

# ECDC HIV Modelling Tool User Manual

*Version 1.3.0*

*European Centre for Disease Prevention and Control*

*20 December 2017*

## 1 Table of Contents

2	Introduction .....	3
2.1	Incidence Method .....	3
2.2	London Method .....	3
3	Technical details and installation.....	3
3.1	Prerequisites .....	4
3.2	Installation .....	4
3.3	Support.....	5
4	Input data sets .....	5
4.1	Datasets.....	5
4.1.1	Before creating datasets.....	5
4.1.2	Populations.....	6
4.1.3	Preparation of datasets .....	7
4.2	General format of data sets.....	9
4.2.1	Header row.....	10
4.2.2	Data row .....	10
4.2.3	Excel and regional settings .....	11
5	Starting the tool .....	12
5.1	SETTINGS .....	12
5.1.1	Interface .....	12
5.1.2	Model.....	13
5.2	HELP.....	13
5.3	ABOUT .....	14
5.4	introduction.....	14
5.5	models.....	14
5.5.1	Population list .....	15
5.5.2	Managing populations .....	15
5.5.3	Parameter settings and outputs .....	16
6	Creating a new model .....	16
6.1	Step 1 of 5: paths to input data and output folder .....	17

6.2	Step 2 of 5: input data availability.....	18
6.3	Step 3 of 5: Incidence Method data availability.....	19
6.4	Step 4 of 5: Incidence Method checks.....	20
6.5	Step 5 of 5: London Method checks.....	21
7	Parameter specification.....	21
7.1	Meta Information.....	22
7.1.1	Combinations of populations.....	23
7.2	Incidence Method.....	24
7.2.1	Diagnosis probability.....	24
7.2.2	Specifying time intervals.....	25
7.2.3	Specifying presumed shape.....	26
7.2.4	Examples.....	27
7.3	London Method.....	32
7.3.1	Calculation based on.....	33
7.4	Incidence Method advanced parameters.....	33
7.4.1	Calendar year ranges.....	34
7.4.2	Full/partial data.....	36
7.4.3	Confidence intervals.....	37
7.4.4	Incidence curve.....	38
7.4.5	Maximum likelihood.....	43
7.4.6	Diagnosis rate.....	44
7.4.7	Country-specific settings.....	44
7.5	London Method advanced parameters.....	44
7.5.1	Calendar year ranges.....	45
7.5.2	95% confidence intervals.....	45
8	Outputs.....	45
8.1	Goodness-of-fit.....	45
8.2	Tables.....	48
8.3	Graphs.....	48
8.3.1	Incidence Method.....	49
8.3.2	London Method.....	49
8.4	CSV and Excel files.....	50
9	Appendix.....	50
9.1	Result_main.....	50
9.2	Result_BS.....	54
9.3	Result_main_ConfIntervals.....	54
9.4	Param_BS.....	54
9.5	Result_LM.....	56

## 2 Introduction

The ECDC HIV Modelling Tool is a tool developed by ECDC in collaboration with international partners to provide estimates of the number of people living with HIV, including those not yet diagnosed. The tool can also estimate the annual number of new HIV infections, the average time between infection and diagnosis, and the number of people in need of treatment according to CD4 counts.

To achieve all of this, the tool needs only routinely collected HIV surveillance data. Nearly all countries in the European region report annual HIV and AIDS diagnoses to the TESSy database hosted at ECDC.

The tool produces estimates based on two different methods. The first method is the so-called Incidence Method, which requires the most data but can also provide the most detailed estimates<sup>1</sup>. This method first estimates HIV incidence over time as well as time to diagnosis by CD4 count strata and then estimates the undiagnosed HIV-positive population.

The second method is the London Method<sup>2</sup>. This method requires less data, typically surveillance data for just one year, but, as a result, can only provide less detailed estimates for the undiagnosed population. In particular, the method only generates estimates of the number of undiagnosed individuals in immediate need of antiretroviral treatment, based on CD4 count thresholds of 200 or 350 cells/mm<sup>3</sup>, which are also given by the Incidence Method.

### 2.1 Incidence Method

The Incidence Method uses a mathematical model to estimate the annual number of HIV infections, the probability of being diagnosed with HIV depending on CD4 counts, and the time between infection and diagnosis. This mathematical model describes the progression of HIV from the time of infection to diagnosis or development of AIDS in the absence of antiretroviral treatment.

Once the number of HIV infections and the time from infection to diagnosis have been estimated, the Incidence Method calculates other outcomes of interest such as the number of people living with HIV, including those not yet diagnosed.

### 2.2 London Method

The London Method is based on the assumption that undiagnosed HIV-positive individuals who develop AIDS or other HIV-related symptoms of sufficient severity will present for care and as a result be diagnosed with HIV<sup>2</sup>. The rate at which such symptoms develop depends on the level of CD4 counts and is determined from data on untreated HIV-positive individuals. From the observed number of symptomatic HIV diagnoses in a specific CD4 interval, the tool is able to estimate the total number of HIV-positive individuals for that particular CD4 count interval. This method only works for CD4 count levels below 350 cells/mm<sup>3</sup> for which the rate of symptoms is sufficiently large.

