SmartVA Documentation

This document is an instruction manual for using the *IHME Smart Verbal Autopsy* (VA) application. This application implements the Tariff Method for computer certification of VA. It takes verbal autopsy interview data as input and produces cause of death estimates at the individual and population levels.

Table of contents

1. System requirements
2. General description of Tariff 2.0
3. Instructions for use of SmartVA
    a. Prepping input data
    b. Selecting input data
    c. Selecting output location
    d. Defining input parameters
    e. Running SmartVA
    f. Analyzing output files
4. Frequently asked questions

System requirements

- Windows 7 or 8
- At least 2 GB RAM

General description of Tariff 2.0

The IHME Tariff 2.0 Verbal Autopsy cause of death assignment system was designed and validated with the Population Health Metrics Research Consortium (PHMRC) Gold Standard VA database. However, with proper mapping, it can be applied to any VA survey. The program uses tariff scores and ranking against the PHMRC Gold Standard Dataset to assign individual causes of death. Tariffs are cause of death-specific normalized endorsement rates for each symptom reported in the PHMRC Gold Standard dataset. The formula for a tariff for cause/symptom pair (*i*,*j*) is the following:

$$Tariff_{cause\ i,\ \ symptom\ j} = \frac{Endorsement\ Rate_{cause\ i,\ \ symptom\ j} - Median\ Endorsement\ Rate_{symptom\ j}}{Interquartile\ Range_{symptom\ j}}$$

The tariff scores of VAs are calculated by taking the sum of all of the tariff scores for the symptoms that were endorsed by that VA.

Once the tariff scores are calculated for all of the VAs in your dataset, they are compared to the tariff scores for VAs whose true cause of death is known from the PHMRC Gold Standard VA Dataset. The cause of death with the best tariff score when compared to the Gold Standard VAs for that cause of death is then assigned to that VA as the Tariff-Method-assigned cause of death.

The results are then assessed for prediction quality, and low-scoring predictions are marked as indeterminate for individual-level estimates. These causes are redistributed based on country-specific cause fractions for population-level estimates.

Instructions for use of Smart VA

Step 1 – Prepping input VA Data

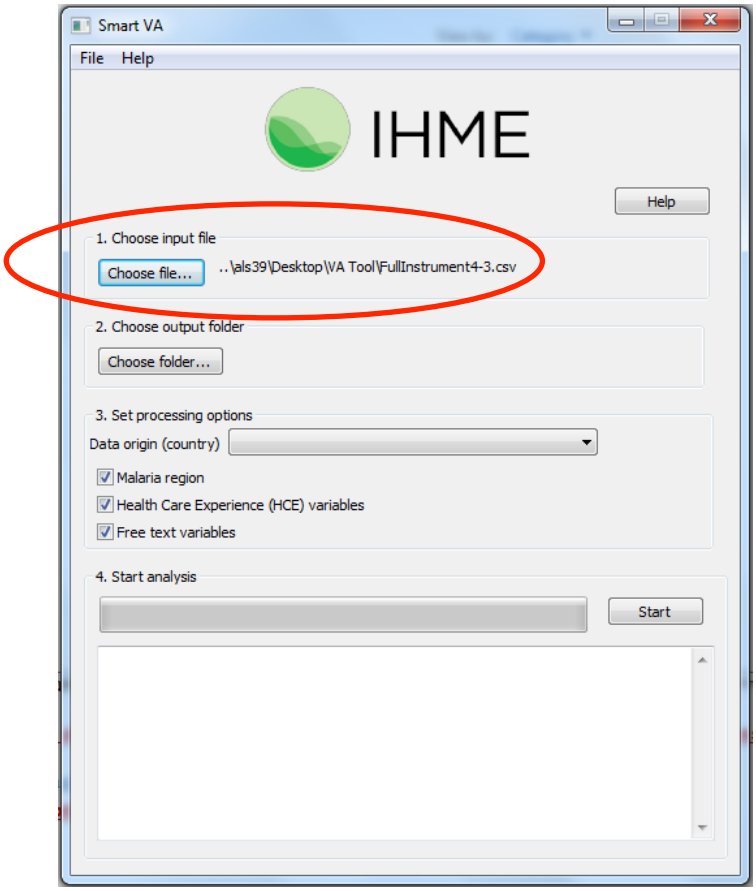
The SmartVA system is currently capable of analyzing VA data collected electronically using the PHMRC instrument on the ODK Collect system on Android devices. The SmartVA system requires as an input the .csv file output from the ODK Briefcase Software for processing such data. ODK Briefcase can be downloaded from: http://opendatakit.org/use/briefcase/

