COVID-19 Modelling

PHAC age-stratified SEIR compartmental model: User Notes for Version 15

Overview of steps required to run the model

1	Introduction to the PHAC age-stratified SEIR compartmental model	2
2	Downloading the required software	3
3	Walking through the example workbooks	3
3.1	Documentation worksheet	4
3.2	Model specification (OOS)	5
3.2.1	"Model specification" sheet (OOS)	5
3.2.2	"Intermediate calculations" sheet (OOS)	6
3.3	Parameters **	6
3.3.1	"Parameters any time any age" sheet	6
3.3.2	"Parameters any age" sheet	7
3.3.3	"Parameters by age" sheet	7
3.3.4	"Parameters by Age x Age" sheet	8
3.4	"Initial conditions" sheet **	9
3.5	"Post processing" sheet (OOS)	9
3.6	"HyperCube Sampling Specs" sheet (OOS)	
3.6.1	Including parameter variability	9
4	Running the model	11
5	Appendix: Tables of compartments, parameters and list of outputs	14
	Table 1. Compartment definitions	14
	Table 2. Parameter definitions	15
	System and versions of R, EpiSim and other packages last used with the examples in this guide	17

1 Introduction to the PHAC age-stratified SEIR compartmental model

This user guide presents an overview of the steps required to use the PHAC age-stratified COVID-19 SEIR model. The model is an updated version (v15) of the age-stratified dynamic deterministic compartmental model using a susceptible, exposed, infected, removed (SEIR) framework applied to the Canadian population stratified into six age groups, presented in Ludwig et al. 2020. It uses ordinary differential equations with transition rates as the inverse of time in reservoirs to simulate changes in infectious and disease states of individuals in the population.

Performing simulations with the model is executed in R but the model structure, input values, some intermediate calculations and many post-processing steps are defined in an Excel worksheet. The objective of this document is to provide the reader with the information required to: (1) run the model in its present form; and, (2) to modify the parameter values to perform simulations of COVID-19 transmission according to the population characteristics they may want to simulate. Although the epidemiologic structure of the model, the list of outputs and many post-processing steps can be modified by the user, an explanation of those changes is beyond the scope of this user guide. The post-processing step, in particular requires a more thorough understanding of the EpiSim package structure than can be provide here. Sections in this user guide that correspond to what a user may be expected to change are marked with ** while the others are followed by OOS (out of scope). Explanations provided for modifications deemed OOS will be brief but are provided to assist the user to understand the general mechanics of the model.

The model structure, including transmission equations and intermediate calculations, is defined within an Excel workbook. The workbook containing the PHAC compartmental model can be downloaded **here**. The parameter inputs values should be provided by the user although to help the user with this task, two examples are provided.

The first example allows for a basic epidemic simulation. It uses the "Example1(single parameter set). xls" workbook and "Example1(single parameter set).r" R code. The required files can be downloaded **here**.

The second example allows the user to perform multiple simulations while varying values for some parameters. It uses the "Example2(multiple parameter sets).xls" and "Example2(multiple parameter sets).r" R code. The required files can be downloaded **here**.

2 Downloading the required software

Running the model requires the use of the EpiSim package, which can be downloaded at https://github.com/statcan/EpiSim. The model is run in R, a language and environment for statistical computing and graphics commonly used in many scientific fields. R is available as Free Software and can be downloaded from https://www.r-project.org/.

The use of RStudio, another Free Software, in addition to R is recommended but not obligatory. It provides a better interface and helps with the generation of R markdown document among other benefits.

Once the user has R or RStudio installed on his computer, the following R command should be run to install package EpiSim.

```
devtools::install_github("https://github.com/statcan/EpiSim", upgrade = "never")
```

This installation requires the R package devtools. The user will need to install it if it is not already on so.

