openVA Analysis with R Shiny

2018 May 25

Vignette

- *Introduction to VA Analysis using openVA the R Shiny App in Windows 10*
  - 2017 September 28
    - http://openva.net/openVA_files/shiny-openVA-vignette.pdf
  - **2017 October 11**
    - https://github.com/jarathomas/openVA-Pipeline/blob/master/ShinyDownload/shiny-openVA-vignette.pdf

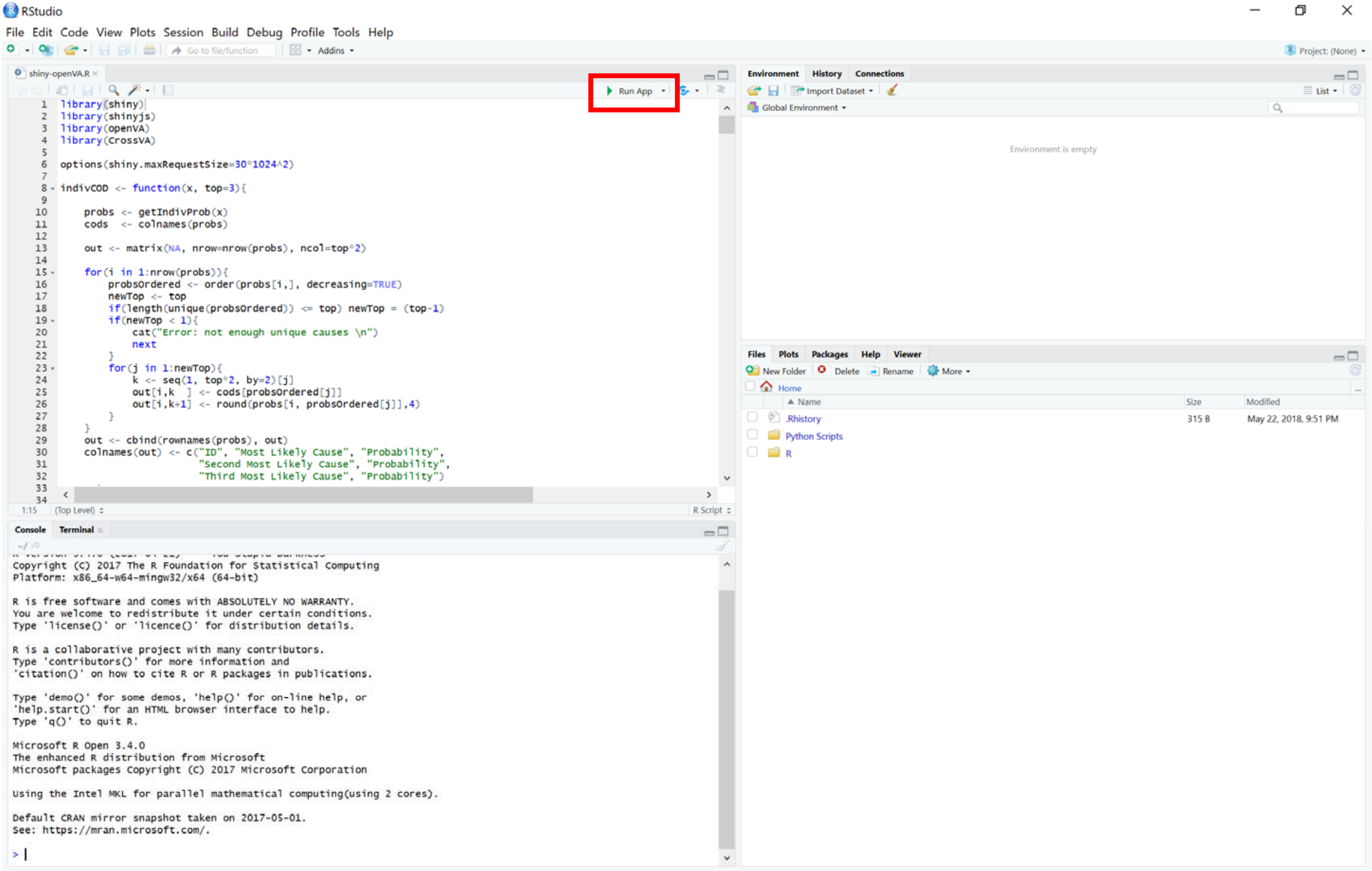
- Currently only works on Windows
  - I’m working on a Linux method

