

Readme - Rubella model overview

Summary

This repository holds the code and input files used for the rubella model runs for the [Vaccine Impact Modelling Consortium](#) (VIMC) during 2023-24 (referred to as the "UKHSA model" or, previously the "PHE model"). It also includes examples of output files. The files are as follows:

- Code files: wraprubella2324.zip and rubelladynamics2324.zip.
- Input files: inputfiles1.zip, inputfiles2.zip, inputfiles3.zip and foi_bootstraps.zip.
- Example output files: Example output.zip.

The model code authors are Emilia Vynnycky, with contributions from Timos Papadopoulos and Kostas Angelis over the past 10+ years.

Contents

	Page no
1 Folder structure	2
2 Description of the files.....	2
2.1 Code files and compilation	2
2.2 Input files	3
2.3 Foi_bootstrap files.....	4
3 Running the model	5
3.1 Option 1 - Running the model using the wrapping program	5
3.2 Option 2 - Running the rubella model without using the wrapping program	6
4 Output files	7

1 Folder structure

When unzipped, the folder structure needs to be as follows, where Model is the name of the folder in which you plan to run the code.

```
Model\ -  
  
Model\inputfiles  
  
Model\foi_bootstraps  
  
Model\output\detailed  
  
Model\output\summary  
  
Model\output\imperial
```

The code needs to go into the “Model” folder. The files in inputfiles1.zip, inputfiles2.zip and inputfiles3.zip need to go into Model\inputfiles; the files in foi_bootstraps need to go into Model\foi_bootstraps. You can use a different name from “Model” for the main folder, but the subfolders need to be given the names above for the model to run. To comply with sharing restrictions, we have not included several files which would have been in the foi_bootstraps folder when the model was run for VIMC and once compiled, the program will not run properly for the countries which need to access the files.

2 Description of the files

2.1 Code files and compilation

The code is written in C/C++. There are two zip files containing code. The file wraprubella2324.zip contains code for a wrapping program, which is intended to simplify running the rubella model for many different countries and repetitions. The file rubelladynamics2324.zip contains the code for producing the rubella dynamics and generating the output required by VIMC.

The files in the zip files contain the following files

wraprubella2324.zip:

- wraprubella2324.c
- unrutil.c

rubelladynamics2324.zip:

- main_rubella_2324.c
- rubella_readin2324.c
- rubella_tynamics2324.c
- rubella_sendout2324.c
- unrutil.c

For convenience, the file unrutil.c in both zip files is identical and comprises code relating to allocating and freeing up memory.

The two files in wraprubella2324.zip need to be compiled together to generate the corresponding executable. For example, if you're using g++, the wrapping code would be compiled as follows to produce an executable called wraprubella2324:

```
g++ -o wraprubella2324 wraprubella2324.c unrutil.c
```

The files in the rubelladynamics2324.zip contain code relating to the main running of the program (main_rubella2324.c), reading in values from input files (rubella_readin2324.c), updating the size of compartments over time (rubella_todynamics2324.c), sending output to files (rubella_sendout2324.c) and allocating and freeing up and memory (unrutil.c). The rubella dynamics code would be compiled as follows to produce an executable called main_rubella2324:

```
g++ -o main_rubella2324 main_rubella_2324.c rubella_readin2324.c
rubella_todynamics2324.c rubella_sendout2324.c unrutil.c
```

2.2 Input files

The input files folder contains the files described below. In general, comments in the file are indicated by an asterisk at the start of the line and the end of the file is indicated by "!" (demography and vaccination coverage files) or by "Z" (BS summary file).

1. Demography files

- a) The files with the name imp*.inp contain the demographic inputs provided by VIMC, which have been formatted to allow them to be read by the model and modified slightly to include the UN country code for each country. The files are
 impaspfert.inp, imppopmales2023.inp, imppopfem2023, impcbrate2023.inp,
 impsurvprop_females_2023, impsurvprop_males_2023.inp, implexpectancy2023.inp.
- b) The file demog_summary2024.inp summarises the names of the demography input files used in the model.