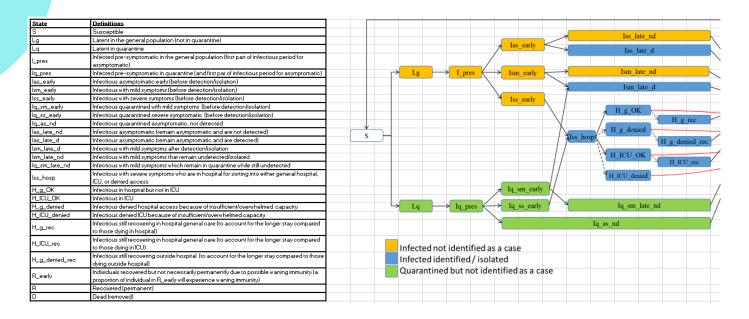
Information on versions of R and packages that were used to produce the examples in this guide are provide in the Appendix. If the user is having difficulties running the examples, it could be because an improper version of some packages is used.

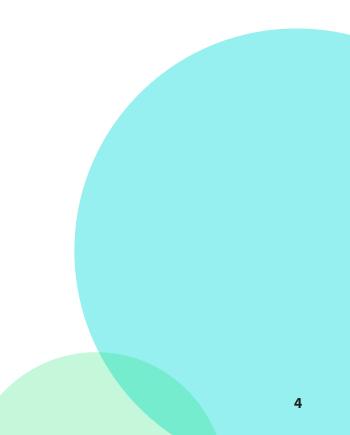
3 Walking through the example workbooks

Modifications to the Excel sheets in the Excel workbook may be required depending on the simulation objectives. How to do this from scratch is out of scope (OOS). This brief guide and examples focus on using the model "as is", with the possibility of changing parameter values. Changes to the model structure are OOS of the current document.

3.1 Documentation worksheet

This Excel sheet present the conceptual structure of the model and is for documentation purposes only and, as such, is not read by the R code. The flow diagram of the model is presented along with the names and definitions of the compartments and parameters.





3.2 Model specification (OOS)

3.2.1 "Model specification" sheet (OOS)

The model uses ordinary differential equations with transition rates as the inverse of time in reservoirs. In this sheet, all of the equations corresponding to the differential equations are written in a systematic way using the "From", "To" and "expression" columns.

The "activation" and "lazy" columns are a mandatory requirement of the EpiSim package and refer to additional functionalities which are OOS for this user guide: these do not require modification for the examples presented.

The "expression" column involves expressions that are a function of the compartments (e.g., Lg) and parameters (e.g., sigma) as well as any variable (e.g., tot_pop_Dexcluded_by_age) constructed in the "Intermediate calculations" sheet briefly discussed later.

Modifications to this sheet have implications for subsequent dependent sheets in the workbook and are OOS for this user guide.

A	B	С	D	E	F
Rate	From	То	lazy	activation	expression
Susceptible to Latent (S to LXXX)					
	S	Lg	0) 1	(S*beta/ tot_pop_Dexcluded_by_age)*(l-lambda*delta)* contact_chur
	S	Lq	0) 1	(S*beta/ tot_pop_Dexcluded_by_age)* lambda *delta * contact_chu
Latent to Ig pres or Ig pres					
	Lg	I_pres	0	1	Lg / sigma
	Lq	lq_pres	0	1	Lq / sigma
Transitions out of Ig_pres or Iq_pres					
manations out on ig_pres on ig_pres	I pres	las_early	0	1	I pres*(tau) /t pres
	l pres	Ism early	0		I pres* (1-tau) * (1-alpha) /t pres
	I_pres	lss_early	0	1	I_pres* (l-tau)* alpha /t_pres
	lq pres	lq_sm_early	0	1 4	Iq_pres*(1-tau) * (1-alpha) /t_pres
	Iq pres	Iq ss early	0		Iq pres* alpha *(1-tau) /t pres
	Iq_pres	lq_as_nd	0		Iq pres* tau /t pres
		- <u>-</u>			
Transition out of las_early					
	las_early	las_late_nd	0) 1	Ias_early* (l-delta) / tsm_early
	las_early	las_late_d	0	1	Ias_early* (delta) / tsm_early
Transition out of Ism_early					
	Ism_early	lsm_late_nd	0	1	Ism early * (l-delta) / tsm early
	Ism_early	lsm_late_d	0	1	Ism early * (delta) / tsm early
Transition out of lss early					
	Iss early	lss hosp	0	1	Iss early / tss early