To Download

- openVA Shiny Package
  - https://github.com/jarathomas/openVA-Pipeline/tree/master/ShinyDownload/shiny-openVA.zip
- Java SE 7 or higher
  - http://www.oracle.com/technetwork/java/javase/downloads/index.html
- Download R
  - https://cran.r-project.org/bin/windows/base/
  - Microsoft R Open also works
    - https://mran.microsoft.com/open
- **RStudio (Desktop works fine)**
  - https://www.rstudio.com/products/rstudio/download/

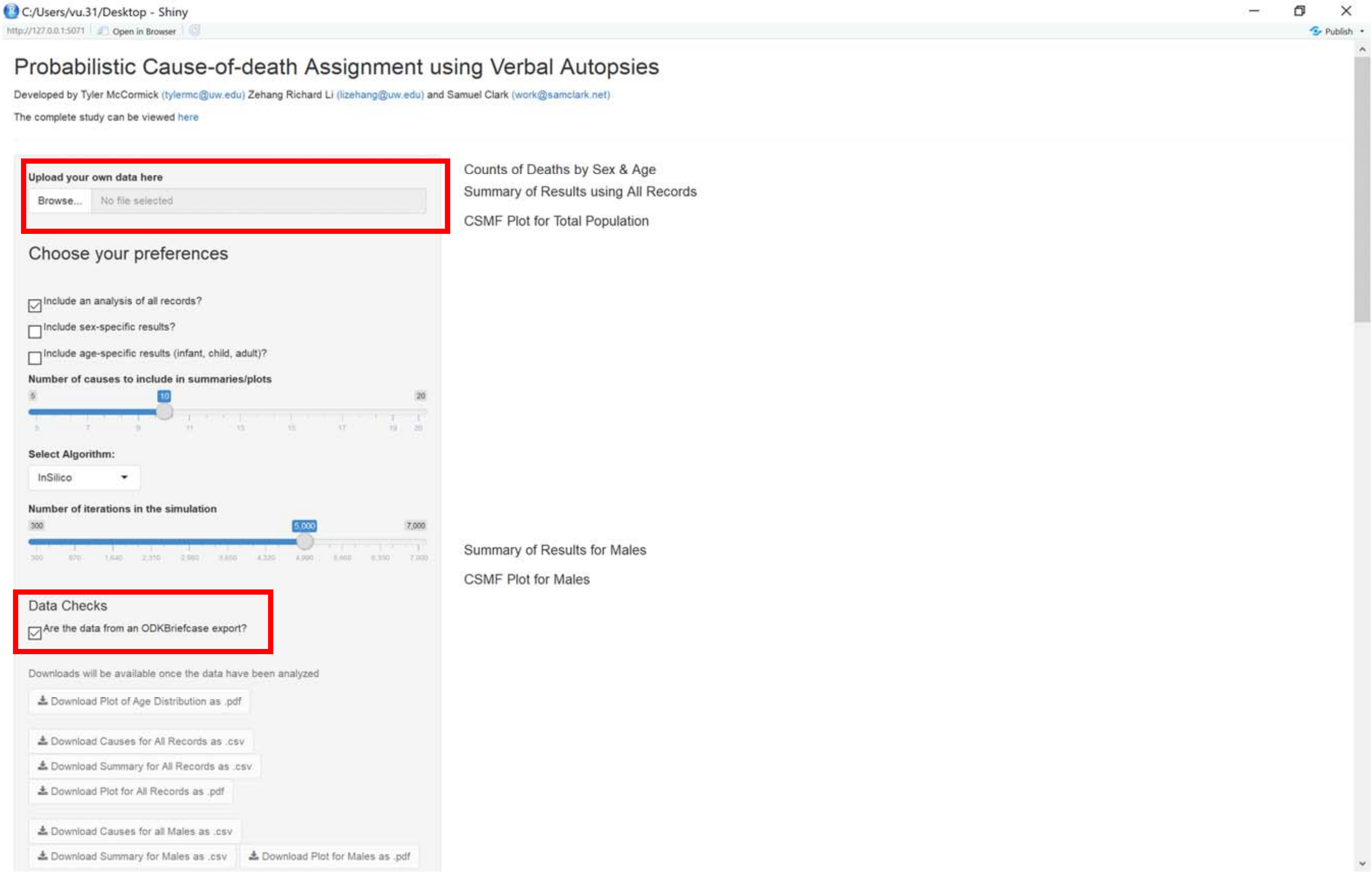
In RStudio

- Select **File** then **Open File** `...`
- Open your `shiny-openVA.R` script



```r
library(shiny)
library(shinyjs)
library(openVA)
library(CrossVA)

options(shiny.maxRequestSize=30*1024^2)

indivCOD <- function(x, top=3){

    probs <- getIndivProb(x)
    cods  <- colnames(probs)

    out <- matrix(NA, nrow=nrow(probs), ncol=top*2)

    for(i in 1:nrow(probs)){
        probsOrdered <- order(probs[i,], decreasing=TRUE)
        newTop <- top
        if(length(unique(probsOrdered)) <= top) newTop = (top-1)
        if(newTop < 1){
            cat("Error: not enough unique causes \n")
            next
        }
        for(j in 1:newTop){
            k <- seq(1, top*2, by=2)[j]
            out[i,k  ] <- cods[probsOrdered[j]]
            out[i,k+1] <- round(probs[i, probsOrdered[j]],4)
        }
    }
    out <- cbind(rownames(probs), out)
    colnames(out) <- c("ID", "Most Likely Cause", "Probability",
                       "Second Most Likely Cause", "Probability",
