SmartVA Analyze Outputs Interpretation Sheet

SmartVA-Analyze uses an algorithm called 'Tariff 2.0' to assign the cause of death based on the details of the verbal autopsy (VA) interview. The output from SmartVA-Analyze is organised in four folders and what follows is a brief description of the files within these folders, their purpose and how they may be used/interpreted.

VA results should be analysed on a population basis. It should be noted that a large number of VAs (>1000) are necessary to provide a good picture of the population cause distribution. In addition, VA may be collected from specific locations and have an age distribution that may not match that of the general population. Therefore, analysis of VA results should be done in consideration of these factors.

1 – individual-cause-of-death

This folder contains information on individual verbal autopsy results. There are four files, one for each of the three modules (**adult-predictions.csv, child-predictions.csv, neonate-predictions.csv**) and one combining the all-age results (**individual-cause-of-death.csv**)

The information in this folder contains data that could be useful for linking/integrating with other health information or civil registration systems. It includes individual records of VA-ID, name of deceased, geographical location (up to 4 levels), cause name, SmartVA code, ICD-10 code[1], age and sex (1 for male and 2 for female) as well as other information for identification purposes such as date of birth, date of death and interview date.

These files can be used for detailed analysis of VA results, e.g. by different age-groups, location etc. However, caution should be used when analysing the all-age results (individual-cause-of-death.csv) file since child and neonatal deaths are often under-represented in verbal autopsy data and analysing by all-age may not reflect the true scale of the problem in these age-groups.

2 – CSMF

This file contains the cause-specific mortality fraction (CSMF) or the aggregated population level distribution of cause of death from the VA data that has been analysed. There are five files within this folder. Three files correspond to the CSMF for the age-group defined by verbal autopsy for adults (12 years and above), child (29 days – 11 years) and neonate (0-28 days). One file (**csmf.csv**) provides an aggregated CSMF file for all the VA data. As above, caution should be used with interpreting these results which may under-estimate the importance of neonatal and child cause of death. Cause of death, SmartVA code and ICD-10 code with results are provided for male and female separately and combined.

Unlike in the individual cause files, for CSMF there are no 'Undetermined' causes of death. This is because SmartVA-Analyze redistributes the VAs with undetermined cause of death using a combination of two methods[2]. Firstly, a VA with Undetermined COD is fractionally distributed amongst all VA causes, with weights proportional to the likelihood that the particular cause was assigned to undetermined in the gold standard database. The gold standard database is the dataset on which the

[1] For SmartVA, single three-digit ICD-10 codes represent a range of codes. See table at the end of this document which provides an explanation of the cause category for adults, child and neonate.

[2] Serina P, et al. Improving performance of the Tariff Method for assigning causes of death to verbal autopsies. *BMC Medicine*. 2015; 13:291

tariff analysis is based and includes VAs done on 12,542 deaths, for which the true cause of death was known[3]. Certain deaths (such as pneumonia) are more likely to return an Undetermined COD because this cause is inherently more difficult to diagnose using VA methods than a cause like Road traffic accident. The redistribution addresses this by applying a higher weighting to such deaths. Secondly, this fractional redistribution weight is averaged with a proportional redistribution weight selected according to the Global Burden of Disease age and sex cause of death distribution for the country[4].

To produce a CSMF file that includes Undetermined category, you should process your data through SmartVA Analyze *without* specifying the country.

The file **gbd-level1-csmf.csv** provides the breakdown of CSMF by the global burden of disease broad categories (Communicable, maternal, neonatal and nutritional disease; Non-communicable diseases and; Injuries). This broad breakdown can be a useful first check of your VA data to assess whether it follows expectations in terms of these three categories.

3 – graphs-and-tables

Whilst graphs and tables can be developed manually using the files contained in the individual-cause-of-death and CSMF folder, SmartVA Analyze produces some ready-made graphs.