3.2.2 "Intermediate calculations" sheet (OOS)

This sheet includes intermediate expressions that are more practical to formulate outside of the compartments and parameters defined in the "Model specification" sheet. It is possible to construct additional variables in this sheet that can be made available for expressions included in the "Model specification" sheet. Such additional variables will be functions of compartments and parameters.

show.plotperiod.cutoff.1period.cutoff.213060

3.3 Parameters **

The model involves parameters that are briefly described in the Documentation sheet (see section 3.1) and in the Appendix of this user guide. The parameters may depend on age groups as well as time periods. Age groups are numbered from 1 to 6 (refer to appendix for definition). Time periods are specified through the use of two variables, tmin and tmax, in the last three sheets listed in Table 1 below. The simulation will run until the largest tmax value.

Parameter dimension(s)	Sheet in workbook
None	Susceptible
Time era	Latent in the general population (not in quarantine)
Time era x Age group	Latent in quarantine
Time era x Age group x Age group	Infected pre-symptomatic in the general population (first part of
	infectious period for asymptomatic)

3.3.1 "Parameters any time any age" sheet

This sheet is used to define parameters which are neither influenced by time nor age group.

1	Α	В	С	D	E	F	G	Н		J	k
1	beta	lambda	delta	tsm_early	sigma	tau	t_pres	tsm	tss_early	t_sorting	th_I
2	0.041	0.4	0.4	2	4.12	0.2	2	9	1	1	1
3											

3.3.2 "Parameters any age" sheet

Baseline values for parameters which do not vary between age groups but could vary by time period are defined in this sheet.

To illustrate this, in the two examples provided, the parameter "Cgg_multiplier" take a different value at different times in the simulations but is the same for all age groups. This parameter is a coefficient that can be used to modify the value of the parameter "Cgg" by a certain amount to simulate control measures. It is a multiplier of the contact rates within and between age groups.

	Α	В	С		
1	tmin	tmax	Cgg_multiplier		
2	0	30		1	
3	30	60		0.5	
4	60	730		0.8	
5					

3.3.3 "Parameters by age" sheet

Parameters stratified by age groups can be modified in the "Parameters by age" sheet. For example, the user may implement different mortality rates for different age groups. A common value for all age groups can be set if age-specific data are not yet available. There is one row for each of the six age groups and each time period. If multiple periods are specified, a value must be entered for each period (eg. See example below including three periods).

	А	в	С	D	E	
1	tmin	tmax	agegrp	alpha	m_g	
2	0	30	1	0.0086	0	
3	30	60	1	0.0086	0	
4	60	730	1	0.0086	0	
5	0	30	2	0.0091	0.0076	
6	30	60	2	0.0091	0.0076	
7	60	730	2	0.0091	0.0076	
8	0	30	3	0.0179	0.015	
9	30	60	3	0.0179	0.015	
10	60	730	3	0.0179	0.015	
11	0	30	4	0.0086	0.055	
12	30	60	4	0.0086	0.055	
13	60	730	4	0.0086	0.055	
14	0	30	5	0.0086	0.18	

3.3.4 "Parameters by Age x Age" sheet

The purpose of this sheet is to establish the contact level between individuals, both within and between age groups.

Values must be entered for every combination of age groups because contact levels are not considered symmetric. In the current model, the population is stratified into six age groups: 0-9 years, 10-19 years, 20-39 years, 40-59 years, 60-74 years, and 75 years and older. Default contact values presented in the workbook are derived from Prem et al. 2017, but the user should input the values that best represent the contact matrix for the area modelled.

A long and skinny format is used in this sheet. For example, ragegrp=2 and cagegrp =1 refer to the 2nd row and 1st column of the parameters of interest (i.e., Cgg_early or Cgq or beta_mat) which are 6x6 matrices. Note that values for ragegrp=2 and cagegrp =1 are not the same as those for ragegrp=1 and cagegrp =2, which speaks to the earlier comment that Cgg_early and Cgq are not symmetrical matrices. If multiple periods are specified, a value must be entered for each period.