                       "Third Most Likely Cause", "Probability")
```

```
Copyright (C) 2017 The R Foundation for Statistical Computing
Platform: x86_64-w64-mingw32/x64 (64-bit)

R is free software and comes with ABSOLUTELY NO WARRANTY.
You are welcome to redistribute it under certain conditions.
Type 'license()' or 'licence()' for distribution details.

R is a collaborative project with many contributors.
Type 'contributors()' for more information and
'citation()' on how to cite R or R packages in publications.

Type 'demo()' for some demos, 'help()' for on-line help, or
'help.start()' for an HTML browser interface to help.
Type 'q()' to quit R.

Microsoft R Open 3.4.0
The enhanced R distribution from Microsoft
Microsoft packages Copyright (C) 2017 Microsoft Corporation

Using the Intel MKL for parallel mathematical computing(using 2 cores).

Default CRAN mirror snapshot taken on 2017-05-01.
See: https://mran.microsoft.com/.

> 
```



Probabilistic Cause-of-death Assignment using Verbal Autopsies

Developed by Tyler McCormick (tylermc@uw.edu) Zehang Richard Li (lizehang@uw.edu) and Samuel Clark (work@samclark.net)

The complete study can be viewed here

Upload your own data here

Choose your preferences

☑ Include an analysis of all records?

☐ Include sex-specific results?

☐ Include age-specific results (infant, child, adult)?

Number of causes to include in summaries/plots: 10

Select Algorithm: InSilico

Number of iterations in the simulation: 5,000

Data Checks

☑ Are the data from an ODKBriefcase export?