## 3 Technical details and installation

The ECDC HIV Modelling Tool is a standalone application executed directly on the user's computer running a Windows<sup>®</sup> operating system. It is a 32-bit application and supports both 32- and 64 bit versions of Windows Vista SP2, Windows 7 SP1, Windows 8 and later.

## 3.1 Prerequisites

ECDC HIV Modelling Tool combines an interface written in C# utilizing Microsoft .Net technology with a calculation engine written in C. This imposes certain dependencies on the operating system environment. Two packages must be installed prior to running the tool (both Microsoft products):

1. **Microsoft .Net Framework 4.5.1 or later (4.5.2 available)**  
Download link 4.5.1: <http://www.microsoft.com/en-us/download/details.aspx?id=40779> (this version is included in Windows 8.1 by default).  
Download link 4.5.2: <http://www.microsoft.com/en-us/download/details.aspx?id=42642>.
2. **Visual C++ Redistributable Packages for Visual Studio 2017 (version x86)**  
Download link: <https://go.microsoft.com/fwlink/?LinkId=746571>.

Chances are that these two packages are already installed as prerequisites to other software.

## 3.2 Installation

The ECDC HIV Modelling Tool can be downloaded from <http://ecdc.europa.eu/en/publications-data/hiv-modelling-tool>. The latest version (v1.3.0 as of 20 December 2017) is mentioned on the top of that list. The software is distributed as a zip-archive of size around 4.5MB and requires only a few simple installation steps requiring no computer administration rights:

1. Create a folder in which to install the tool, for example `D:\My Documents\HIV Modelling Tool`.
2. Download the zip-file with the latest version of the tool and save it to the folder created in step 1.
3. Unzip the downloaded zip file by selecting [Extract here](#) as an option available with a right-click on the file.

The tool zip package contains 6 files in the root folder: one executable `HIVModellingGui.exe` and five auxiliary library files with extension `dll`:

Name	Date modified	Type	Size
Documentation	29/6/2016 15:37	File folder	
Examples	27/6/2016 22:26	File folder	
Templates	27/6/2016 22:26	File folder	
EPPlus.dll	29/6/2016 15:33	Application extens...	1,097 KB
FirstFloor.ModernUI.dll	29/6/2016 15:33	Application extens...	242 KB
HIVModellingGui	29/6/2016 15:33	Application	1,047 KB
HIVModellingLibrary.dll	29/6/2016 15:33	Application extens...	112 KB
OxyPlot.dll	29/6/2016 15:33	Application extens...	495 KB
OxyPlot.Wpf.dll	29/6/2016 15:33	Application extens...	147 KB

The tool can be started by double-clicking the file `HIVModellingGui.exe`. Once the tool is running one can inspect the [about](#) tab. Successful installation can be confirmed by inspecting if the version numbers for the GUI and Library are printed:

## ECDC HIV Modelling Tool

introduction models

### ABOUT

GUI version: 1.3.0  
Library version: 1.3.0  
Build date: 2017/11/27 09:32:38

A missing Library version number indicates that package 2 of the prerequisites listed above (Visual C++ Redistributable Packages for Visual Studio 2017) is not installed. It should be installed prior to running calculations with the tool.

### 3.3 Support

For technical support and reporting problems please contact [HIV.Modelling@ecdc.europa.eu](mailto:HIV.Modelling@ecdc.europa.eu).

## 4 Input data sets

This section describes the data sets that need to be prepared before running the tool. Data are typically retrieved from national or regional HIV surveillance systems. Two examples of input data sets are available in the zip-archive that also contains the tool executable.