2. Vaccination files

- a) The files with the name vimc*rcv1.inp, vimc*rcv2.inp and vimc*sia.inp contain the RCV1, RCV2 and SIA vaccination coverage respectively for each of the countries and vaccination scenarios, provided by VIMC. The "*" in the name reflects the assumed vaccination scenario (BS=blue skies, DF=default, NV=no vaccination, IA= IA2030). The files have been modified slightly from those provided by VIMC to include the UN country code for each country in the first column.
- b) The file vcov_summary2024.inp summarises the names of the vaccination files used in the model for each of the scenarios. Each scenario is allocated a number in this file. This number is entered in the command line and the wrapper code uses this number to provide the corresponding files to the model, based on the information in the file vcov_summary2024.inp. Further details for running the code are in section 3.

3. BS ("bootstrap") summary file

The file used for VIMC runs for 2023-24 is called bsdalcc2024.inp and provides details about each of the countries that are used when running the model, specifically:

- a) Country_num – each of the countries is assigned a unique identifier between 1 and 117 (the number of countries for which output was required by VIMC). This identifier is used in the command line to specify the country for which output is being generated (see section 3).
- b) Country_name – the country name.
- c) Country_abbrev – the 3 letter country abbreviation.
- d) UN_Country_code – this is the numeric UN country code, which is used to extract the appropriate demographic inputs.
- e) DALY – this is the DALY (Disability-Adjusted Life Year) associated with CRS for the given country.
- f) Source_filename_for_foi_reps - The name of the file containing the equilibrium (pre-vaccination) force of infection for each of the repetitions between 1 and 200. These have the extension *.mbio. These files are in the folder foi_bootstraps.
- g) Source_filename_for_foi_pointest - The name of the file containing the equilibrium (pre-vaccination) force of infection if output is just required for a single point estimate (average). These have the extension *.med or *.ave.

4. Files for specifying the parameters being varied

When the model is run for repetition n, unless otherwise stated, the program extracts the corresponding parameter values from the nth line in the following files:

bootstrap - the nth line in this file (excluding the comments, which are indicated by an *) specifies the line from which the force of infection in the *.mbio file should be taken.

dis.mort – this specifies the CRS-related mortality rate.

risk.crs – this specifies the risk of a child being born with CRS if infection in the mother occurs during the first 16 weeks of pregnancy.

scale.factor - this specifies the factor by which the vaccination coverage should vary from that provided (accounting for chance variation). In the VIMC model runs for 2023-24, this was hard-coded to equal 1.0, so this file was not used.

vacc.eff – this specifies the rubella vaccine efficacy.

2.3 Foi_bootstrap files

The folder foi_bootstraps contains the files which are used to determine the equilibrium pre-vaccination force of infection for each of the countries and repetitions. The values were generated by fitting catalytic models to seroprevalence datasets. Further details about the fitting process can be found in the following references:

Vynnycky E, Knapp JK, Papadopoulos T, Cutts FT, Hachiya M, Miyano S, Reef SE. Estimates of the global burden of Congenital Rubella Syndrome, 1996-2019. *Int J Infect Dis*. 2023 Dec;137:149-156.

Vynnycky E, Adams EJ, Cutts FT, Reef SE, Navar AM, Simons E, Yoshida LM, Brown DW, Jackson C, Strebel PM, Dabagh AJ. Using Seroprevalence and Immunisation Coverage Data to Estimate the Global Burden of Congenital Rubella Syndrome, 1996-2010: A Systematic Review. *PLoS One*. 2016 Mar 10;11(3):e0149160.

The files have been used in other applications (beyond those for VIMC) and only the columns relating to the force of infection (columns labelled `Foi_p_T_sus_young` and `Foi_p_T_sus_old`) are extracted during the model runs for VIMC; the other columns can be ignored.

3 Running the model

3.1 Option 1 - Running the model using the wrapping program

In a Windows environment, the easiest way to run the program is to open up a Command prompt and use the compiled wrapping program from the command line, ensuring that you are in the same folder that the program is located. The program is hard-coded to use the name `main_rubella2324.exe` for the rubella dynamics model (the executable produced after compiling contents of `rubelladynamics2324.zip`). If the name of the rubella dynamics model needs to be different (e.g. because it's being run in a different environment), the wrapping code would need to be recompiled after replacing `main_rubella2324` with the new name of the model in the code.