Downloads will be available once the data have been analyzed

Counts of Deaths by Sex & Age

Summary of Results using All Records

CSMF Plot for Total Population

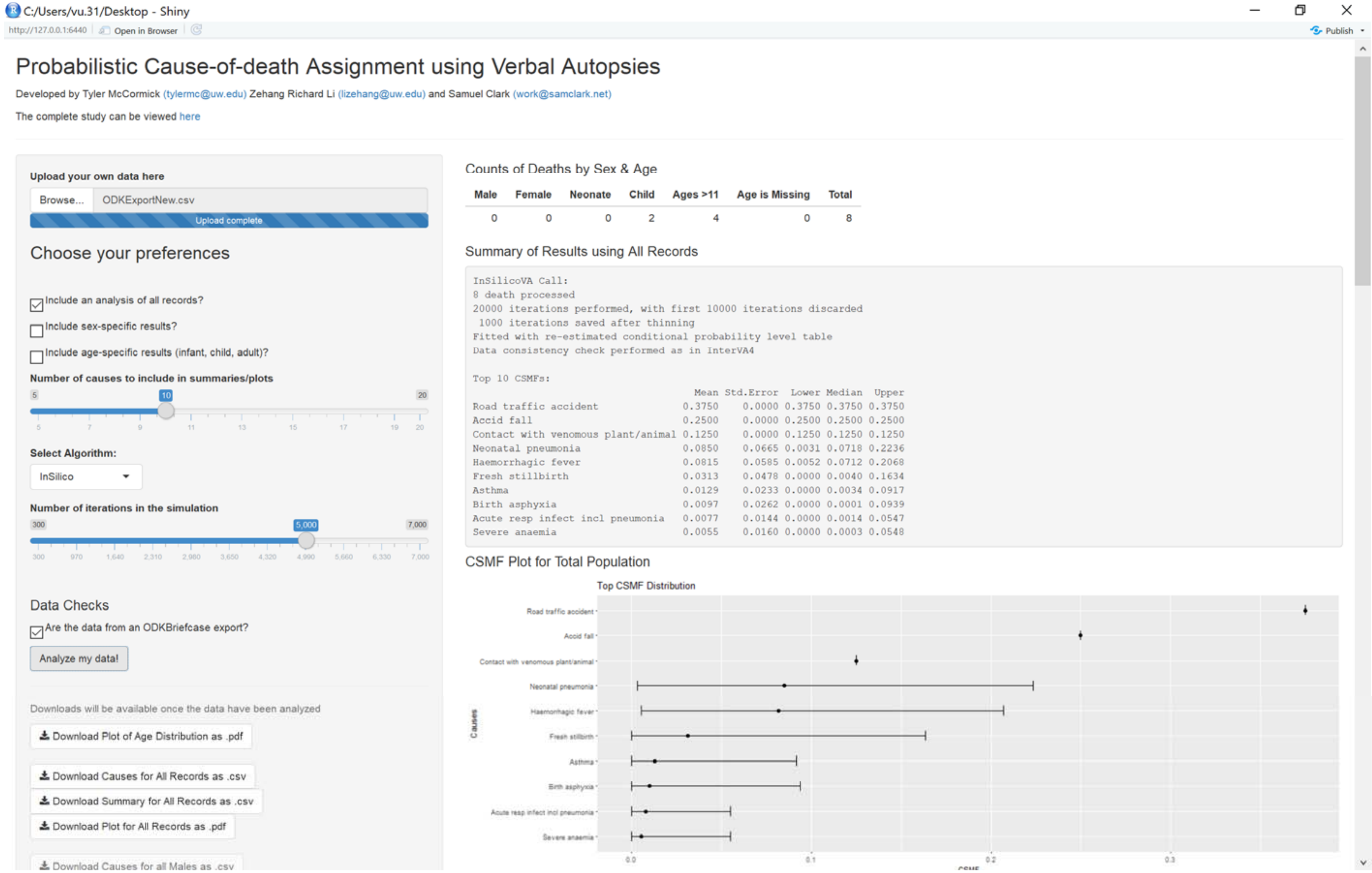
Summary of Results for Males

CSMF Plot for Males

Pipeline Specific Look

- The `shiny-openVA.zip` you will have downloaded includes a file `who_va_output.csv` that the vignette uses.
  - Fifteen records

- Running the pipeline will create a file named `ODKExportNew.csv` under a subfolder `ODKExport` via the ODK Briefcase `.jar`.
  - Our current demo contains 8 records



Probabilistic Cause-of-death Assignment using Verbal Autopsies

Developed by Tyler McCormick (tylermc@uw.edu) Zehang Richard Li (lizehang@uw.edu) and Samuel Clark (work@samclark.net)

The complete study can be viewed here

Upload your own data here

Choose your preferences

☑ Include an analysis of all records?

☐ Include sex-specific results?

☐ Include age-specific results (infant, child, adult)?

Number of causes to include in summaries/plots: 10

Select Algorithm: InSilico

Number of iterations in the simulation: 5,000

Data Checks

☑ Are the data from an ODKBriefcase export?

Downloads will be available once the data have been analyzed

Counts of Deaths by Sex & Age

Male	Female	Neonate	Child	Ages >11	Age is Missing	Total
0	0	0	2	4	0	8

Summary of Results using All Records

```
InSilicoVA Call:
8 death processed
20000 iterations performed, with first 10000 iterations discarded
 1000 iterations saved after thinning
Fitted with re-estimated conditional probability level table
Data consistency check performed as in InterVA4

Top 10 CSMFs:
                                   Mean Std.Error  Lower Median  Upper
Road traffic accident            0.3750    0.0000 0.3750 0.3750 0.3750
Accid fall                       0.2500    0.0000 0.2500 0.2500 0.2500
Contact with venomous plant/animal 0.1250    0.0000 0.1250 0.1250 0.1250
Neonatal pneumonia               0.0850    0.0665 0.0031 0.0718 0.2236
Haemorrhagic fever               0.0815    0.0585 0.0052 0.0712 0.2068
Fresh stillbirth                 0.0313    0.0478 0.0000 0.0040 0.1634
Asthma                           0.0129    0.0233 0.0000 0.0034 0.0917
Birth asphyxia                   0.0097    0.0262 0.0000 0.0001 0.0939
Acute resp infect incl pneumonia 0.0077    0.0144 0.0000 0.0014 0.0547
Severe anaemia                   0.0055    0.0160 0.0000 0.0003 0.0548
```