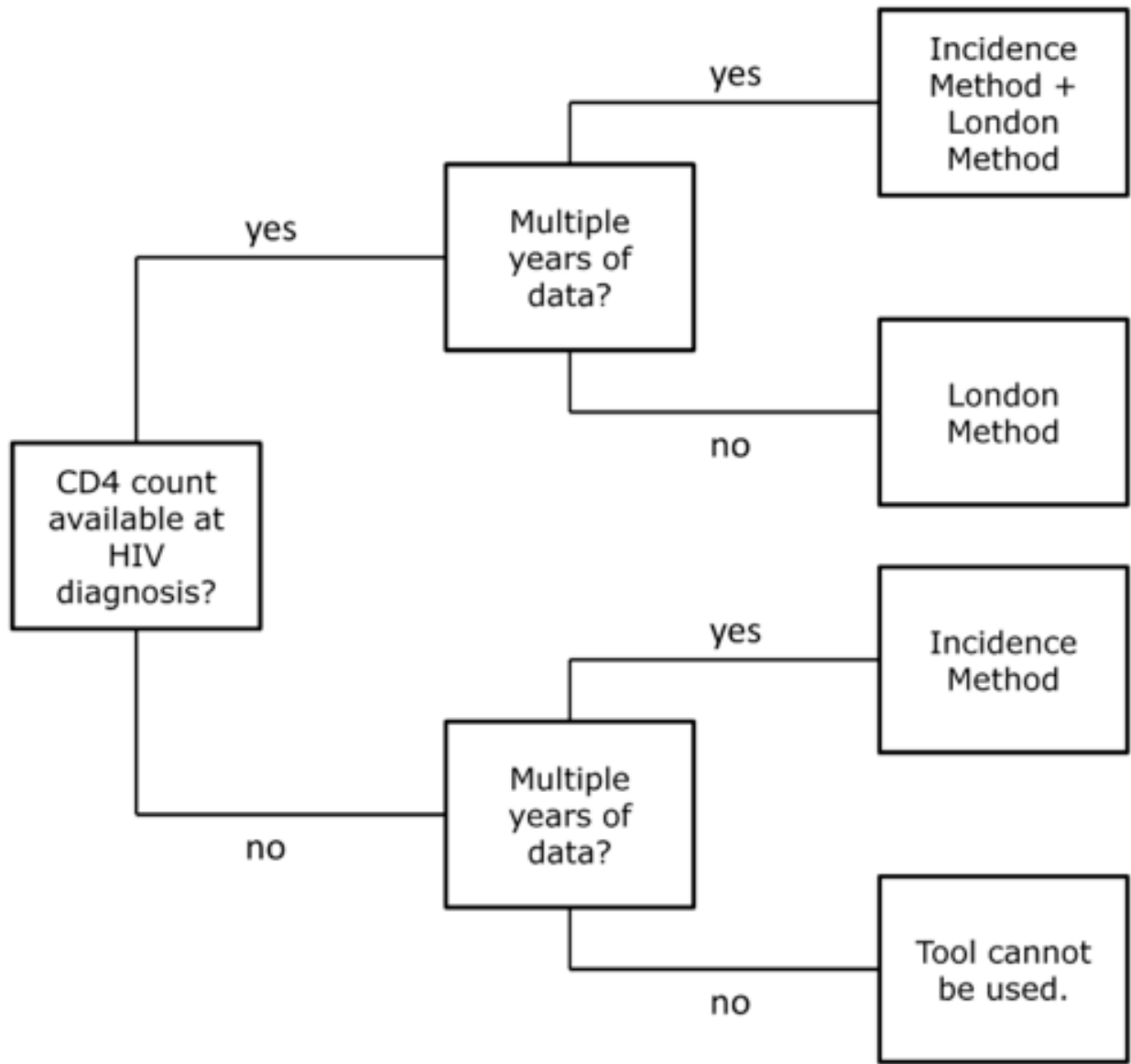
### 4.1 Datasets

#### 4.1.1 Before creating datasets

Creating the datasets that are required by the two methods may be a considerable amount of work. To avoid unnecessary preparation of datasets there are two main issues to consider:

- Are there data available on **CD4 counts at the time of HIV diagnosis**?
- Are **surveillance data** available for **multiple calendar years**?

Depending on the answers to these questions, one or both methods may be applicable according to the scheme below.



The Incidence Method requires surveillance data over multiple years, ideally covering the duration of the HIV epidemic in a country. This method will work with or without CD4 counts at the time of HIV diagnosis, although the first option is preferred. The Incidence Method will also work if data on CD4 counts are only available for several years.

The London Method requires at least one year of data, including CD4 counts at the time of HIV diagnosis. This method will also work with multiple years of data.

Both the Incidence Method and the London Method are based on CD4 cell decline in adult HIV-1-positive individuals. Therefore, *HIV-2-positive individuals and HIV-positive children below 16 years of age should be excluded* from the datasets.

**4.1.2 Populations**

Before using the tool, the user may define *populations* in which the total national or regional HIV population can be divided. Distinguishing one or more populations may be appropriate if the user expects major differences in time between infection and diagnosis between the populations. Differences in time to

diagnosis may be indicated by differences in the mean or median CD4 count at the time of diagnosis. Still, the tool will also work when all HIV-positive individuals in a country are considered as one single population. In that case, however, estimates of time to diagnosis in the Incidence Method will be an average over the total population.

Populations can be based, for instance, on route of transmission: men who have sex with men, heterosexual men and women, injecting drug users. Other examples of populations include native HIV-positive individuals and migrants, or people living HIV in a specific city. The choice of populations will depend on the nature of the HIV epidemic in a given country.

When defining populations, the user should also consider the population size. The outcomes of the ECDC HIV Modelling Tool are harder to interpret for smaller populations with only a few new HIV diagnoses per year. A good approach may therefore be to first consider all HIV-positive individuals as a single population and then as a next step disaggregate this single population into smaller populations.

### 4.1.3 Preparation of datasets

The Incidence Method and the London Method each require their own datasets. There is no overlap between these datasets, although these data will most likely be extracted from the same surveillance system.

#### Incidence Method

The following datasets are necessary for the Incidence Method:

##### *Required*

Data set	Description
<b>HIV</b>	total annual number of HIV diagnoses
<b>AIDS</b>	total annual number of AIDS cases; can be omitted if the calendar years in <b>HIVAIDS</b> cover the duration of the HIV epidemic in a country
<b>HIVAIDS</b>	annual number of HIV diagnoses with a concurrent AIDS diagnosis, i.e., an AIDS diagnosis within, for instance, 3 months of HIV diagnosis