The wrapping program requires input on the following on the same line:

- a) `start_country_num` – the first country for which output is required, which has to be in the range 1-117.
- b) `fin_country_num` – the last country for which output is required, which has to be higher than `start_country_num` and less than 117.
- c) `first_rep` – the first repetition number required, which has to be in the range 0-200
- d) `final_rep` – the final repetition number required, which has to be higher than `first_rep` and less than or equal to 200.
- e) `scenario_num` – the vaccine scenario number. Possible values are 101, 102, 103, 201, 202, 203, 301, 302, 303, 401, as specified in the vaccine summary file.
- f) `bssumfile` – the file containing the epidemiological parameters for each of the countries. For the VIMC runs, this is `bsdalcc2024.inp`.
- g) `vcov_summary_file` – the name of the file containing the names of the vaccination coverage files for each scenario.
- h) `demog_summary` – the name of the file containing the names of the demography files used in the runs.

To run the wrapping program, you would type the following from the command line, replacing each of the arguments after the program name with the corresponding argument:

```
wraprubella2324 start_country_num fin_country_num first_rep final_rep
scenario_num bssumfile vcov_summary_file demog_summary
```

For example, to use the wrapping program to run the main rubella program for scenario 201 for countries 1 and 2 for repetitions 3-4, you would type the following at the command line:

```
wraprubella2324 1 2 3 4 201 bsdalcc2024.inp vcov_summary2024.inp
demog_summary2024.inp
```

3.2 Option 2 - Running the rubella model without using the wrapping program

In a Windows environment, the rubella dynamics model can be run from the command line without using the wrapping program, after ensuring that you are in the same folder that the program is located. The program needs the values for the following inputs on the command line, which should all be on the same line:

- a) cnum – this is the country number for the country for which output is required, which has to be in the range 1-117.
- b) fin_year – this is the final year for the model runs, which has to be 2101.
- c) scenario_num - – the vaccine scenario number. Possible values are 101, 102, 103, 201, 202, 203, 301, 302, 303, 401, as specified in the vaccine summary file.
- d) first_rep – the first repetition number required, which has to be in the range 0-200.
- e) final_rep – the final repetition number required, which has to be higher than first_rep and less than or equal to 200.
- f) sia_covfile – the name of the file containing SIA coverage.
- g) rcv1_covfile – the name of the file containing RCV1 coverage.
- h) rcv2_covfile - the name of the file containing RCV2 coverage.
- i) bssumfile - – the file containing the epidemiological parameters for each of the countries. For the VIMC runs, this is bsdalcc2024.inp.
- j) pvacc_again_RCV1RCV2 – proportion of RCV2 recipients who have received RCV1. This equalled 1.0 for VIMC runs.
- k) frac_import – importation fraction. This equalled 0.0 in VIMC runs.
- l) brate_file – name of the file containing the birth rate.
- m) male_surv_file – name of the file containing the survival data for males.
- n) fem_surv_file – name of the file containing the survival data for females.
- o) male_popfile – name of the file containing the age-specific population for males.
- p) fem_popfile – name of the file containing the age-specific population for females.
- q) lexpectancy_file - name of the file containing the life expectancy for females.
- r) agespec_fert_file – name of the file containing the age-specific fertility rates.