CSMF Plot for Total Population

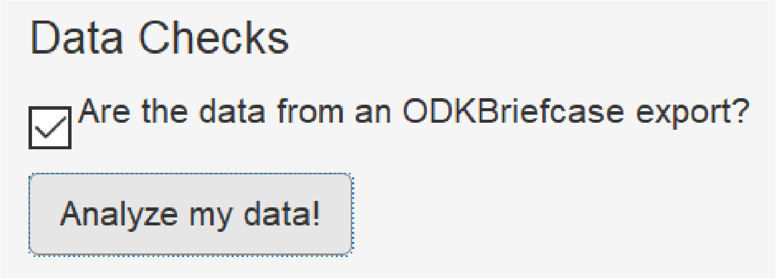
Choose your preferences

- **Include an analysis of all records?** – This option is selected by default, but it is not necessary (e.g., you could solely request sex-specific results).
- **Include sex-specific results?** – Simply click in the empty box to produce separate results for males and females.
- **Include age-specific results?** – Simply click in the empty box to produce separate results for infants (less than 1 year old), children (aged 1-4 years old), and adults (ages 5 and above).
- **Number of causes to include in summaries/plots** – The Shiny app will display the VA results in several tables and plots showing a list of the top (i.e., most prevalent) causes of death. Slide the bar right (or left) to include more (less) causes in the summaries of the results. The default is to include 10 causes, with a minimum and maximum of 5 and 20, respectively – this was simply a design choice for the app, openVA can include any number of causes the user wants.

Select Algorithm

- **InSilico** – The algorithm computes the causes of death (COD) by performing a specific number of iterations, which can be chosen with the **Number of iterations in the simulation** slider (the default is 5,000). Choosing a larger number of iterations will result in a longer waiting time for the algorithm to finish, but will also help improve the approximate results. Part of the InSilico algorithm is implemented in Java and thus a separate windows will appear with a Java Progress Bar indicating the percentage of iterations that have been completed.
- **InterVA4** – The user can specify the **Level of HIV prevalence** and **Level of malaria prevalence** which are the same parameters required by the InterVA-4 software developed by Peter Byass.

Analyze my data!



Data Checks

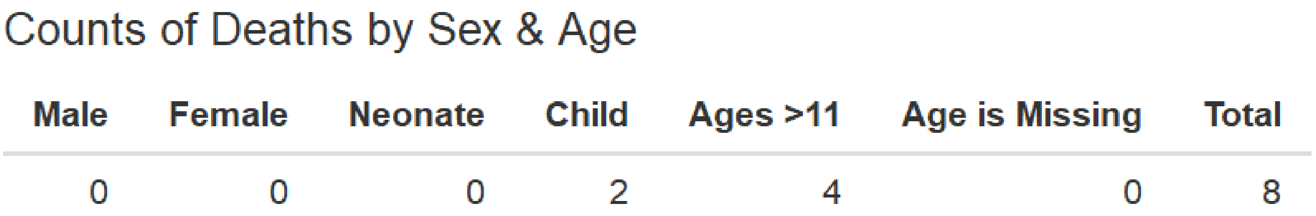
☑ Are the data from an ODKBriefcase export?

- With 500 cases it may take anywhere from 5 to 30 minutes depending on the algorithm (InSilico typically takes longer, and the waiting time will obviously increase with the number of iterations for the simulation

Download Options

- **Download Plot of Age Distribution as .pdf** – Contains a histogram of the number of deaths in the three age groups (infants, children, and adults). The corresponding counts are displayed in the table (appearing in the right side of the browser) entitled "Counts of Deaths by Sex & Age".
- **Download Causes for All Records as .csv** – This file includes the record ID (which is simply the row number in the original data file) and the three most likely causes of death and the associated probabilities; note that some records may be excluded due to data errors (i.e., missing necessary data/indicators).
- **Download Summary for All Records as .csv** – This file contains the information appearing in the web browser under the title "Summary of Results using All Records".
- **Download Plot for All Records as .pdf** – This file contains the CSMF displayed on the right side of the web browser under the title "CSMF Plot for Total Population".
- The options for downloading causes and summaries (as a .csv files) and the plots (as a .pdf) are also available for each group included in the analysis.
- **Download warnings as .txt** – This option is located at the very bottom of the panel on the left side of the web browser. The associated text file includes the warning and error messages generated by InSilico and the issues encountered and, in some cases, the steps taken to reconcile the inconsistencies in the data.