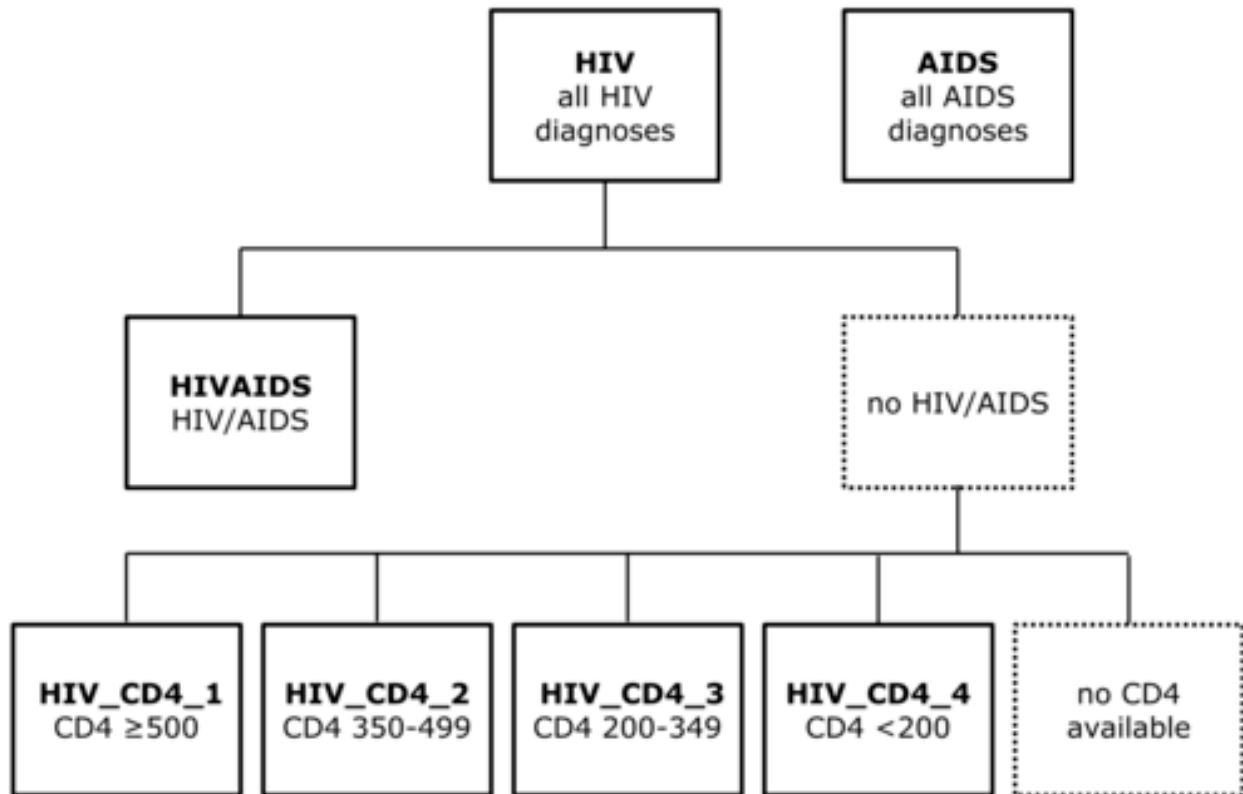
##### *Recommended*

Data set	Description
<b>HIV_CD4_1</b>	annual number of HIV diagnoses with CD4 $\geq$ 500 cells/mm <sup>3</sup> and no concurrent AIDS diagnosis
<b>HIV_CD4_2</b>	annual number of HIV diagnoses with CD4 350-499 cells/mm <sup>3</sup> and no concurrent AIDS diagnosis
<b>HIV_CD4_3</b>	annual number of HIV diagnoses with CD4 200-349 cells/mm <sup>3</sup> and no concurrent AIDS diagnosis
<b>HIV_CD4_4</b>	annual number of HIV diagnoses with CD4 < 200 cells/mm <sup>3</sup> and no concurrent AIDS diagnosis

##### *Optional*

Data set	Description
<b>Dead</b>	annual number of deaths (of any cause) among diagnosed individuals

These datasets should be prepared according to the scheme below. Only the datasets in the boxes with a solid border need to be provided; data items in the boxes with a dashed border are derived by the tool informed by other data items.



The dataset **HIV** contains all HIV diagnoses per year.

**AIDS** contains all AIDS diagnoses per year, *including* those that are in **HIVAIDS**.

**HIVAIDS** contains all HIV/AIDS diagnoses, i.e., HIV diagnoses with a concurrent AIDS diagnosis, *irrespective of the CD4 cell count at the time of diagnosis*.

**HIV\_CD4\_1** to **HIV\_CD4\_4** contain HIV diagnoses with *no* concurrent AIDS diagnosis and with CD4 counts at the time of diagnosis in the specified range.

#### Notes

- It is possible to run the Incidence Method if there are no data on CD4 counts at the time of diagnosis, but this is not recommended.
- The Incidence Method calculates the number of HIV-positive individuals who are still alive by subtracting the number who died, as given in dataset **Dead**, from the estimated number of individuals ever infected. Therefore, if the annual number of deaths among diagnosed individuals is (partly) missing, the Incidence Method cannot correctly determine the number of HIV-positive individuals who are still alive. This does, however, not affect estimation of the annual number of new infections.

#### London Method

The London Method is based on the assumption that undiagnosed HIV-positive individuals who develop AIDS or other HIV-related symptoms of sufficient severity, or which are sufficiently specific to HIV, will present for care and as a result be diagnosed with HIV<sup>2</sup>. Such HIV-related symptoms would typically refer to those listed as category B and C conditions (CDC-B and -C events) in the 1993 revised CDC classification system<sup>3</sup>. Only symptoms which are assumed to be caused by HIV are of interest, so symptoms related to a bacterial sexually transmitted infection, for example, should not count.