It is necessary to provide this information separately for contacts between individuals in the general population and individuals in quarantine at home.

Ref: Prem K, Cook ARA, Jit M. Projecting social contact matrices in 152 countries using contact surveys and demographic data. PLoS Comput Biol. 2017;13. doi:10.1371/journal.pcbi.1005697

	А	В	С	D	E	F
1	tmin	tmax	ragegrp	cagegrp	Cgg_early	Cgq
2	0	30	1	1	4.599534812	0.467799287
3	0	30	1	2	0.890526663	0.090571711
4	0	30	1	3	2.594102626	0.263835237
5	0	30	1	4	1.375859514	0.139932868
6	0	30	1	5	0.335597055	0.034132161
7	0	30	1	6	0.036661986	0.003728736
8	0	30	2	1	1.025782623	0.061104424
9	0	30	2	2	10.26981369	0.61175832
10	0	30	2	3	2.795061477	0.166497871
11	0	30	2	4	2.453452343	0.146148696
12	0	30	2	5	0.209460263	0.012477253
13	0	30	2	6	0.033800288	0.002013436
14	0	30	3	1	1.152654388	0.074715133
15	0	30	3	2	1.667010138	0.108055707
40	•	20	2	2	0.475000650	0 50005 6501

3.4 "Initial conditions" sheet **

Initial values must be provided by the user for all compartments and for all age groups (i.e., one column per age group). In the example provided, the values indicated are used to study percolation between age groups as a theoretical experiment: the user will want to provide initial conditions corresponding to the simulation of interest.

	А	В	С	D	Е	F	G
1	NAME	agegrp1	agegrp2	agegrp3	agegrp4	agegrp5	ageg
2	S	3982527	4146397	10286131	10069708	6315255	2789
3	Lg	200	0	0	0	0	
4	Lq	0	0	0	0	0	
5	I_pres	100	0	0	0	0	
6	lq_pres	0	0	0	0	0	
7	las_early	150	0	0	0	0	
8	Ism_early	150	0	0	0	0	
9	lss_early	50	0	0	0	0	
10	lq_sm_early	0	0	0	0	0	
11	lq_ss_early	0	0	0	0	0	
12	lq_as_nd	0	0	0	0	0	
13	las late nd	100	0	0	0	0	

3.5 "Post processing" sheet (OOS)

This sheet defines post processing work, including the output for the model. The content of this sheet is intertwined with multiple elements of the EpiSim package. The model currently outputs values for all model compartments and additional parameters of interest and modifying this is OOS for this user guide.

3.6 "HyperCube Sampling Specs" sheet (OOS

3.6.1 Including parameter variability

It is possible for the user to set up multiple simulations and have the values for some parameters drawn from a distribution. This can be used, for example, to assess the sensitivity of the modelling results to uncertainty in the values of some parameters.

When using this functionality, values can be drawn from either a triangular or a uniform distribution and are attributed for the entire duration of the simulation run.

Using this functionality requires the user to add information to the "HyperCube Sampling Specs" sheet.

	Α	В	С	D	E	F
1	hypercube	parameter.name	lower.bound	upper.bound	apex	
2	1	delta	0.2	1	0.4	
3						

In order to use this functionality, a call to the relevant functions must be made in the main R code. The code for "Example2(multiple parameter sets)" allows for the use of hypercube sampling and does require minimal information from the user. The inputs required are:

- The line beginning with "hypercube.apex.mode" can be commented (add # before the line) if the user prefers a uniform distribution
- The line beginning with "n.repeat.within.hypercube" indicates the number of repeat (a new set of parameter values will be drawn for each repetition)
- The line beginning with "racine" require a seed number for the pseudo random process
- Lines beginning with "write.csv" require that names must be provided for "csv" output files

Note that, if values are provided in the hypercube sampling sheet, the model will run as if parameter values of the workbook were overwritten (even though the workbook itself is not modified).

Modifications to the others elements of this code section are OOS for this user guide.

