Demonstration of InterVA-5

There are three stages to using InterVA-5 successfully:

1. Preparing the input data for InterVA-5 in the correct format

2. Running InterVA-5 on the data that have been prepared

3. Processing and interpreting the InterVA-5 output

Verbal autopsy data comes from many sources, which generally have their own specific formats. InterVA-5 has a very specific and standardised format for its input data, in CSV (comma separated variable) format, and data that are not in the correct format will be rejected.

The exact format and content for InterVA-5 data is specified in detail in the InterVA-5 User Guide. There are 353 data columns (specific items from the verbal autopsy interview) plus an identifier column, and each row represents an individual death. The top left corner of an input data file looks like this:

	i004a	i004b	i019a	i019b	i022a	i022b	i022c	i022d	i022e	i022f	i022g
1	–	–	–	Y	–	–	–	–	Y	–	–
2	Y	–	Y	–	–	–	Y	–	–	–	–
3	Y	–	Y	–	–	–	–	–	–	–	Y
4	Y	–	–	Y	Y	–	–	–	–	–	–
5	–	Y	–	Y	–	–	–	–	–	Y	–
6	–	Y	Y	–	Y	–	–	–	–	–	–
7	–	Y	–	Y	–	–	–	–	–	Y	–
8	–	Y	Y	–	Y	–	–	–	–	–	–

3

The important question is how to prepare an input data file in the correct format. Most standard data processing software can be used for this – these examples illustrate the process using Stata.

It's best to write a script that reads the existing verbal autopsy data file, and writes a new file in the InterVA-5 format.

For example, a script can start by reading the input data, and defining new variables for the output file, starting like this:

```
* interVA5_Afghan_in.do - script file to create input for InterVA5

use AMSdeaths_cl, replace

* generate an unique ID from cluster/household/column
gen str10 id=string(10000000+qhclust*10000+qhnumber*10+qh303,"%8.0f")

order id

* set up InterVA5 indicators
gen   i004a ="-"
gen   i004b ="-"
gen   i019a ="-"
gen   i019b ="-"
gen   i022a ="-"
gen   i022b ="-"
gen   i022c ="-"
```

… and continuing like this:

```
gen  i458o ="-"
gen  i459o ="-"

* code InterVA indicators from VA data
replace  i004a ="Y"  if qv305n<=4
replace  i004b ="Y"  if qv305n>=5 & qv305n<=12
replace  i019a ="Y"  if qv30a==1
replace  i019b ="Y"  if qv30a==2
replace  i022a ="Y"  if qv304v==1 & qv304o>=65
replace  i022b ="Y"  if qv304v==1 & qv304o>=50 & qv304o<65
replace  i022c ="Y"  if qv304v==1 & qv304o>=15 & qv304o<50
replace  i022d ="Y"  if (qv304v==1 & qv304o>=5 & qv304o<15) | (qv304v==2 & qv304o>=60 & qv304o<99)
replace  i022e ="Y"  if (qv304v==1 & qv304o>=1 & qv304o<5) | (qv304v==2 & qv304o>=12 & qv304o<60)
```

… until all the output variables are coded and the file can be written:

```
replace  i458o ="N"  if qh604d==2
replace  i459o ="Y"  if strpos(qh517,"A")>0 | strpos(qh517,"B")>0

* drop non-InterVA variables
drop q*
order id, after(i459o)

* save input file
save afghan_iv5, replace

outsheet using afghan_iv5.csv, comma noquote replace
```

The file afghan_iv5.csv is now ready to use as InterVA-5 input.

Having prepared the input data file, the next step is to run InterVA-5 to process the data. InterVA-5 opens with this screen:

InterVA-5 Verbal Autopsy Interpretation System

Free software issued under the terms of the GNU General Public License Version 3

running using probbase v17 180907
expecting 353 input indicators per case
coding to 3 pregnancy status outcomes, 64 cause of death categories and 7 circumstances of death categories

It is necessary to set parameters for local malaria and HIV/AIDS prevalence:

    enter H(igh) or L(ow) or V(ery low) malaria mortality H

    enter V(ery low) or L(ow) or H(igh) HIV/AIDS mortality L

Choose the malaria and HIV/AIDS levels for the location, then enter the name of the input data file:

change or accept batch input file name: afghan_iv5.csv

Next a checking process is run on the input data – which may take some time if there are a lot of cases in the file.

```
delimited output will be stored in afghan_iv5_out_20190424.csv

checking data...

any warning messages will be stored in file afghan_iv5_warnings_20190424.txt

check of afghan_iv5.csv at 04/24/19 19:09:56
```

After the checking process is complete, the software starts processing case-by-case, and writing the output files.

The software shuts down when processing is complete.

InterVA-5 output data is produced in a standard CSV output file, with the same name as the data input file plus a date stamp of its InterVA-5 processing – in this example, afghan_iv5_out_20190424.csv – which can be handled by any standard data processing software.
Again, it's best to write a script that reads the InterVA-5 output and processes it as required. It's important to understand that the InterVA-5 output comes in this format:

iv5_id	malprev	hivprev	pregstat	preglik	cause1	lik1	cause2	lik2	cause3	lik3	indet	comcat	comnum
10010481	V	V	n/a		01.07 Men	47					53	Culture	85
10010611	V	V	n/a		06.02 Live	87					13	Health sys	99
10011871	V	V	n/a		10.02 Birth	99					1	Health sys	78
10020101	V	V	n/a		04.02 Stro	97					3	Resources	90
10030831	V	V	n/a		05.01 Chrc	75					25	Health sys	95
10031351	V	V	n/a		12.09 Assa	49	04.01 Acut	24			27	Emergenc	90
10040161	V	V	n/a		11.01 Fres	99					1	Inevitable	99
10040791	V	V	n/a		11.01 Fres	99					1	Inevitable	99
10040961	V	V	n/a		01.04 Diar	99					1	Knowledg	86
10041071	V	V	n/a		02.03 Resp	98					2	Health sys	55
10041072	V	V	n/a		01.02 Acut	58					42	Culture	78

Since InterVA-5 can generate up to 3 likely causes for each case, and the likelihood of the likely causes may not reach 100%, for analysis it is often useful to transform this format into one record for each cause for each case:

iv5_id	malprev	hivprev	pregstat	preglik	cause	lik	comcat	comnum
10010481	V	V	n/a	.	99 Indeterminate	.53	Culture	85
10010481	V	V	n/a	.	01.07 Meningitis and encephalitis	.47	Culture	85
10010611	V	V	n/a	.	06.02 Liver cirrhosis	.87	Health systems	99
10010611	V	V	n/a	.	99 Indeterminate	.13	Health systems	99
10011871	V	V	n/a	.	99 Indeterminate	.01	Health systems	78
10011871	V	V	n/a	.	10.02 Birth asphyxia	.99	Health systems	78
10020101	V	V	n/a	.	04.02 Stroke	.97	Resources	90
10020101	V	V	n/a	.	99 Indeterminate	.03	Resources	90
10030831	V	V	n/a	.	05.01 Chronic obstructive pulmonary dis	.75	Health systems	95
10030831	V	V	n/a	.	99 Indeterminate	.25	Health systems	95
10031351	V	V	n/a	.	12.09 Assault	.49	Emergency	90
10031351	V	V	n/a	.	12.09 Assault	.24	Emergency	90
10031351	V	V	n/a	.	99 Indeterminate	.27	Emergency	90

In this format, the number of records is unimportant – but the total of the lik column will correspond to the total number of cases in the file. This can then be used as a weighting factor in multivariable analyses.