For adult deaths, the CSMF for all adults (**adult-csmf-figure.png**) and for female (**female-csmf-figure.png**) and male (**male-csmf-figure.png**) separately are available. For children (child-csmf-figure.png) and neonates (neonate-csmf-figure.png). An all age CSMF file also available (**csmf-figure-png**)

The **gbd-level1-csmf.png** file is a pie-chart that illustrates in graphical form the breakdown by broad disease group.

The **causes-of-death.csv** file provides a breakdown by defined age-group and sex for the different causes in SmartVA. This can be useful to develop simple tables and graphs for specific age-groups of interest.

4 – monitoring and quality

This folder provides one main file (**adult-likelihoods.xlsx; child-likelihoods.xlsx; neonate-likelihoods.xlsx**) for each of the modules (adult, child and neonate) for inspecting the individual VA data. This file outputs the VA-ID, age, sex and up to three causes of death predicted by SmartVA Analyze. These causes have a likelihood score with colour-coding (Very Likely, Likely, Somewhat Likely, Possible) and the key endorsed symptom questions that contributed to this assigned cause of death along with a column containing all endorsed symptom questions for the particular VA record.

In an intermediate folder within the main monitoring-and-quality folder there are additional files that available. These include files containing the VA raw data (**adult-raw-data.csv; child-raw-data.csv, neonate-raw-data.csv**), files that have converted raw data into binary (0 and 1) for analysis (**adult-symptom.csv; child-symptom.csv; neonate-symptom.csv**) and files that shows the endorsement rates, or the percentage of respondents who answered "yes" to a particular question for each predicted cause category(**adult-endorsement-rates.csv; child-endorsement-rates.csv; neonate-**

[3] Murray CJL, Lopez AD, et al. Population Health Metrics Research Consortium gold standard verbal autopsy validation study: design, implementation, and development of analysis datasets**.** *Population Health Metrics*. **2011; 9:27.**

[4] http://www.healthdata.org/gbd

endorsement-rates.csv). The csv likelihood files (**adult-likelihoods.csv; child-likelihoods.csv; neonate-likelihoods.csv.** These have identical information to the corresponding Excel file in the main folder and can be used if there is a problem with the Excel file. In addition, a file called **VA-data-age-groupings.csv** includes summary output data with counts for male and female by standard age groupings. This file can be used to compare VA data with other sources of mortality data.

A text file report file (**report.txt**) that reports the following:

Analysis parameters: where the input and output files are located, the country, whether it was stated as a malaria and HIV area.

Quality summary: Number of entries (rows) that have duplicate or missing sids, number of entries (rows) where interview was declined. Number of entries (rows) that did not have valid age data and therefore could not be analysed.

The list of problem entries with their respective row numbers is listed below this. The row numbers relate to the rows in the csv file under analysis.

Cause list for SmartVA against ICD-10 codes

	Code to ICD-10	**WHO ICD definition and comments**
ADULT CAUSES		
GBD Cause Group A: Communicable, maternal, neonatal and nutritional disorders		
AIDS	B24	Unspecified human immunodeficiency virus [HIV] disease
Diarrhea/Dysentery	A09	Other gastroenteritis and colitis of infectious and unspecified origin
Malaria	B54	Unspecified malaria
Maternal	O95	Obstetric death of unspecified cause: Maternal death from unspecified cause occurring during pregnancy, labour and delivery, or the puerperium
Other Infectious Diseases	B99	Other and unspecified infectious diseases
Pneumonia	J22	Unspecified acute lower respiratory infection
TB	A16	Respiratory tuberculosis, not confirmed bacteriologically or histologically
GBD Cause Group B: Non-communicable diseases		
Acute Myocardial Infarction	I24	Other acute ischaemic heart diseases (as for WHO 2014)
Breast Cancer	C50	Malignant neoplasm of breast
Chronic Respiratory Diseases	J44	Other chronic obstructive pulmonary disease
Cervical Cancers	C53	Malignant neoplasm of cervix uteri (WHO VA has C55 for all female reproductive neoplasms)
Cirrhosis	K74	Fibrosis and cirrhosis of liver
Colorectal Cancer	C18	Malignant neoplasm of colon
Diabetes	E14	Unspecified diabetes mellitus
Esophageal Cancer	C15	Malignant neoplasm of oesophagus