To run the main rubella program for scenario 201 for country 3 for repetitions 5-7, you would type the following at the command line:

```
main_rubella2324 3 2101 201 5 7 vimcDFsia.inp vimcNVrcv1.inp vimcNVrcv2.inp
bsdalcc2024.inp 1.0 0.0 impcbrate2023.inp impsurvprop_males_2023.inp
impsurvprop_females_2023.inp impopmales2023.inp impopfem2023.inp
implexpectancy2023.inp impaspfert2023.inp
```

Whilst running, the model sends output to the screen, which can be used to keep track of whether the model is running successfully. Once the message `Program running...starting v_ind=0` appears as shown below, the model has read in all the inputs that it needs from the input files and is starting to run the rubella dynamics.

```
Program running...starting v_ind=0  
year=1800  
year=1801  
year=1802  
year=1803  
year=1804  
year=1805  
year=1806  
year=1807  
year=1808  
year=1809  
year=1810  
year=1811  
year=1812  
year=1813  
year=1814  
year=1815  
year=1816  
year=1817  
year=1818  
year=1819  
year=1820  
year=1821  
year=1822  
year=1823
```

If the above output has not appeared soon (usually within 30 seconds) of starting to run the model, it is likely that an error has occurred. If the folder structure is correct (see section 1), the error may lie with the input file. In this case, the file leading to the error can often be identified from the error message that the model generates.

The model stops running the current repetition once it has reached year=2100, as follows:

```
year=2074  
year=2075  
year=2076  
year=2077  
year=2078  
year=2079  
year=2080  
year=2081  
year=2082  
year=2083  
year=2084  
year=2085  
year=2086  
year=2087  
year=2088  
year=2089  
year=2090  
year=2091  
year=2092  
year=2093  
year=2094  
year=2095  
year=2096  
year=2097  
year=2098  
year=2099  
year=2100  
year=2101
```

For each model run, the model generates several output files, which are described in the next section.

4 Output files

For each repetition, the model generates an output file in each of the folders output\detailed, output\imperial and output\summary. The extension of the files in these folders is determined by the scenario, replicate, bootstrap number and country numbers. The files in the zip file Example output.zip

contain the output files that are produced after running the code using the commands on pages 5 and 6.

The files in the imperial subfolder contains the age-specific output that is provided for VIMC, after the line starting "Created on,," is deleted and the files are concatenated across all repetitions and countries. The contents of the columns in the file for runs for repetitions 1-200 are as follows:

Column 1: Antigen name.

Column 2: Repetition number.

Column 3: Year.

Column 4: Age (years).

Column 5: 3 letter country code.

Column 6: Country name.

Column 7: Population size in the year and age group in columns 3 and 4.

Column 8: Number of DALYs associated with CRS cases born to mothers in the year and age group in columns 3 and 4 respectively.

Column 9: Number of CRS cases born to mothers in the year and age group in columns 3 and 4 respectively.

Column 10: Number of CRS deaths among children born to mothers in the year and age group in columns 3 and 4 respectively.

Column 11: Number of infections among people in the year and age group in columns 3 and 4 respectively.

Column 12: Number of years of life lost among CRS cases born to mothers in the year and age group in columns 3 and 4 respectively.

The columns for repetition 0 (which is run when the point estimate is required), are the same, except that there are only 11 columns and the column for the repetition number is omitted.

The files in the summary subfolder have annual output aggregated across all ages, which can be used for checking output between successive runs or, after reformatting, for plotting. The output for a given run is on a single line and the name of the variable is provided in the column before the output for that variable for the first year starts. The output for each of the columns is as follows:

Columns 1-24 – Date of the run, input parameters and names of files used in the run.

Columns 26-146 Numbers of individuals vaccinated in each year during 1980-2100.

Columns 148-253 Annual number of CRS cases, 1995-2100.

Columns 255-360 Annual number of infns, 1995-2100.

Columns 362-467 Annual population size, 1995-2100

Columns 469-574 Age-standardised number of infections per 100,000 population, 1995-2100 (standardised to the population in 2010).

Columns 576-681	Total DALYs, 1995-2100.
Columns 683-788	Number of CRS cases per 100,000 livebirths, 1995-2100.

The contents of the detailed folder can be ignored or used for cross-checking. During the run, the code generates a file called `crssum.out`, which can be ignored or used for cross-checking. Both of these files just contain the parameters and names of the input files used in the run.

If the model is run using the wrapping program, the code also generates a file called `test*.out` which contains output that is sent to the screen during the run, which can be used for diagnostic purposes if the model fails to run.

If the output folders already contain output from the repetition that is currently being run, the model appends the output to the end of the existing file.