Ideally, the London Method would require data on the number of symptomatic HIV diagnoses. However, quite often the only data that are available in surveillance systems, are data on HIV/AIDS diagnoses, which may underestimate the true number of HIV diagnoses as a result of HIV-related symptoms.

The datasets for the London Method need to be specified as given below. Datasets starting with **HIV\_CD4\_LM\_1** contain data on HIV/AIDS diagnoses, while datasets starting with **HIV\_CD4\_LM\_2** contain data on symptomatic HIV diagnoses, i.e., HIV/AIDS diagnoses plus HIV diagnoses with other HIV-related symptoms. The tool will run the London Method when either or both are provided. The datasets should contain data for at least one calendar year.

*Required*

Data set	Description
<b>HIV_CD4_LM_1_0</b>	number of HIV/AIDS diagnoses without CD4 count
<b>HIV_CD4_LM_1_1</b>	number of HIV/AIDS diagnoses with CD4 < 20 cells/mm <sup>3</sup>
<b>HIV_CD4_LM_1_2</b>	number of HIV/AIDS diagnoses with CD4 20-49 cells/mm <sup>3</sup>
<b>HIV_CD4_LM_1_3</b>	number of HIV/AIDS diagnoses with CD4 50-99 cells/mm <sup>3</sup>
<b>HIV_CD4_LM_1_4</b>	number of HIV/AIDS diagnoses with CD4 100-149 cells/mm <sup>3</sup>
<b>HIV_CD4_LM_1_5</b>	number of HIV/AIDS diagnoses with CD4 150-199 cells/mm <sup>3</sup>
<b>HIV_CD4_LM_1_6</b>	number of HIV/AIDS diagnoses with CD4 200-249 cells/mm <sup>3</sup>
<b>HIV_CD4_LM_1_7</b>	number of HIV/AIDS diagnoses with CD4 250-299 cells/mm <sup>3</sup>
<b>HIV_CD4_LM_1_8</b>	number of HIV/AIDS diagnoses with CD4 300-350 cells/mm <sup>3</sup>
<b>HIV_CD4_LM_1_9</b>	number of HIV/AIDS diagnoses with CD4 > 350 cells/mm <sup>3</sup>

*Required*

Data set	Description
<b>HIV_CD4_LM_2_0</b>	number of HIV/AIDS diagnoses without CD4 count
<b>HIV_CD4_LM_2_1</b>	number of HIV/AIDS diagnoses with CD4 < 20 cells/mm <sup>3</sup>
<b>HIV_CD4_LM_2_2</b>	number of HIV/AIDS diagnoses with CD4 20-49 cells/mm <sup>3</sup>
<b>HIV_CD4_LM_2_3</b>	number of HIV/AIDS diagnoses with CD4 50-99 cells/mm <sup>3</sup>
<b>HIV_CD4_LM_2_4</b>	number of HIV/AIDS diagnoses with CD4 100-149 cells/mm <sup>3</sup>
<b>HIV_CD4_LM_2_5</b>	number of HIV/AIDS diagnoses with CD4 150-199 cells/mm <sup>3</sup>
<b>HIV_CD4_LM_2_6</b>	number of HIV/AIDS diagnoses with CD4 200-249 cells/mm <sup>3</sup>
<b>HIV_CD4_LM_2_7</b>	number of HIV/AIDS diagnoses with CD4 250-299 cells/mm <sup>3</sup>
<b>HIV_CD4_LM_2_8</b>	number of HIV/AIDS diagnoses with CD4 300-350 cells/mm <sup>3</sup>
<b>HIV_CD4_LM_2_9</b>	number of HIV/AIDS diagnoses with CD4 > 350 cells/mm <sup>3</sup>

## 4.2 General format of data sets

Data sets need to be prepared in comma-separated values (CSV) text format. CSV files can easily be created from packages like Microsoft Excel or SAS, or in a text editor like Notepad.

In a CSV file, the data are arranged in records, or rows in a spreadsheet, and the records are made up of a set of fields that represent epidemiological variables. Each field is separated from the next by a so-called list separator, except for the last field in the record, which is followed by the [RETURN] character. The tool automatically recognizes list separator used in the input data files, thus files created on computers with different regional settings are still recognized and properly interpreted.

Each data set contains epidemiological data for a certain number of mutually exclusive populations that together form the total HIV population in a country or region.

### 4.2.1 Header row

Each data set needs a header row. In a CSV file, the header row is simply a list of variable names and may look like this with comma (",") as list separator:

- in Notepad

```
Year,population_1,population_2,population_3
```

- in Excel

```
Year population_1 population_2 population_3
```

In this case the CSV file contains information on 3 groups. The names of these groups are arbitrary but should be same in all data sets that are necessary for the tool.

### 4.2.2 Data row

The header row is followed by rows containing epidemiological data, one row for each calendar year. Each data row contains a calendar year followed by one or more numbers, for instance the number of AIDS diagnoses. Numbers do not necessarily have to be integers, which could be the case, for instance, when a correction for reporting delay is made.