The electronic version of the PHMRC instrument and how to use ODK can be downloaded here:

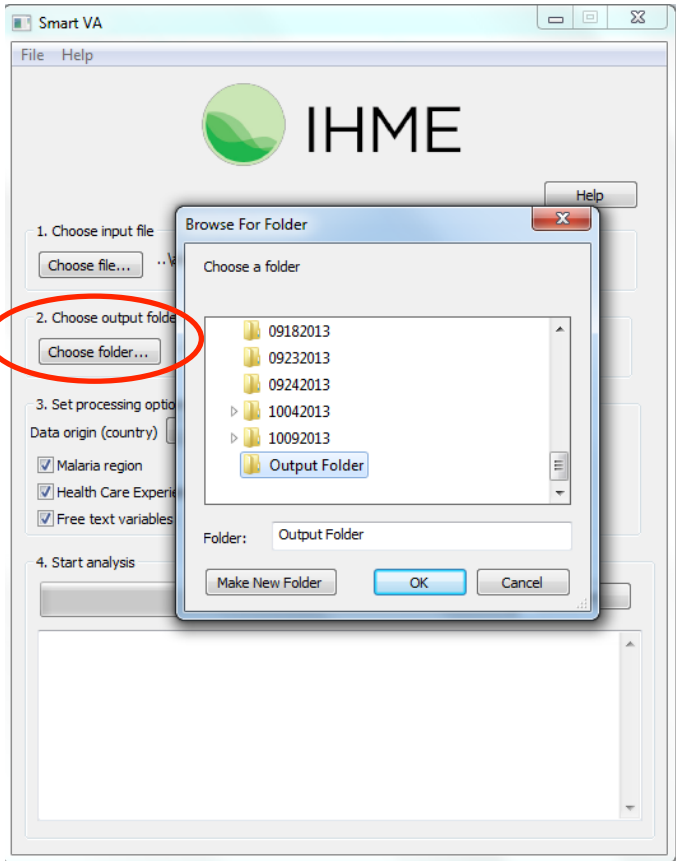
http://www.healthmetricsandevaluation.org/smart-va-application

Step 2 – Selecting input data

Once your data have been processed by ODK Briefcase, you can open SmartVA and select the location of your input data.


Smart VA
File Help
IHME
Help
1. Choose input file
Choose file...
..\als39\Desktop\VA Tool\FullInstrument4-3.csv
2. Choose output folder
Choose folder...
3. Set processing options
Data origin (country)
Malaria region
Health Care Experience (HCE) variables
Free text variables
4. Start analysis
Start