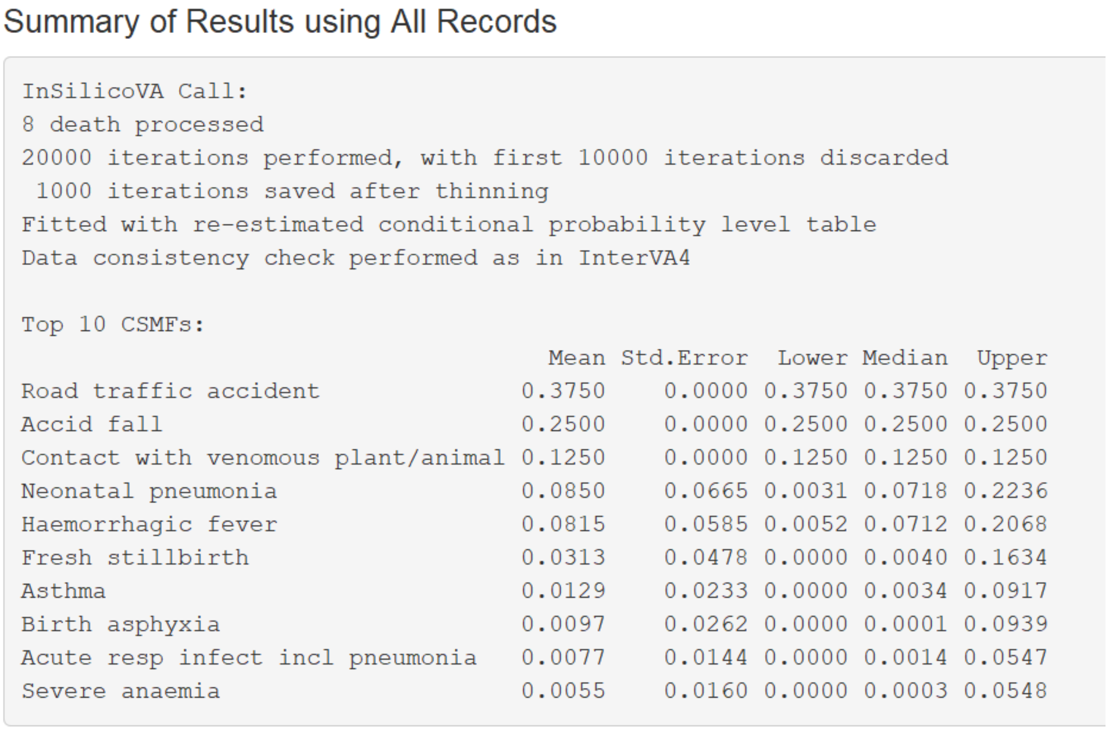
Counts of Deaths by Sex & Age



Counts of Deaths by Sex & Age

Male	Female	Neonate	Child	Ages >11	Age is Missing	Total
0	0	0	2	4	0	8

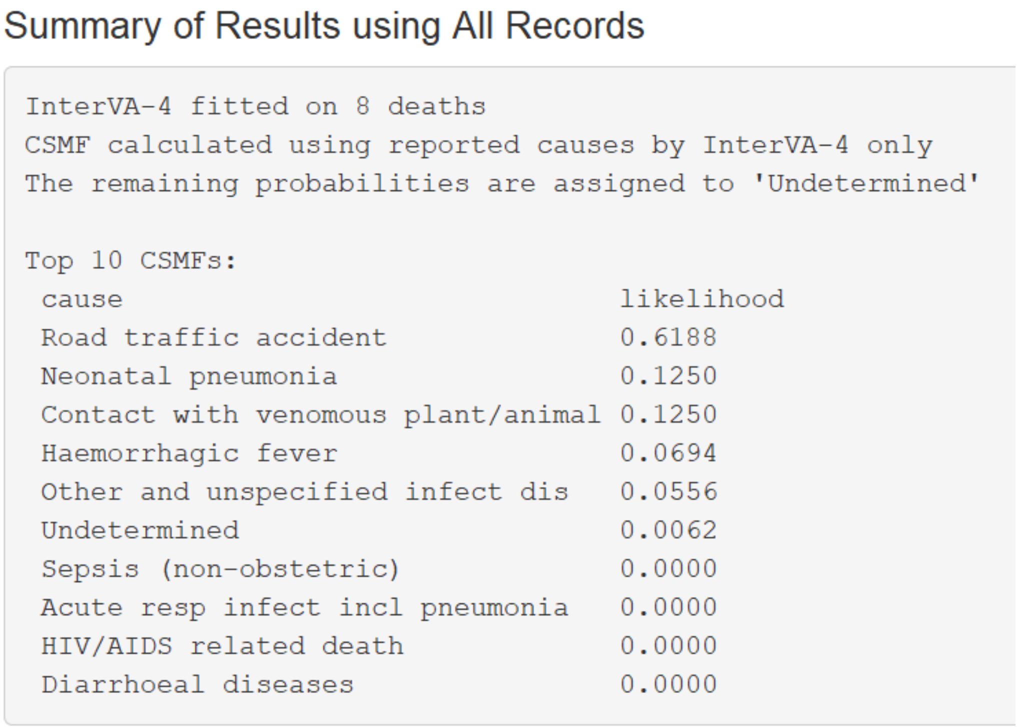
Summary of Results using All Records (InSilico)