- in Notepad

```
1982,0,0,0
1983,0,0,0
1984,2,1,3
1985,17,15,20
1987,20,23,20
...
2013,51.5,30,25
```

- in Excel

```
1982    0    0    0
1983    0    0    0
1984     2    1    3
1985   17   15   20
1987   20   23   20
...
2013  51.5  30   25
```

In the example above, for the first population there are zero diagnoses in 1982 and 1983, 2 in 1984, 17 in 1985, and so on. For 2013, there are 50 observed diagnoses, but corrected for a reporting delay of 5% assuming for instance that 5% of the diagnoses has not yet been reported the expected number would be 51.5.

For calendar years in which the surveillance system captured a specific data item but no data (diagnoses) were observed, there should be a corresponding row with 0 diagnoses in the CSV file (as in the example above for 1982 and 1982). In contrast, for calendar years in which the surveillance system did not yet capture a specific data item, there should be *no* corresponding row (as in the example above for 1981

and earlier years). Putting a 0 in this case may lead to wrong results, because the tool will treat a number of 0 diagnoses the same was a any other number.

Note that the tool will assume 0 observations or diagnoses for *intermediate* years that are missing in the input datasets (as in the example above for 1986).

### 4.2.3 Excel and regional settings

As mentioned already the tool automatically detects both field and decimal separators used in the data sets and can properly interpret values. Output data sets are in CSV format as well and the tool will use the field and decimal separators from the operating systems' regional settings.

However, we would like to comment here about the behaviour of Excel when it comes to dealing with CSV files. Many users will most likely use that software to prepare input data sets and examine results.

Sharing CSV files for opening in Excel where different regional settings are applied can be troublesome. Excel reads the list and decimal separators from the operating system's regional settings and expects the opened CSV file to follow the exact same convention with respect to the separators. Very often Windows sets comma (",") as the list separator and a period (".") as the decimal separator. It follows this convention for many Western European countries, including United Kingdom. However, for other countries, for instance Germany, it applies a different convention - semi-colon (";") as the list separator and comma (",") as the decimal separator. Therefore the same data set will be saved by Excel in different formats depending on the regional settings applied:

- in the UK

```
1982,0,0,0
1983,0,0,0
1984,2,1,3
1985,17,15,20
1987,20,23,20
...;...;...;...
2013,51.5,30,25
```

- in Germany

```
1982;0;0;0
1983;0;0;0
1984;2;1;3
1985;17;15;20
1987;20;23;20
...;...;...;...
2013;51,5;30;25
```

It can be easily checked which setting is currently applied in Excel by opening one of the CSV files in the examples distributed with the tool. The examples include data sets using a comma (",") as commas separator. If Excel does not put each value in to a separate cell, then it means that it did not recognize the list separator. Most likely the data sets that use semi-colon (";") as the list separator will be displayed correctly.

If it is intended to share data between operating systems with different regional settings, then we advise to use Excel file format (file extensions ".xls" or ".xlsx"), rather than CSV (file extension ".csv").

## 5 Starting the tool

To start the ECDC HIV Modelling Tool:

1. go to the folder where the tool was installed.
2. double-click HIVModellingGui.

The tool will open as shown in the figure below and is ready for use.



### 5.1 SETTINGS

In the **SETTINGS** menu, the user can specify several general settings for the tool.

#### 5.1.1 Interface

##### Show RunLog tab

If this box is ticked, the tool will show the **run log** tab in the **models** window. The **run log** can be used for monitoring the models' progression, but this is mainly for development purposes.

##### Charts as png files

- **Export after run**  
If this check box is ticked, charts generated by the tool will be exported as PNG files.
- **Width**  
Specify the width of the output charts (default: 800 pixels).
- **Height**  
Specify the height of the output charts (default: 400 pixels).

## Charts as Excel template

- **Export after run**  
If this check box is ticked, charts generated by the tool will be exported as Excel files.
- **Template file name**  
By default, charts will be exported to macro-enabled Excel files using the `Charts.xlsxsm` template. If macros are disabled and cannot be enabled by the user, please put here `Charts_withoutMacro.xlsx`.

## Results as csv files

- **List separator**  
Default value is the list (field) separator currently specified in the Windows settings. The value can be overridden by the user, but the output CSV files may not be displayed properly in Excel.
- **Decimal separator** Default value is the decimal separator currently specified in the Windows settings. The value can be overridden by the user, but the output CSV files may not be displayed properly in Excel.

## 5.1.2 Model

This section specifies several default settings for the Incidence Method and London Method.

### Range of calculations

'Range of calculations' is the range visible in the slider bars in the `advanced` tab in the `models` window. There is generally no need to change the default settings.

- **Minimum year**  
Specifies the minimum year of the sliders bars (default: 1975).
- **Maximum year**  
Specifies the maximum year of the sliders bars (default: 2020).

### Fit parameters

'Fit parameters' concern the fitting process for the Incidence Method.

- **Maximum number of iterations**  
Specifies the maximum number of iterations in the fitting process (default: 30). If this value is set too small, the fit may not converge, whereas for large values the fit may take a long time.
- **Start random**  
If this box is ticked, the tool will start the fitting process at a randomly generated set of parameters. This can be used to check that the final set of estimated parameters and model outcomes are always the same (up to a few decimals) when the tool is run on the same input data with the same parameter settings.