Leukemia/Lymphomas	C96	Other and unspecified malignant neoplasms of lymphoid, haematopoietic and related tissue
Lung Cancer	C34	Malignant neoplasm of bronchus and lung
Other Cardiovascular Diseases	I99	Other and unspecified disorders of circulatory system
Other Non-communicable Diseases	UU1*	
Prostate Cancer	C61	Malignant neoplasm of prostate
Chronic Kidney Disease	N18	Chronic Kidney Disease
Stomach Cancer	C16	Malignant neoplasm of stomach
Stroke	I64	Stroke, not specified as haemorrhage or infarction
Other Cancers	C76	Malignant neoplasm of other and ill-defined sites
GBD Cause Group C: Injuries		
Bite of Venomous Animal	X27	Contact with other specified venomous animals
Drowning	W74	Unspecified drowning and submersion
Falls	W19	Unspecified fall
Fires	X09	Exposure to unspecified smoke, fire and flames
Homicide (assault)	Y09	Assault by unspecified means
Other Injuries	X58	Exposure to other specified factors
Poisonings (accidental)	X49	Accidental poisoning by and exposure to other and unspecified chemicals and noxious substances
Road Traffic	V89	Motor- or nonmotor-vehicle accident, type of vehicle unspecified
Suicide (intentional self-harm)	X84	Intentional self-harm by unspecified means
CHILD CAUSES		
GBD Cause Group A: Communicable, maternal, neonatal and nutritional disorders		
AIDS	B24	Unspecified human immunodeficiency virus [HIV] disease
Diarrhea/Dysentery	A09	Other gastroenteritis and colitis of infectious and unspecified origin
Encephalitis	G04	Encephalitis, myelitis and encephalomyelitis
Hemorrhagic fever	A99	Unspecified viral haemorrhagic fever
Malaria	B54	Unspecified malaria
Measles	B05	Measles
Meningitis	G03	Meningitis due to other and unspecified causes
Other Infectious Diseases	B99	Other and unspecified infectious diseases
Pneumonia	J22	Unspecified acute lower respiratory infection
Sepsis	A41	Other sepsis
GBD Cause Group B: Non-communicable diseases		
Child Cancers	C76	Malignant neoplasm of other and ill-defined sites
Child Cardiovascular Diseases	I99	Other and unspecified disorders of circulatory system
Other Defined Causes of Child Deaths	UU2*	Other ill-defined and unspecified causes of mortality
Other Digestive Diseases	K92	Other diseases of digestive system
GBD Cause Group C: Injuries		
Bite of Venomous Animal	X27	Contact with other specified venomous animals
Drowning	W74	Unspecified drowning and submersion

Falls	W19	Unspecified fall
Fires	X09	Exposure to unspecified smoke, fire and flames
Poisonings	X49	Accidental poisoning by and exposure to other and unspecified chemicals and noxious substances
Road Traffic	V89	Motor- or nonmotor-vehicle accident, type of vehicle unspecified
Homicide (assault)	Y09	Assault by unspecified means
NEONATE CAUSES		
Birth asphyxia	P21	Birth asphyxia
Congenital malformation	Q89	Other congenital malformations, not elsewhere classified
Neonatal Meningitis/Sepsis	P36	Bacterial sepsis of newborn
Neonatal Pneumonia	P23	Congenital pneumonia/Unspecified acute lower respiratory infection
Preterm Delivery	P07	Disorders related to short gestation and low birth weight, not elsewhere classified
Stillbirth	P95	Fetal death of unspecified cause

*Non ICD-10 Code to signify other NCD/Other defined causes of childhood death not otherwise included in the SmartVA cause list.