Summary of Results using All Records

```
InSilicoVA Call:
8 death processed
20000 iterations performed, with first 10000 iterations discarded
 1000 iterations saved after thinning
Fitted with re-estimated conditional probability level table
Data consistency check performed as in InterVA4

Top 10 CSMFs:
                                    Mean Std.Error  Lower Median  Upper
Road traffic accident             0.3750    0.0000 0.3750 0.3750 0.3750
Accid fall                        0.2500    0.0000 0.2500 0.2500 0.2500
Contact with venomous plant/animal 0.1250    0.0000 0.1250 0.1250 0.1250
Neonatal pneumonia                0.0850    0.0665 0.0031 0.0718 0.2236
Haemorrhagic fever                0.0815    0.0585 0.0052 0.0712 0.2068
Fresh stillbirth                  0.0313    0.0478 0.0000 0.0040 0.1634
Asthma                            0.0129    0.0233 0.0000 0.0034 0.0917
Birth asphyxia                    0.0097    0.0262 0.0000 0.0001 0.0939
Acute resp infect incl pneumonia  0.0077    0.0144 0.0000 0.0014 0.0547
Severe anaemia                    0.0055    0.0160 0.0000 0.0003 0.0548
```

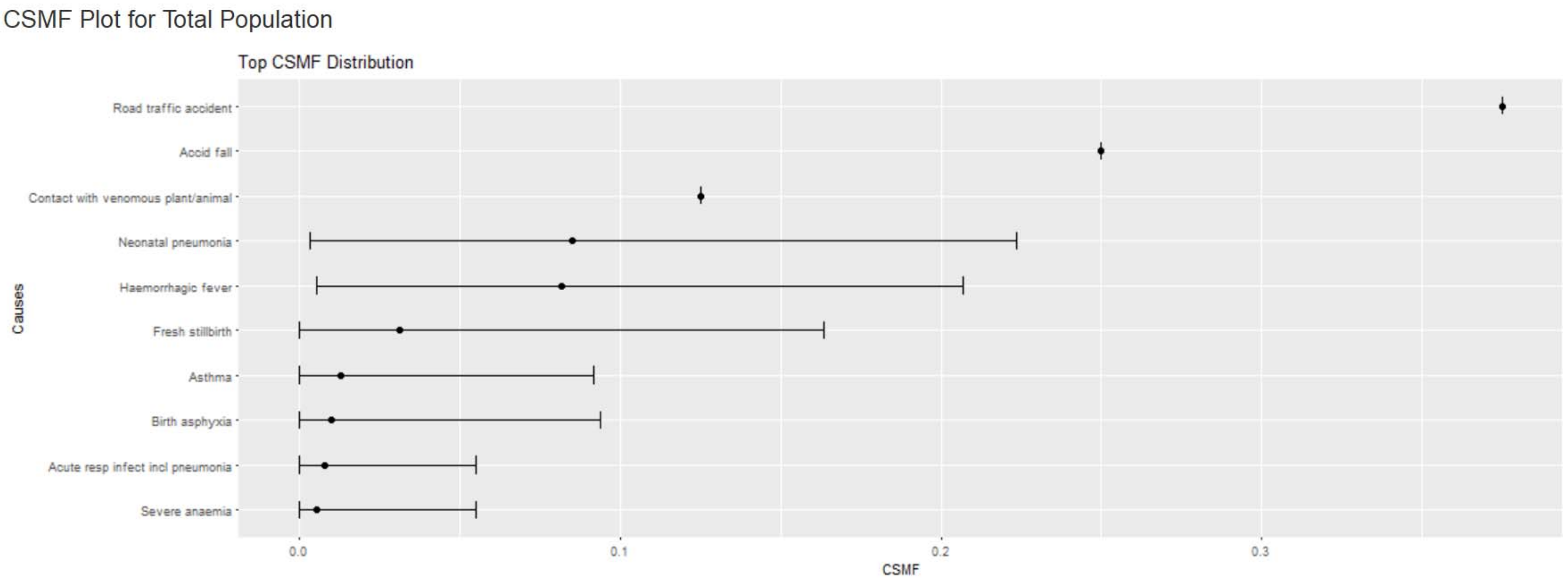
Summary of Results using All Records (InterVA)



