VA Data Quality Review Protocol and Report

Bloomberg Philantropies Data for Health Verbal Autopsy Community of Practice (Martin Bratschi, )

November 12, 2018

## Purpose and Use of Protocol / Report

The purpose of this docuemnt is to provide generic guidance on the protocol for the data quality reveiw of electronically (ODK) collected verbal autopsy (VA) data uisng the WHO or PHMRC short questionnaire, which is to be analyzed automatically using openVA or SmartVA-Analyze. The protocol outlines generic data quality reveiw steps that should be taken before the analysis of VA data using the algorithms and it also outlines data quality reveiw steps to be taken after the assignment of a cause-of-death (COD) using the algorithm(s). The document contains the data quality reveiw steps and also produces the report from the data quality reveiw for relevant stakholders to take the next steps. The review protocol / report can be used as a stand alone docuemnt for the review of VA data and the document is also part of the openVA Peipline which automatically transfers data from ODK Aggregate to DHIS2. As requiered for use as part of the openVA Pipelien, all of the data quality reveiw steps catogorise recorsd into one of the folowing three catergories: “ready4Analysis”, “needReview”, and “notSuitable4Analysis”. This protocol / report is produced as part of the Bloomberg Philantropies Data for Health Initiative. It is available from <https://github.com/verbal-autopsy-software> under a GNU General Public License v3.0.

## General Information about the Data Analyzed

#Time of this run.  
Sys.time()

## [1] "2018-11-26 10:03:26 +08"

#Number of rows and columns in the data set.  
dim(VAdata)

## [1] 117 527

## Pre-Algorithm VA Data Quality Review

The following set of check will check the data quality of the collected VA data before the data is to be processed by an algorithm.

### Time taken for interview

The following will check the time taken for the interview and catagorise VAs which have taken less then 10 minutes as “notSuitable4Analysis”.

#Generate variables for reveiw  
  
#Plot / table output of the reveiw  
  
#Generate new variable for output of this check  
  
#Table with sorting of VAs into the three categories

## Algorithm

The following is the command used to run the algorithm. If uisng PHMRC short questionnaire use that algorithm and import data back into R and merge to with algorithm input data set.

#Code for running the algorithm.

## Post-Algorithm VA Data Quality Review

The following set of check will check the data quality of the collected VA data after the data is to be processed by an algorithm.

### Non-sensical CODs (just an example)

The following will check if there are some VAs that have a non sensical CODs (just an example).

#Generate variables for reveiw  
  
#Plot / table output of the reveiw  
  
#Generate new variable for output of this check  
  
#Table with sorting of VAs into the three categories

## Conclusion of VA Data Quality Reveiw

The following shows the conclusion of the VA Data Quality Review.

#Table with sorting of VAs into the three categories