Step 3 – Select output location

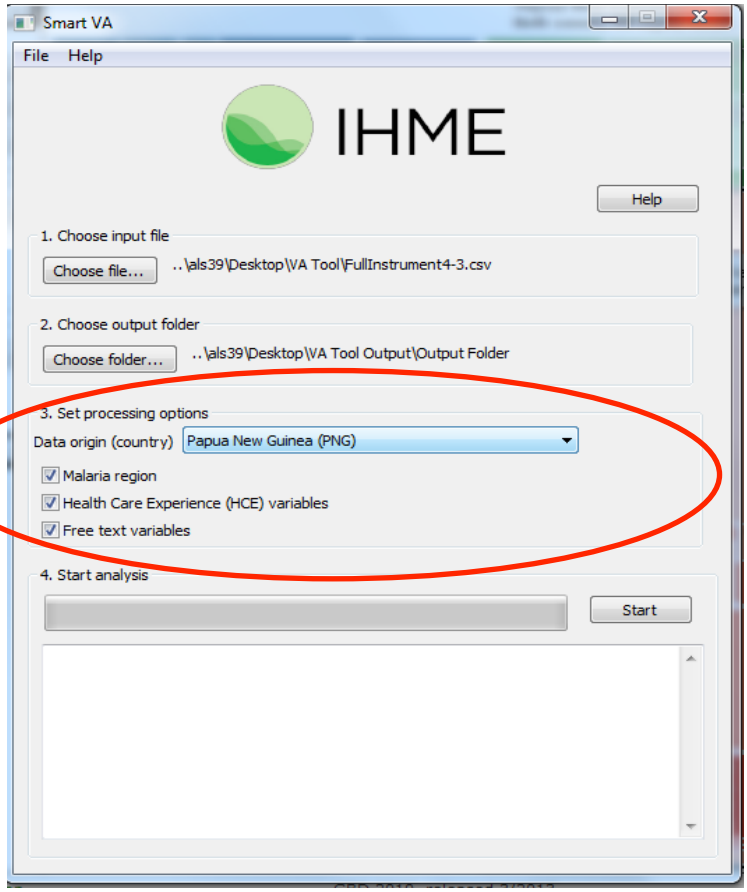


Select where you would like the output from the analysis to be saved.

The output from the Tariff Method will be saved in different subfolders within the folder you select in this step.

Step 4 – Defining input parameters

Four additional options can be specified:



Country of origin

The user may select the country where the VA data were collected. This information is used for reallocation of indeterminate VAs to present results for the entire population of VAs.

Individual observations from the data are not reallocated. Instead, the age and sex distribution of the "indeterminate" VAs in your sample are used to adjust the estimated population-level cause-specific mortality fractions (CSMFs) based on the Global Burden of Disease estimates for the country of VA origin, weighted according to the Tariff Method performance for each of the causes. Since the Tariff Method was developed using validated VAs, it is known which causes of death it underestimates and will adjust the CSMFs of those causes in the final CSMF step.

If no country of origin is specified, the indeterminate VAs will not be reallocated, and an additional category of "indeterminate" will be shown on the final CSMF graphs and CSV files.

Malaria region

The user must determine whether malaria is a possible cause of death in the population from which the VAs were collected. If this the box next to "**Malaria region**" is not selected, the Tariff Method will not assign malaria is a cause of death.

Health care experience variables

The user should determine whether, as part of the survey, questions regarding the health care experience (HCE) of the deceased or his/her family are asked. If the box next to "**Health Care Experience (HCE) variables"** is not checked, these variables are not included in the analysis, and the software will use appropriate training data which are not enhanced with HCE variables.

The following questions in the PHMRC instrument are considered "health care experience:"

- For adults, the question, "Did the deceased have any of the following?" followed by a list of chronic conditions.
- Any data that were transcribed from health records. (This is section 6 of the adult module and section 5 of the child/neonate module).
- For all age modules, responses to the question, "Could you please summarize, or tell us in your own words, any additional information about the illness and/or death of your loved one?"

Free text variables

The Tariff Method has the capability of analyzing open response portions of the VA by turning them into "free text" variables.

In the PHMRC instrument, the open response questions are the following:

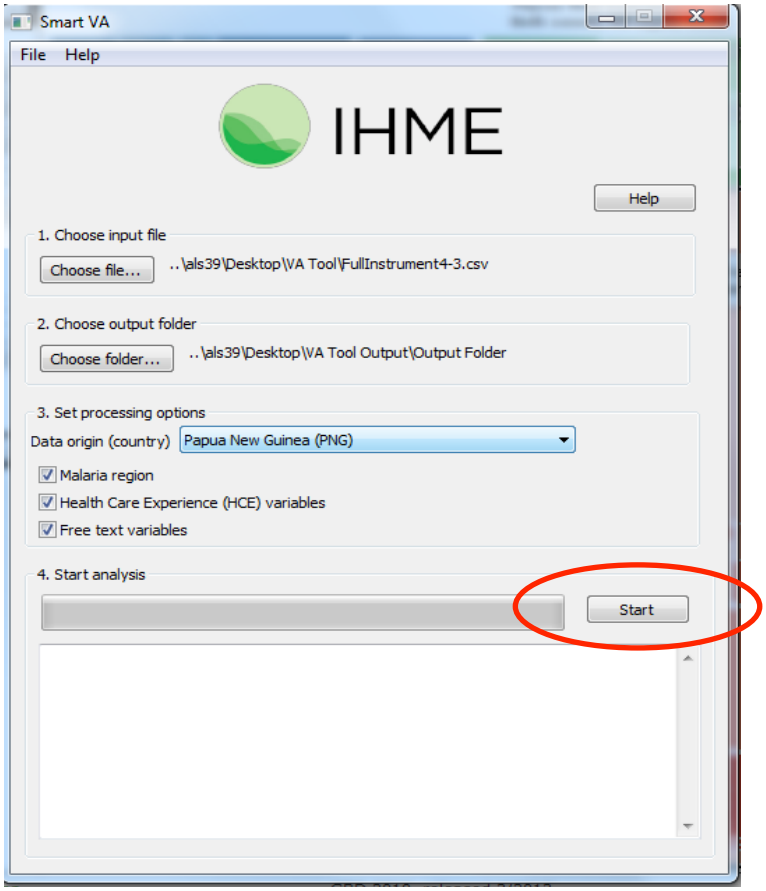
1. "Could you please summarize, or tell us in your own words, any additional information about the illness and/or death of your loved one?"
2. Transcription of medical records and death certificates that are available at the time of interview.

If your data have an open response component and you would like this to be analyzed by the Tariff Method, make sure the box next to "**Free text variables**" is selected.

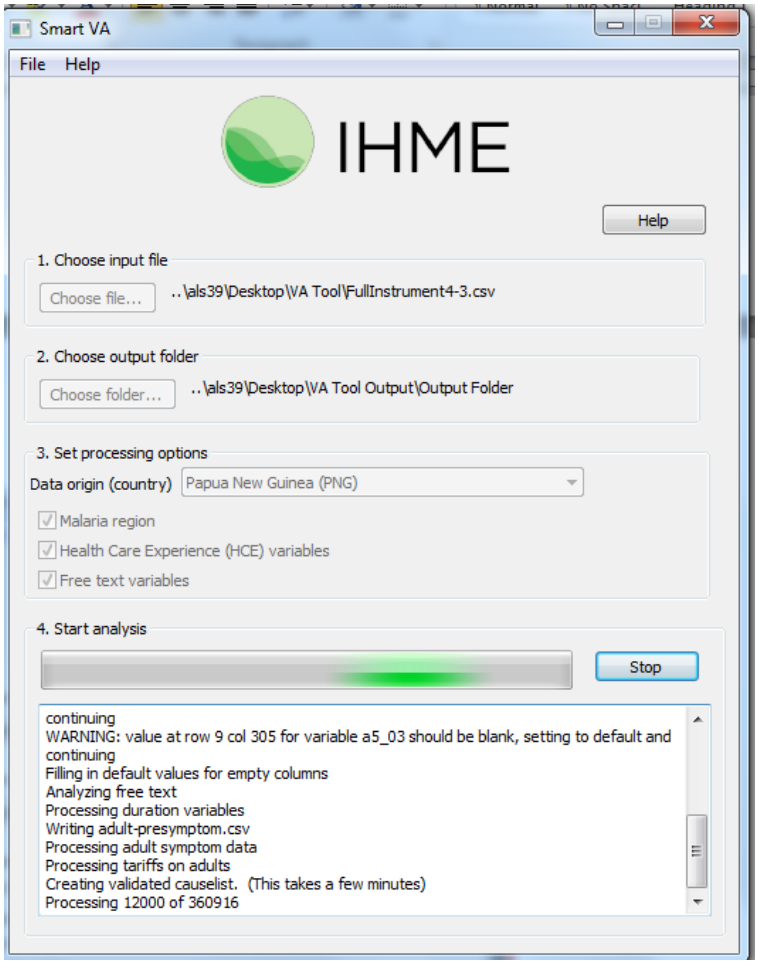
The Tariff Method currently is capable of analyzing open response data only in English.

