# Set up and run the Simple HIV model

All the project files are stored in the main directory and 3 main sub-directories. This main README file describes the overall project.

The project code is written in `R` version 4.3.2 as R or R markdown scripts with `Rstudio` version 2024.04.2+764.

All model inputs stored as either `.csv` and outputs stored as `.csv` files.

## `0\_Setupmodel.Rmd`:

This script is designed to create a new project for the Simple HIV model.

The example used is project investigating the impact of COVID-19 (COVID-19impact).

The script creates a project directory in the "projects" directory and relevant sub-directories and copies template files in "templates" directory to the "data" folder (sub-folder).

These template files are pre-filled with example data from the `COVID-19impact` project to demonstrate how the model can be configured and run. You can modify these templates to fit your own project requirements.

3 csv files and 1 .R file included in this sub-folder.

\* `project\_specs.csv`: For specifying project name, starting year, number of time-step, step size, and beta option.

\* `hiv\_time\_series\_data.csv`: For specifying time series parameters.

\* `hiv\_fixed\_data.csv`: For specifying fixed parameters.

\* `GenerateScenarios.R`: For specifying scenarios.

## `1\_input and run model.Rmd`:

This script is designed to input project-specific data and run the SIMPLE HIV model. The script allows users to configure the model for different scenarios and generate relevant outputs such as infection results, cumulative infections, diagnoses, and more. The example used here is for the "COVID-19impact" project.

### How to Use:

#### 1. Prerequisites:

Before running this script, ensure the following step is complete:

- You have successfully run the script `0\_Setupmodel.Rmd` to 1). create the "projects" directory and relevant sub-directories and 2). copy .csv files in the "templates" folder to the "data" folder and `GenerateScenarios.R` to the "projects" directory.

#### 2. Initialization:

- Set the working directory and ensure all required libraries (`dplyr\_1.1.4`) and functions (`BetaOption.R`, `Parameters.R`, and `SimpleHiv.R`) are loaded. This step also clears the workspace to prevent any conflicts.

#### 3. User Inputs:

- Modify the `selectedproject` variable to select the project for which you want to run the model.

- Ensure that the project folder, data folder, and output folder paths are correctly set.

#### 4. Project Data:

The script will load the following .csv files from the project's "data" sub-folder:

- `project\_specs.csv`: Defines project specifications such as start year, number of time-steps, and beta options.

- `hiv\_time\_series\_data.csv`: Contains time-series data for parameters like condom use, testing rates, and more.

- `hiv\_fixed\_data.csv`: Holds fixed parameter values for the model.

- `prep\_estimate\_predicted\_use.csv` and `condom\_estimate\_predicted\_use.csv`: Scenario-specific data for PrEP use and condom use.

#### 5. Scenarios:

The following scenarios are generated and simulated:

- \*\*With COVID-19 impact\*\*: It was shown as "COVID-19 scenario" in the paper, where all changes in monthly parameter values occur and reflect the impact of COVID-19.

- \*\*Without COVID-19 impact\*\*: It was shown as "no COVID-19 scenario" in the paper, a counterfactual scenario where the input parameters remained at their value in December 2019 and corresponded to pre-COVID-19 values.

- \*\*Without COVID-19 impact and with PrEP promotion\*\*: It was shown as "no COVID-19 plus PrEP scenario" in the paper, an alternative counterfactual no COVID-19 scenario where PrEP scale-up and the corresponding decrease in condom use continued during 2020-2020 following the pre-2019 trends, with the other parameters remaining at their December 2019 value.

#### 6. Running the Model:

- The script sets up multiple scenarios based on user-defined parameter sets. Each scenario represents a different possible outcome, for example, with or without the impact of COVID-19 or with PrEP promotion.

- For each scenario, the model simulates the transmission of HIV over the specified time period and outputs results such as the number of new infections, the total number of people living with HIV (PLHIV), and the number of new diagnoses.

