interva5= 'Breast neoplasms' then ICD = 'C50';
If interva5= 'Chronic obstructive pulmonary dis' then ICD = 'J44';
If interva5= 'Congenital malformation' then ICD = 'Q89'; If interva5= 'Contact with venomous plant/animal' then ICD = 'X29'; If interva5= 'Dengue fever' then ICD = 'A90';
If interva5= 'Diabetes mellitus' then ICD = 'E14'; If interva5= 'Diarrhoeal diseases' then ICD = 'A09'; If interva5= 'Digestive neoplasms' then ICD = 'C26';
If interva5= 'Ectopic pregnancy' then ICD ='O00'; If interva5= 'Epilepsy' then ICD = 'G40'; If interva5= 'Exposure to force of nature' then ICD = 'X39';
If interva5= 'Fresh stillbirth' then ICD = 'P95'; If interva5= 'Female reproductive neoplasms' then ICD = 'C57';
If interva5= 'HIV/AIDS related death' then ICD = 'B24'; If interva5=  'Haemorrhagic fever' then ICD = 'A99';
If interva5= 'Indeterminate' then ICD = 'R99'; If interva5=  'Intentional self-harm' then ICD = 'X84'; If interva5= 'Liver cirrhosis' then ICD = 'K74';
If interva5= 'Macerated stillbirth' then ICD = 'P95'; If interva5= 'Malaria' then ICD = 'B54'; If interva5= 'Male reproductive neoplasms' then ICD = 'C63';
```

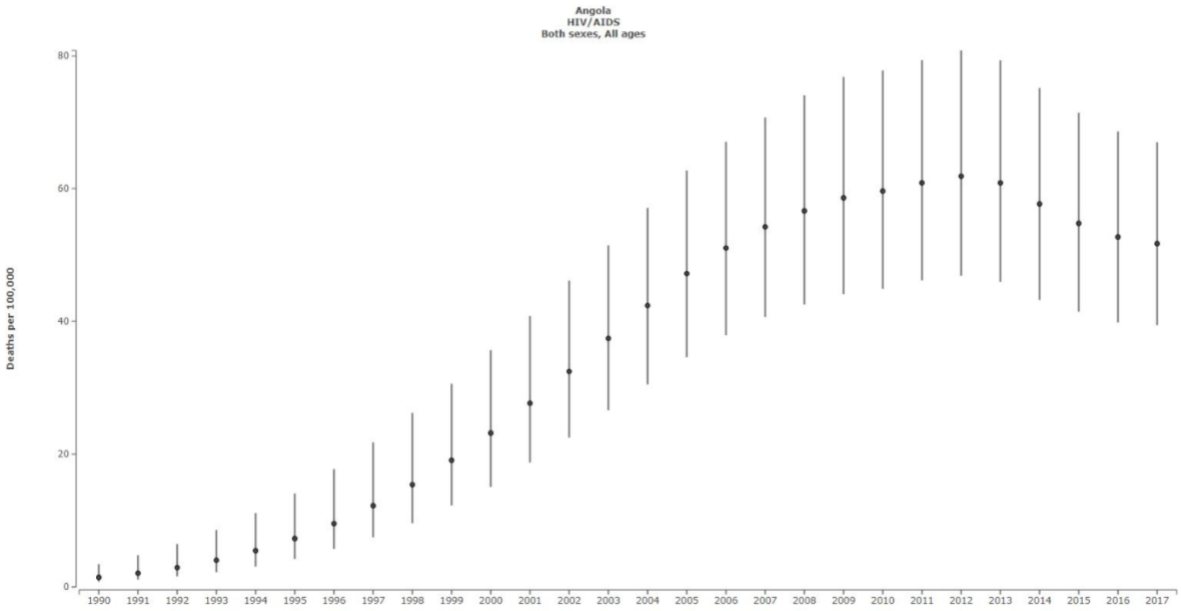
```
If interva5= 'Measles' then ICD = 'B05'; If interva5= 'Meningitis and encephalitis' then ICD =
'G03';
If interva5= 'Neonatal pneumonia' then ICD = 'P23'; If interva5= 'Neonatal sepsis' then ICD =
'P36'; If interva5 = 'Neonatal tetanus' then ICD = 'A33';
If interva5= 'Obstetric haemorrhage' then ICD= 'O72'; If interva5= 'Obstetric labour' then ICD
= 'O66'; If interva5= 'Oral neoplasms' then ICD = 'C06';
If interva5= 'Other and unspecified NCD' then ICD = 'R99'; If interva5= 'Other and unspecified
cardiac dis' then ICD = 'I99';
If interva5= 'Other and unspecified external CoD' then ICD = 'X59'; If interva5= 'Other and
unspecified infect dis' then ICD = 'B99';
If interva5= 'Other and unspecified maternal CoD' then ICD = 'O05'; If interva5= 'Other and
unspecified neoplasms' then ICD = 'C80';
If interva5= 'Other and unspecified perinatal CoD' then ICD = 'P96'; If interva5= 'Other
transport accident' then ICD = 'V99';
If interva5= 'Pertussis' then ICD = 'A37'; If interva5= 'Pregnancy-induced hypertension' then
ICD = 'O13'; If interva5= 'Pregnancy-related sepsis' then ICD = 'O75';
If interva5= 'Prematurity' then ICD = 'P07'; If interva5= 'Pulmonary tuberculosis' then ICD =
'A16'; If interva5= 'Pertussis' then ICD = 'A37';
If interva5= 'Renal failure' then ICD = 'N19'; If interva5= 'Reproductive neoplasms MF' then
ICD = 'C63'; If interva5= 'Respiratory neoplasms' then ICD = 'C39';
If interva5= 'Road traffic accident' then ICD = 'V89'; If interva5= 'Ruptured uterus' then ICD =
'O71'; If interva5= 'Sepsis (non-obstetric)' then ICD = 'A41';
If interva5= 'Severe anaemia' then ICD = 'D64'; If interva5= 'Severe malnutrition' then ICD =