Step 5 – Running SmartVA

After selecting all of the required inputs, press the “Start” button to begin analysis.



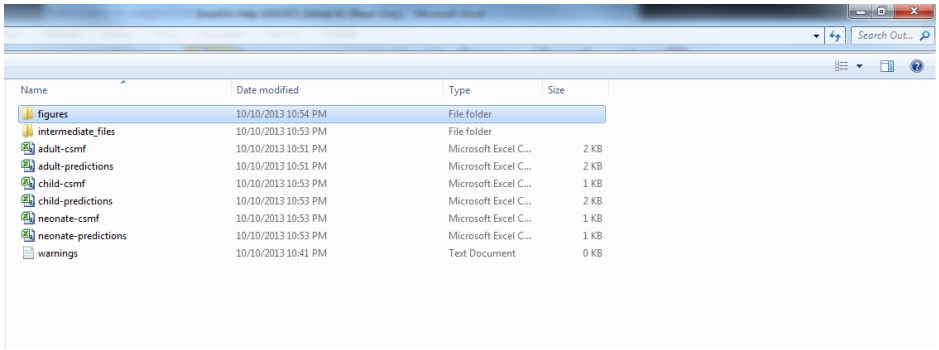
Example of running SmartVA



Step 6 – Analyzing output files

The SmartVA software has several output files. The file containing the individual-level cause of death assignments for your data are the files called “adult-predictions.csv,” “child-predictions.csv,” and “neonate-predictions.csv.”

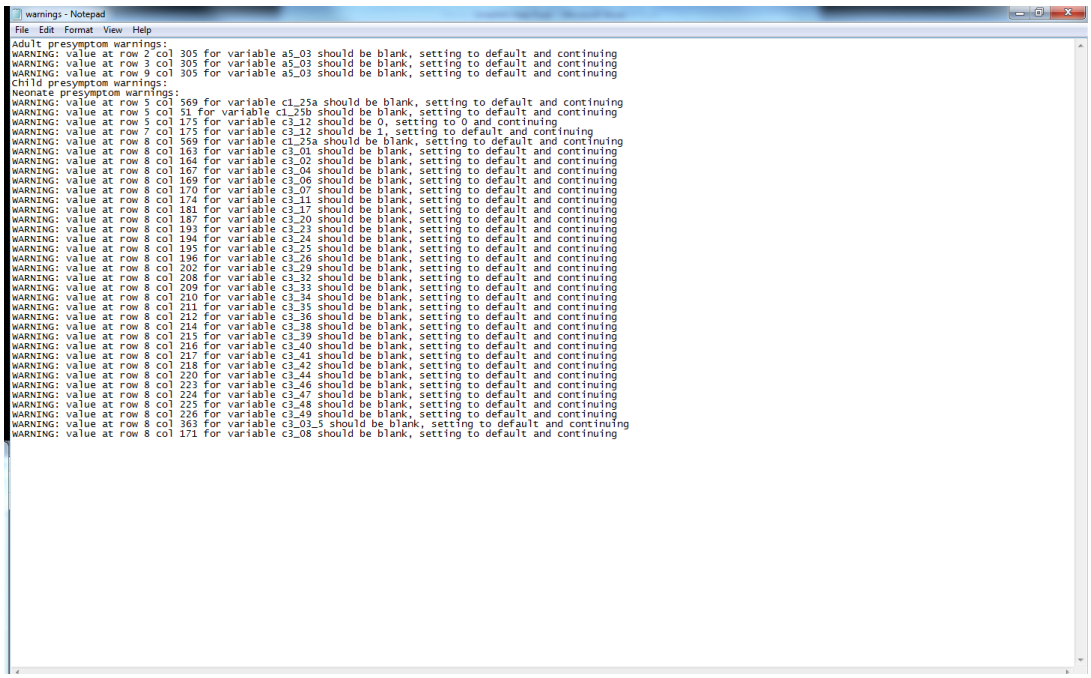
These files can be found in the output folder you specified earlier:



As denoted, each row contains information about one death, including its unique identifier (“sid”), the estimated cause of death, and the age and sex reported on that VA.