#### 7. Simulation with Uncertainty:

The script allows for 1000 simulations using sampled parameter sets to assess uncertainty. Percentile Intervals (PIs) are calculated based on these simulations.

#### 8. Saving Results:

The final output results are saved in the project’s "output" folder as .csv files. The script checks if a file already exists before saving to prevent overwriting previous results.

#### 9. Output Data:

The script generates several output datasets, which are saved in CSV format in the project’s "output" folder. Key outputs include:

- 'infectionresults\_date.csv' \*\*Infection Results\*\*: The dataset contains comprehensive results of the project for each time-step and scenario, including the number of new HIV infections, the total number of PLHIV, the proportion of diagnosed, the number of new diagnoses, cumulative numbers of new HIV infections, cumulative numbers of new diagnoses, and cumulative numbers of new HIV infections and new diagnoses for each year.

- 'infectionresults\_sim\_date.csv' \*\*Infection Results with 1000 simulations\*\*: The dataset contains results across 1000 simulations for each time-step and scenario. It includes the number of new HIV infections, the total number of PLHIV, the proportion of diagnosed, the number of new diagnoses, cumulative numbers of new HIV infections, cumulative numbers of new diagnoses, and cumulative numbers of new HIV infections and new diagnoses for each year.

- 'monthly\_results\_95PI\_date.csv' \*\*Monthly Results with 95% PIs\*\*: The dataset provides monthly results for new HIV infections along with 95% PIs to illustrate the uncertainty in the estimates.

- 'summarized\_results\_95PI\_date.csv' \*\*Summarised Results with 95% PIs\*\*: The dataset summarizes key outcomes for each scenario, including the cumulative numbers of new HIV infections and diagnoses, as well as the total number of PLHIV at the end of the simulation period, all presented with 95% PIs.

- 'yearly\_cumulative\_results\_date.csv' \*\*Yearly Cumulative Results with 95% PIs\*\*: The dataset shows yearly cumulative numbers of new HIV infections and diagnoses, with 95% PIs provided for each scenario at the end of the simulation period.

\*\*"date" above is the date we saved the .csv file.\*\*

## `2\_Generate\_figures.Rmd`: This script is to generate figures for Simple HIV model.

This script is designed to generate figures based on the results from the Simple HIV model. It creates visualizations that help illustrate key outputs such as the number of new HIV infections and annual diagnoses. The script leverages the ggplot2 package and related libraries to create these figures.

### How to Use:

#### 1. Prerequisites:

Before running this script, ensure the following steps are complete:

- You have successfully run the script `1\_input and run model.Rmd` to generate the necessary results.

- All required data files are available in the "output" folder, including `infectionresults\_date.csv`, `monthly\_results\_95PI\_date.csv`, and `yearly\_cumulative\_results\_date.csv`.

#### 2. Generating Figures:

The script generates several figures based on the simulation results:

##### a. \*\*New HIV Infections Over Time\*\*:

- This section generates a figure showing the number of new HIV infections over time for different scenarios (e.g., with and without COVID-19 impact, and with PrEP promotion).

- \*\*Figure Characteristics\*\*:

- Smooth lines with shaded intervals (bands) represent 95% percentile intervals (PIs).

- Scenarios are color-coded for clarity.

- The figure is saved as `newinfectionfig\_sim\_band\_date.png`. "date" here is the date we saved the figure.

##### b. \*\*Annual Diagnoses with Model Estimates vs. Surveillance Data\*\*:

- This figure compares the annual number of diagnoses from the model with real-world HIV diagnoses from surveillance reports.

- \*\*Figure Characteristics\*\*:

- Bar plots show the modeled diagnoses and actual surveillance data for 2020, 2021, and 2022.

- Error bars represent the 95% PIs for the model predictions.

- The figure is saved as `diagnosis\_peryear\_fig\_date.png`. "date" here is the date we saved the figure.

#### 3. Saving Figures:

The script includes a `SaveFigure` function that ensures figures are saved in the "figures" folder with the correct file name. If the file already exists, it will not be overwritten.