**BASIC STEPS FOR CAUSE OF DEATH (COD) ANALYSIS**

**INTRODUCTION:**

Physician-coded Verbal Autopsy (PCVA) has been used historically to determine a cause of death (COD) from the signs and symptoms reported during Verbal Autopsy (VA) interviews. To account for bias, multiple physicians assess each VA questionnaire record. However, PCVA is time-intensive, and repeatability can be low. To address these challenges, and to support scalability for national COD databases, Computer-Coded Verbal Autposy (CCVA) algorithms are used to determine causes of death (CODs). Below three algorithms are implemented: InterVA5, InSilicoVA, and EAVA. InterVA5 and InSilicoVA can assign CODs for any age, but EAVA is limited to CODs for children under five years of age.

**REQUIREMENTS FOR ANALYSIS:**

In addition to 2016 WHO Questionnaire data and R Studio there are two R packages which are generally helpful for data processing: *data.table, tidyverse*

Three R packages are required to run the CCVAs:

1. The CrossVA package transforms 2016 WHO Verbal Autopsy Questionnaire data to the input format used by the openVA package
2. The openVA package is used to run InterVA5 and InSilicoVA
3. The EAVA package is used to run EAVA - one function transforms VA questionnaire data, another function outputs CODs

**OBTAIN COD RESULTS:**

**InterVA5 and InSilicoVA**

* Step 1 🡪 Use R Studio to read in the 2016 VA data and load required packages
* Step 2 🡪 Remove stillbirths and create/reformat variables required to run CCVAs
* Step 3 🡪 Use odk2openVA\_v151 to convert responses into a Y/N binary format and ensure missing values appear as “.”
* Step 4 🡪 Reorder column names if needed
* Step 5 🡪 Run the codeVA function to obtain CODs for InterVA5 and/or InSilicoVA

**InterVA5 and InSilicoVA Example**

library(openVA)
library(CrossVA)

# read in sample data
example\_file <- system.file("sample", "who151\_odk\_export.csv", package = "CrossVA")

# run CrossVA to map variables to openVA algorithm input format
records <- read.csv(example\_file, stringsAsFactors = FALSE)
mapped\_records <- odk2openVA\_v151(records, id\_col = "KEY")

######## get CODs using two openVA algorithms
# InterVA algorithm
set.seed(123)
get\_InterVA5\_CODs <- codeVA(mapped\_records, data.type = "WHO2016",
 model = "InterVA", version = "5",
 HIV = "h", Malaria = "h", write=FALSE)

InterVA5\_CODs <- getTopCOD(get\_InterVA5\_CODs)
InterVA5\_CODs

# InSilicoVA algorithm
set.seed(123)
get\_InSilicoVA\_CODs <- codeVA(mapped\_records, data.type = "WHO2016",
 model = "InSilicoVA", Nsim=10000, version = "5.0",
 HIV = "h", Malaria = "h")

InSilicoVA\_CODs <- getTopCOD(get\_InSilicoVA\_CODs)
InSilicoVA\_CODs

**EAVA**

* Step 1 🡪 Use R Studio to read in the 2016 VA data and load required packages
* Step 2 🡪 Use odk2EAVA to convert questionnaire data to EAVA input format (note: this is similar to odk2openVA\_v151 for InterVA5 and InSilicoVA)
* Step 3 🡪 Run the codEAVA function to obtain CODs for EAVA

**EAVA Example**

use de-identified data which downloads as *Comsa\_VASA\_20240412* and is available publicly: <https://comsamozambique.org/data-access>

# load EAVA package
library(EAVA)
library(stringi)
library(stringr)

# load publicly available data
data\_public <- read.csv(file.path(file,"Data/Comsa\_WHO\_VA\_20240412.csv"), stringsAsFactors = FALSE, na.strings=c("NULL",""))

data <- as.data.frame(data\_public)

# run odk2EAVA to convert questionnaire data to EAVA input format
output <- odk2EAVA(data, id\_col = "comsa\_id")

# run codEAVA to get CODs for neonates and children 1-to-59 months of age
EAVA\_neonate <- codEAVA(output, "neonate")
EAVA\_child <- codEAVA(output, "child")

# check results
head(EAVA\_neonate)
#> ID cause
#> 21 12127 Pneumonia
#> 28 5527 Unspecified
#> 48 7580 Unspecified
#> 52 11224 Sepsis
#> 58 13674 Intrapartum
#> 64 3868 Pneumonia
head(EAVA\_child)
#> ID cause
#> 2 9241 Pneumonia
#> 5 12103 Unspecified
#> 10 6290 Pneumonia
#> 11 15667 Other infections
#> 13 17097 Diarrhea/Dysentery
#> 17 19372 Unspecified