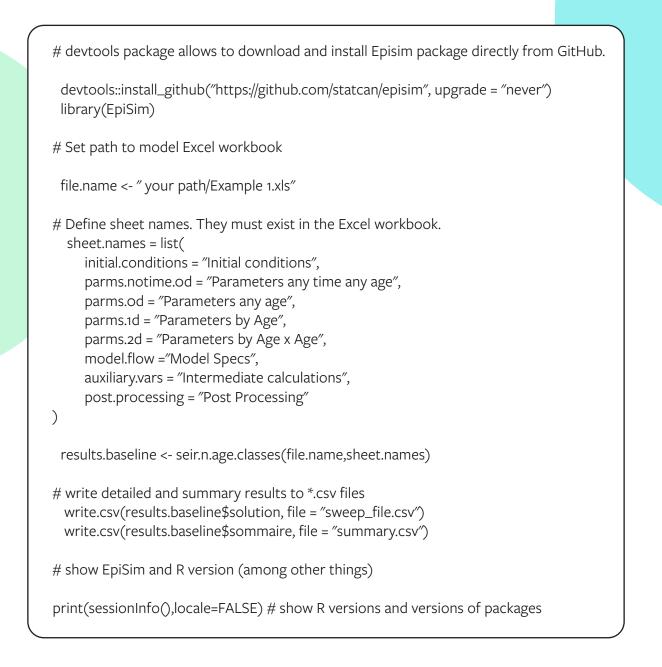


4 Including parameter variability

Two examples of the model simulation in R code are included to demonstrate basic model mechanics and explore the use of multiple parameter values sampled from distributions.

The R code required to use the Example1(single parameter set).xls is presented in Box.1. The R code required to use the Example2(multiple parameter sets).xls. share the same initial sections but necessary modifications are presented in Box.2.

Box 1. Basic simulation approach, for Example1(single parameter set).xls. The code below also found in the R code file "Example1(single parameter set).r".



Starting with the code presented in Box 1, elements presented in Box 2 replace all elements after the line "results.baseline <- seir.n.age.classes(file.name,sheet.names)"

Box 2. Basic simulation approach with added parameter sweeps for Example2(multiple parameter sets).xls. The code below is also found in bottom part of the R code file "Example2(multiple parameter sets).r".

hypercube.specs <- read.hypercube.sampling.specs(file.name, sheet = "HyperCube Sampling Specs")

```
parm.cloud.grid.specs <- list(
  hypercube.lower.bounds = hypercube.specs$lower.bound,
  hypercube.upper.bounds = hypercube.specs$upper.bound,
  hypercube.apex.mode = hypercube.specs$apex , # To be commented if uniform
  n.repeat.within.hypercube = 2, # 10000
  LatinHypercubeSampling = c(FALSE, TRUE)[2],
  racine = 98 , #this is a seed for the pseudo random process
  backend.transformation = function(x) {x},
  reference.alteration = c("overwrite", "add", "multiply")[1],
  tmin.alter.scope = 0:1000
```

```
)
```

various.parms.result = try.various.parms.values(results.baseline, parm.cloud.grid.specs)
various.parms.result\$df.sweep\$which.sim = cumsum(various.parms.result\$df.sweep\$time ==0)

```
# write sweep file and outcome summary to csv
write.csv(various.parms.result$df.sweep, file = "NAME.csv")
write.csv(various.parms.result$outcomes.summary.df, file = "NAME.csv")
```

```
# show EpiSim and R version (among other things)
```

print(sessionInfo(),locale=FALSE) # show R versions and versions of packages

Although the model produces many results, their complete description is OOS for this guide. The following outputs can be used directly by the user:

AR: Attack rate (in summary.csv)

Incl: Incidence of infected (in sweep_file.csv)

cuml: Cumulative incidence of infected (in sweep_file.csv)

cumIconfirmed: Cumulative incidence of identified/isolated cases (in sweep_file.csv)

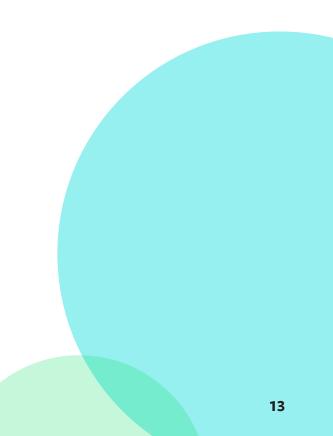
The results can be exported to csv format as illustrated below.

