**UNIVAC\_v161120.R – Documentation & Instructions to run model**

*## Static Cohort Model R version of the UNIVAC Excel Model*

*## Authors: Hira Tanvir, Andrew Clark*

*## Date: 16 November 2020 -- latest version*

The following packages need to be downloaded in a version of R > 3.6.1

* tictoc, pbapply, data.table, extraDistr, foreach, doParallel, parallel, doMPI, tidyverse
* **“mypackage”** – is an R package which I wrote in C++ and has to be complied in R

1. **Instructions to build & compile “mypackage” :**

Instructions can also be found here: **https://kbroman.org/pkg\_primer/pages/build.html**

* **Build** a tar.gz file of the folder “mypackage”
* Go to terminal to the directory containing the package folder
* Enter into terminal: **R CMD build mypackage**
* This will create a tar.gz file maybe called **“mypackage\_1.0.tar.gz”**
* **Compile the package:**
* Enter into terminal: **R CMD INSTALL mypackage**
* Open R and enter library(“mypackage”) to use package
* The packages required for UNIVAC\_v161120.R to run should now be installed on the desktop you are using.

“mypackages” include the following functions:

1. cumulative\_covC
2. impact\_time\_week
3. incremental\_cov\_leakage
4. incremental\_covC
5. rescale\_to\_coverage
6. **Run the Model:**

This model contains options to run both deterministic (base-case) scenarios as well as Probability Sensitivity Analysis (PSA).

* Open R file UNIVAC\_v161120.R
* Set the directory to the **“Sandbox”** folder in R by giving the pathway to it from your desktop to the folder where the model is saved.
* In R: **setwd(“~USERNAME/Documents/UNIVAC\_model/Sandbox/”)**
* Source the functions from a separate R code: “univac\_functions\_v161120.R”
* The rest of the input data should read as it is already set as long as the folders for the inputs have not changed locations – but if they have then you would need to adjust the pathways for these.
* Under “Model set-up” on line 37, select the scenarios for each model run i.e:
  + **p** = Select pathogen Type (Rota, Sp, or Hib)
  + **coverage\_given** = Coverage type (default coverage or best coverage)
  + **dist\_type** = Age distribution type is set to “A1” but “A2” is also an option
  + **psa** – select TRUE to run PSA or FALSE for deterministic run of the model
  + **waning** – can only set to T (TRUE) for Rota
  + **lines 99-102 of code –** if psa is set to T, the model runs 200 times per country, otherwise is psa is set to F the model runs 1 time per country

**univac\_log file:**

To view the progress of the runs, then go to terminal to the folder which contains the model UNIVAC\_161120.R and type in the terminal:

“tail -f univac\_log”

1. **Running PSA scenarios in parallel on the LSHTM cluster:**

* If you wish to run the model on a cluster then it is better to save different files of the model based on the scenario’s you wish to run so that scenario’s are completed faster by running them parallel to one another i.e. if you wish to run Rota for default coverage and Rota for best coverage then it is best to save the code separately for those scenario’s i.e “psa\_Rota\_default.r” and “psa\_Rota\_best.r”
* There are also shell scripts which I will add to the folder. For example if you wish to run rota default scenario on the cluster then you would need to run “Univac\_rota\_default.sh” on the cluster 🡪 this script then in turn calls “psa\_Rota\_default.r”
* On the cluster you would need to move the files and folders on the “cloud” in the same pathway as they are set currently
* The R packages would need to be installed on the cluster and “mypackage” would also need to be compiled again using the tar.gz file that is already saved.

1. **Output Saved**

* Outputs for each scenario are saved under the “Output” folder
* In the Output folder the following 3 folders need to be made for results to be saved:
  + VIMC\_Rota\_Results
  + VIMC\_Sp\_Results
  + VIMC\_Hib\_Results
* If the model is running a deterministic scenario, then the model will automatically create a *“deter”* folder if it doesn’t exist to save the results and save the outcomes by date and vaccination scenario (“no\_vax”, “default\_routine”, “best\_routine”).
* If the model is running a psa scenario, then a *“PSA”* folder will be created if it doesn’t already exist and create the relevant folders based on date, vaccine scenario within.
* Additional output files when running PSA scenario include:
  + *“…InputParams.csv”* – shows the input values for each parameter across all 200 probabilistic runs for each country i.e. the rates (for cases, visits, hospital, deaths), population only for the year 2000.
  + *“…RandomNumbers.csv”* – This output shows the random unique probability used to calculate the input parameter value from a random position between the low-mid-high ranges that exist. This probability is taken from the random number list that is unique for each seed (pathogen) and each input parameter has a unique value for every run of the PSA analysis for every country.

If there is any confusion, then feel free to contact me on: [hira.tanvir@lshtm.ac.uk](mailto:hira.tanvir@lshtm.ac.uk) or [hira.tanvir333@googlemail.com](mailto:hira.tanvir333@googlemail.com)