## 5.2 HELP

In the `HELP` menu, the user can browse through the manual or download the complete manual as a PDF file.

## Browse manual

Here the user can browse through the manual. By clicking items in the table of contents on the right hand side of the window, it is possible to navigate through the manual.

## Open PDF version

Opens a PDF version of the complete manual.

## 5.3 ABOUT

The **ABOUT** tab shows information on the following items:

### About

Current version of the Graphical User Interface (GUI) and the library implementing the Incidence Method and the London Method and the date of release of the current version of the tool. If the version is not displayed, most likely the required software packages are not installed (see [Prerequisites](#)).

### Team in charge

Team of international experts and experts at ECDC who are responsible for developing the tool.

### Acknowledgement

Member States who participated in developing and piloting the tool.

### Funding

Information on how the development of the tool and the Graphical User Interface was funded.

### Further reading

References to publicly accessible literature relating to the tool.

### Contact us

Email address for suggestions and questions regarding the tool.

### Keeping informed

You can choose to

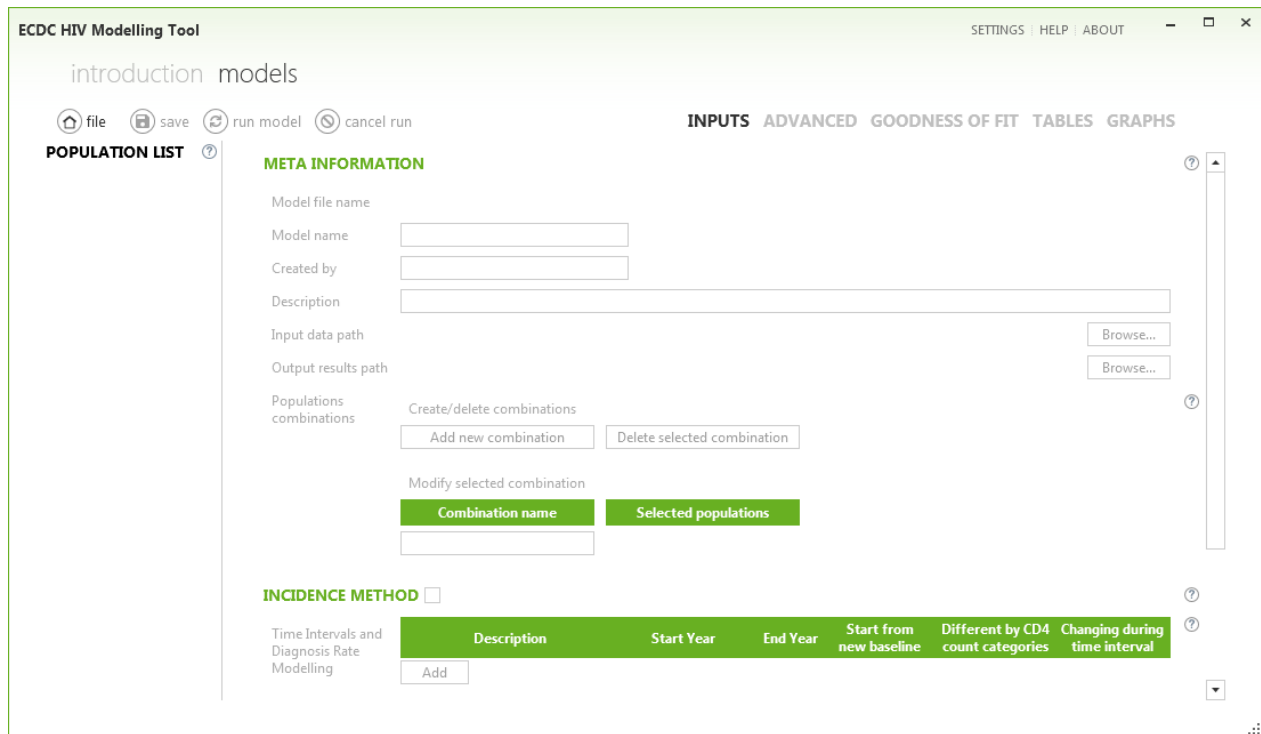
- Receive news about updates of the tool by clicking [Subscribe for updates](#).
- Stop receiving news about the tool by clicking [Unsubscribe from updates](#).

## 5.4 introduction

The **introduction** window gives a brief introduction of the ECDC HIV Modelling Tool.

## 5.5 models

Select **models** to go to the main window of the ECDC HIV Modelling Tool. This window will look like this:



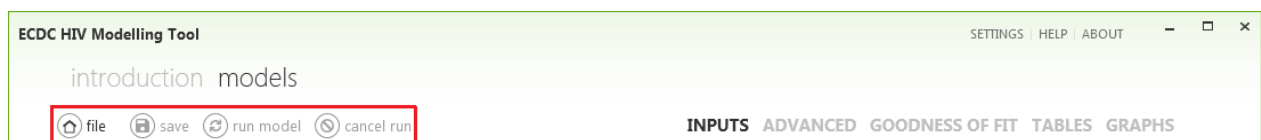
### 5.5.1 Population list

On the left-hand side of the window, **population list** lists all models that were defined by the user and loaded into the tool. Within each model there may be different *populations*.