For Example1(single parameter set).

```
28 # write sweep file and outcome summary to csv
29 write.csv(results.baseline$solution, file = "sweep_file.csv")
30 write.csv(results.baseline$sommaire, file = "summary.csv")
31
```

For Example2(multiple parameter sets).

```
49 # write sweep file and outcome summary to csv
50 write.csv(various.parms.result$df.sweep, file = "sweep_file.csv")
51 write.csv(various.parms.result$outcomes.summary.df, file = "summary.csv")
52
```



5 Appendix: Tables of compartments, parameters and list of outputs

Table 1. Compartment Definition

State	Definitions
S	Susceptible
Lg	Latent in the general population (not in quarantine)
Lq	Latent in quarantine
l_pres	Infected pre-symptomatic in the general population (first part of infectious period for asymptomatic)
lq_pres	Infected pre-symptomatic in quarantine (and first part of infectious period for asymptomatic)
las_early	Infectious asymptomatic early (before detection/isolation)
Ism_early	Infectious with mild symptoms (before detection/isolation)
Iss_early	Infectious with severe symptoms (before detection/isolation)
Iq_sm_early	Infectious quarantined with mild symptoms (before detection/isolation)
Iq_ss_early	Infectious quarantined with severe symptoms (before detection/isolation)
lq_as_nd	Infectious quarantined asymptomatic, not detected
las_late_nd	Infectious asymptomatic (remain asymptomatic and are not detected)
las_late_d	Infectious asymptomatic (remain asymptomatic and are detected)
Ism_late_d	Infectious with mild symptoms after detection/isolation
Ism_late_nd	Infectious with mild symptoms that remain undetected/isolated
lq_sm_late_nd	Infectious with mild symptoms which remain in quarantine while still undetected
lss_hosp	Infectious with severe symptoms who are in hospital for sorting into either general hospital, ICU, or denied access
H_g_OK	Infectious in hospital but not in ICU
H_ICU_OK	Infectious in hospital in ICU
H_g_denied	Infectious denied hospital access because of insufficient/overwhelmed capacity
H_ICU_denied	Infectious in hospital, requiring ICU but denied ICU access because of insufficient/ overwhelmed capacity
H_g_rec	Infectious still recovering in hospital general care (to account for the longer stay compared to those dying in hospital)
H_ICU_rec	Infectious still recovering in hospital ICU (to account for the longer stay compared to those dying in ICU)
H_g_denied_ rec	Infectious still recovering outside hospital (to account for the longer stay compared to those dying outside hospital)
R_early	Individuals recovered but not necessarily permanently due to possible waning immunity (a proportion of individual in R_early will experience waning immunity)
R	Recovered (permanent)
D	Dead (removed)

Parameter Name	Definition
Agegr#	Age group number. By default, the following six age groups are defined: 0-9 years, 10-19 years, 20-39 years, 40-59 years, 60-74 years, and 75 years and older
beta	Probability of transmission when contact made with infectious person
lambda	Proportion of exposed to detected infectious who are traced and quarantined (contact tracing/quarantine)
Cgg	Number of daily contacts between an individual present in the general population and another individual also in the general population (can be equal to Cgg_early or modified to account for physical/social distancing and closures at different times over the simulation)
Cgq	Number of daily contacts between an individual in quarantine and another individual in the general population (can be equal to Cgq_early or modified to account for physical/social distancing and closures at different time over the simulation)
Cgg_early	Initial number of daily contacts between two individuals from the general population as derived from Prem et al. 2017 (see ref below table)
sigma	Latent period (days)
delta	Proportion of presymptomatic infectious who will be identified (or detected)
alpha	Proportion of symptomatic infected who develop severe symptoms
tau	Proportion of infectious that are asymptomatic after the duration of the presymptomatic infectious period
t_pres	Period of time between onset of infectiousness and onset of symptoms in those developing symptoms OR first infectious period for asymptomatic
tsm_early	Period of time between onset of symptoms for mild cases or asymptomatic and detection
tss_early	Period of time between onset of symptoms for severe cases or asymptomatic and detection
tsm_late	Period of time between the possibility of being detected and end of infectious period for asymptomatic and mild cases
tsm	Total period of time for the symptomatic period
t_late_q_sm	Period of time between the possibility of being detected and end of quarantine for mild cases
t_late_q_as	Period of time between end of theoretical presymptomatic infectious period and end of quarantine for asymptomatic
p_ICU	Proportion of hospitalized cases who require ICU in hospital
t_sorting	Period of time for sorting severe cases in hospital (before general service or ICU)
m_g	Mortality rate for severe cases in hospital that do not require ICU or a ventilator (general)
m_ICU	Mortality rate for severe cases in hospital ICU
1	