	A	B	C	D	E
1	sid	cause	cause34	age	sex
2	Example_VA_1	1	AIDS	31	1
3	Example_VA_2	1	AIDS	31	1
4	Example_VA_3	21	Maternal	23	1
5	Example_VA_4	21	Maternal	17	1
6	Example_VA_5	21	Maternal	27	1
7	Example_VA_6	9	Diabetes	43	1
8	Example_VA_7	32	Stroke	60	0
9	Example_VA_8	18	Leukemia/Lymphomas	42	0
10	Example_VA_9	30	Road Traffic	27	0
11	Example_VA_10	4	Breast Cancer	55	1
12					
13					
14					

While the program is running, it updates the user by printing update messages both on the user interface and in a file called “warnings” in the output folder. The warnings file alerts the user to any variables that contain illegal values such as skip patterns, violations of the PHMRC instrument, or values that are out of range or unexpected for each of the variables in the input data. If a row contains an illegal value, the software will reset this value to a default value and continue analyzing the data.



This output is showing that some observations in the VA dataset had values for variables that should have been skipped according to the PHMRC instrument.

The other subfolders in the output folder contain intermediate files that the Tariff Method requires to run and graphs that show the CSMFs for each age and cause.

Frequently asked questions

Q: What are the age cutoffs for "Adult," "Child," and "Neonate" VAs?

A: The age cutoffs are:

Age module	Age range
Adult	12 years and older
Child	28 days – 11 years old
Neonate	<28 days old

Q: How does the Tariff Method assign an indeterminate cause of death?

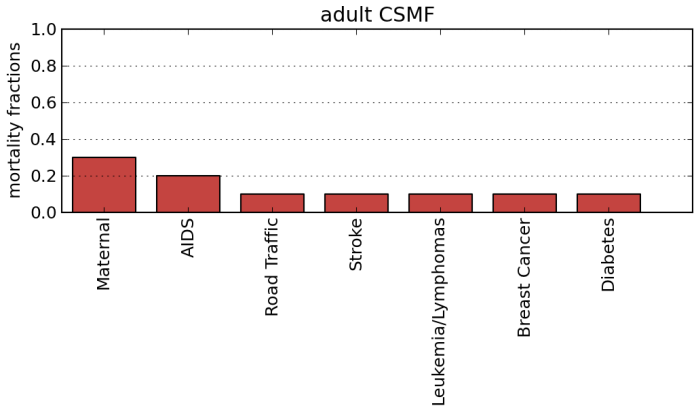
A: The Tariff Method assigns a tariff score to each VA in the input data for every possible cause of death. Each VA's tariff scores are then ranked against the tariff scores from the PHMRC Gold Standard Dataset. If a VA in the input data has tariff scores that are significantly lower than all of the tariff scores in the Gold Standard Dataset, it will receive a cause of death of indeterminate. The tariff scores and ranks can be viewed in the files called "adult-tariff-scores.csv" and "adult-external-ranks.csv," respectively. The "adult-tariff-ranks.csv" files contain the ranks of the VAs after the cutoffs have been applied, which determine which ranks are too high for that VA to be considered for that cause of death.

Q: What are the files "adult-prepped.csv," "adult-presymptom.csv," and "adult-symptom.csv"?

A: These files are the input data in standardized formats that are produced by the software. The "adult-prepped.csv" file contains the raw data from the electronic instrument, the "adult-presymptom.csv" file contains the data in the PHMRC instrument format, and the "adult-symptom.csv" file contains dichotomized or Yes/No variables that are the direct inputs for the Tariff Method analysis.

Q: How do I interpret the graphs?

A: The graphs show bars whose heights are proportional to the estimated cause-specific mortality fraction (CSMF) for each of the causes of death on the cause list for that age module. These graphs include the added weights that were applied from the indeterminate VAs.


adult CSMF
mortality fractions
1.0
0.8
0.6
0.4
0.2
0.0
Maternal
AIDS
Road Traffic
Stroke
Leukemia/Lymphomas
Breast Cancer
Diabetes