'E46'; If interva5= 'Sickle cell with crisis' then ICD = 'D57';
If interva5= 'Stroke' then ICD = 'I64'; If interva5= 'Tetanus' then ICD = 'A35'; If interva5=
'Unspecified infectious disease' then ICD = 'B99';
 /*Keep interva5 ICD id10019 sex ageinyears age;*/
run;
ODS TAGSETS.EXCELXP
file='C:\Users\VA_MCCD_Lusaka\VA_ICD10.xls'
STYLE=minimal
OPTIONS ( Orientation = 'landscape'
FitToPage = 'yes'
Pages_FitWidth = '1'
Pages_FitHeight = '100' );
proc print data=VA_ICD.VA2019;
Run;
```

```
ods tagsets.excelxp close;
```

APPENDIX D: Trend comparisons with mortality data

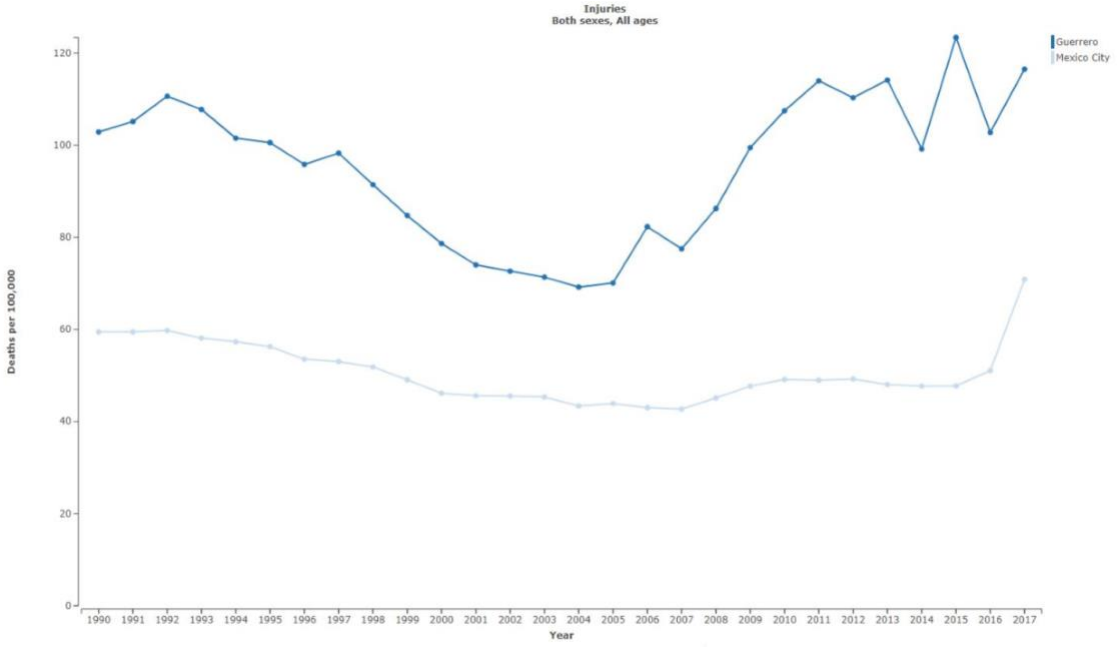
Scatter plots show the distribution of point estimates (i.e. deaths) by another variable, in this case, time. Figure A is a scatter plot of deaths per 100,000 due to HIV/AIDS in Angola from 1992 through 2017. Note that the graph includes error bars, accounting for uncertainty around the point estimate.

Figure A: Deaths per 100,000 due to HIV/AIDS in Angola between 1990 and 2017 (GBD, 2017)



A line graph is an alternative visualization method that can highlight changes in mortality over time more intentionally. Additionally, these graphs can be used to compare change over time by a second variable, including socio-demographic variables or geography. Figure B shows the trends in deaths due to injuries for all ages and both sexes, comparing the state of Guerrero to Mexico City in Mexico.

Figure B: Deaths per 100,000 in two states in Mexico between 1990 and 2017 (GBD, 2017)



These graphs can help identify whether progress is occurring according to government goals and whether more resources or attention is needed in a particular geographic association or socio-demographic group.

Conduct a trend analysis

If there is an interest in testing for significance of a trend, associations with other factors, or forecasting future values, then a trend analysis can be conducted using either record-level data or aggregated data.

When conducting a trend analysis, consider the following:

1. Identify the rationale for the time period chosen in the time trend analysis.
2. If conducting time trend analyses of vital records over time, assume there is minimal or no correlation between each time point (e.g. month-to-month, year-to-year).[26]
3. Use the same VA algorithm consistently for proper comparison of VA data and consider changes in ICD coding for MCCD data.

[26] Ingram DD, Malec DJ, Makuc DM, Kruszon-Moran D, Gindi RM, Albert M, et al. National Center for Health Statistics Guidelines for Analysis of Trends. National Center for Health Statistics. Vital Health Stat 2(179). 2018.

4. Select the appropriate type of analysis. For example, a time series analysis uses regression methods and modeling to identify trends in the data.