Summary of Results using All Records

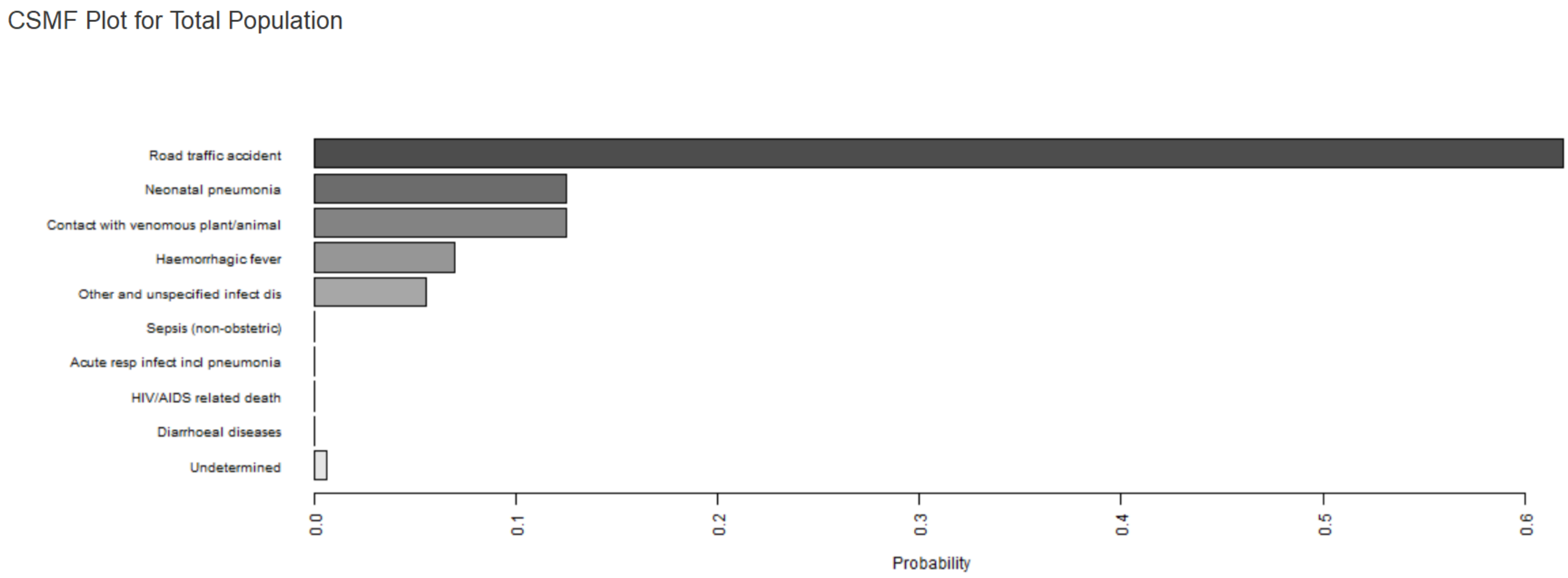
```
InterVA-4 fitted on 8 deaths
CSMF calculated using reported causes by InterVA-4 only
The remaining probabilities are assigned to 'Undetermined'

Top 10 CSMFs:
 cause                              likelihood
 Road traffic accident              0.6188
 Neonatal pneumonia                 0.1250
 Contact with venomous plant/animal 0.1250
 Haemorrhagic fever                 0.0694
 Other and unspecified infect dis   0.0556
 Undetermined                       0.0062
 Sepsis (non-obstetric)             0.0000
 Acute resp infect incl pneumonia   0.0000
 HIV/AIDS related death             0.0000
 Diarrhoeal diseases                0.0000
```

CSMF Plot for Total Population (InSilico)



The results are slightly different because InSilico includes uncertainty (similar to confidence intervals) around the point estimates, whereas InterVA4 does not.

CSMF Plot for Total Population (InterVA4)



The results are slightly different because InSilico includes uncertainty (similar to confidence intervals) around the point estimates, whereas InterVA4 does not.