Table 2. Parameter Definitions

Parameter Name	Definition
t_hr_early	Period of time between first day in hospital (after sorting) and death, for dead
	cases.
th_late	Period of time between second period of hospitalization and recovery, for
	recovered cases.
m_g_denied	Mortality rate for severe cases dying at home because they are not able to
	access hospital care
m_ICU	Mortality rate for severe cases dying in hospital because they are not able to
	access ICU
ICU capacity	Maximum ICU capacity
w	Percent of recovered who loose their immunity
t_im	Duration of immunity for recovered

Ref: Prem K, Cook ARA, Jit M. Projecting social contact matrices in 152 countries using contact surveys and demographic data. PLoS Comput Biol. 2017;13. doi:10.1371/journal.pcbi.1005697



System and versions of R, EpiSim and other packages last used with the examples in this guide:

NOTE: if the user is having difficulties running the examples presented in this user guide, it could be because improper version of some packages are being used (especially those that are bolded below)

R version 4.0.2 (2020-06-22) Platform: x86_64-w64-mingw32/x64 (64-bit) Running under: Windows 10 x64 (build 17134)

locale:

LC_COLLATE=English_Canada.1252 LC_CTYPE=English_Canada.1252 LC_MONETARY<mark>=English_</mark> Canada.1252 LC_NUMERIC=C LC_TIME=English_Canada.1252

attached base packages: stats graphics grDevices utils datasets methods base

other attached packages: EpiSim_0.12.7

Packages loaded via a namespace (and not attached):

adaptivetau_2.2-3	ellipsis_0.3.1	memoise_1.1.0	rprojroot_1.3-2
assertthat_0.2.1	fansi_0.4.1	munsell_0.5.0	rstudioapi_0.11
backports_1.1.7	forcats_0.5.0	openxlsx_4.1.5	scales_1.1.1
callr_3.4.4	fs_1.4.2	pillar_1.4.6	sessioninfo_1.1.1
cellranger_1.1.0	generics_0.0.2	pkgbuild_1.1.0	stringi_1.4.6
cli_2.0.2	ggplot2_3.3.2	pkgconfig_2.0.3	testthat_2.3.2
colorspace_1.4-1	glue_1.4.1	pkgload_1.1.0	tibble_3.0.2
compiler_4.0.2	grid_4.0.2	plotly_4.9.2.1	tidyr_1.1.0
crayon_1.3.4	gtable_0.3.0	prettyunits_1.1.1	tidyselect_1.1.0
curl_4.3	htmltools_0.5.0	processx_3.4.3	tools_4.0.2
data.table_1.12.8	htmlwidgets_1.5.1	ps_1.3.3	triangle_0.12
desc_1.2.0	httr_1.4.2	purrr_0.3.4	usethis_1.6.3
deSolve_1.28	jsonlite_1.7.1	R6_2.4.1	vctrs_0.3.1
devtools_2.3.2	lazyeval_0.2.2	Rcpp_1.0.5	viridisLite_0.3.0
digest_0.6.25	lhs_1.0.2	readxl_1.3.1	withr_2.2.0
dplyr_1.0.0	lifecycle_0.2.0	remotes_2.2.0	zip_2.0.4
DT_0.15	magrittr_1.5	rlang_0.4.7	