The currently selected population is shown in green and further information is shown in the main window.

### 5.5.2 Managing populations

On the top left are four buttons that are used to manage models and to run the Incidence Method and the London Method.



#### file

This menu is used to create, save, and open models in the **population list**.

- **New:**  
Create a new model.
- **Open:**  
Open an existing model and add it to the **population list**.
- **Save as:**  
Save and rename the current selected model and its parameter settings.

- **Close:**  
Remove the currently selected model from the tool. Note that this option will only remove the model from the population list. It will still exist in the Models folder.

### save

Save the currently selected model and its parameter settings. This option will be faded out as long as there were no changes in the settings for the selected model.

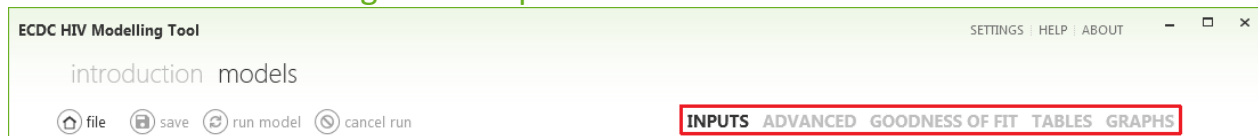
### run model

Run the Incidence Method and/or London Method for the selected population in the **population list**.

### cancel run

Stop running the Incidence Method and/or London Method for the selected population in the list.

## 5.5.3 Parameter settings and outputs



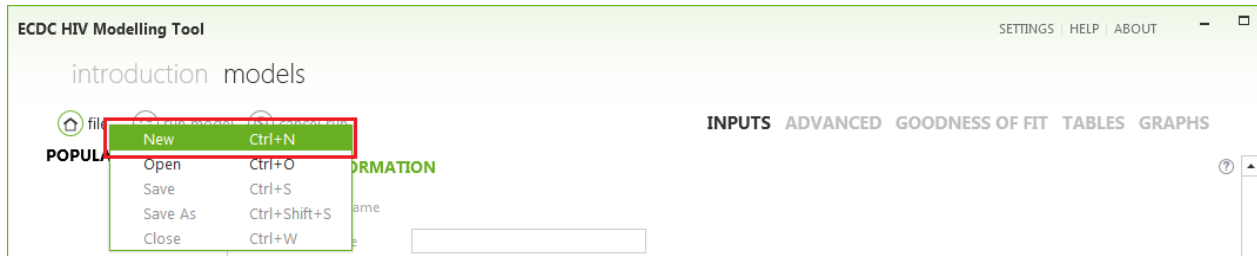
Using the buttons on the right the user can switch between different frames in the **models** window. The selected frame will be indicated in bold font, the other frames will be faded out.

- **Inputs:**  
Basic input parameters and model settings.
- **Advanced:**  
Advanced input parameters and model settings.
- **Run log:**  
(mainly for development purposes) Shows progression of the tool and interim results. By default, this button is hidden unless **INTERFACE->Show RunLog** is checked.
- **Goodness of fit:**  
Tables and figures showing input data with model fits.
- **Tables:**  
Tables with data from the models and model fits.
- **Graphs:**  
Figures with model outcomes.

## 6 Creating a new model

To create a new population in the list select **file->New** on the top left hand side of the **models** window.





A wizard will guide you through the process of creating a new model.

You can use the next and previous buttons to move to the next or previous step in the wizard. Use cancel if you want to stop the wizard.

## 6.1 Step 1 of 5: paths to input data and output folder

The image shows the "NEW MODEL WIZARD" interface, specifically "STEP 1 OF 5: PATHS". The screen is titled "NEW MODEL WIZARD" and "STEP 1 OF 5: PATHS". Under "INPUT PATH", it says "Please, provide the path to the input data folder:" followed by a text input field and a "Browse..." button. Under "OUTPUT PATH", it says "Additionally, please provide the path to the folder where the output files should be stored:" followed by a text input field and a "Browse..." button. At the bottom right, there are four buttons: "cancel", "previous", "next", and "finish".

In the first step of the wizard you are asked to specify

- the location of the input data folder.
- the location of the folder where the outcomes of the tool should be stored.

For each model, which can contain one or more populations, you need a separate input data folder.

## 6.2 Step 2 of 5: input data availability

### NEW MODEL WIZARD

#### STEP 2 OF 5: GENERAL FINDINGS

**A. METHODS**

The specified input data folder contains sufficient information for the following methods to run:

- *Incidence Method*
- *London Method using the following data:*
  - HIV/AIDS diagnosis
  - HIV/AIDS diagnosis + HIV diagnosis with HIV-related symptoms

**B. OUTPUT PATH**

The output results path is valid.

**C. POPULATIONS**

The following populations have been initialized based on the input data provided: category\_1, category\_2, category\_3, category\_4.

cancel previous next finish

### A. Methods

The tool will check whether the required and recommended datasets for each of the two methods are available and correctly specified. For each of the methods, the wizard will show whether the input data files were correctly specified (green) or not (red). More specific information on problems with the input data is given in Step 3 and 5.

Note that the names of the input datasets should be exactly as specified before in [Preparation of datasets](#).

After amending the input data files, you need to go back to Step 1 and continue from there or restart the wizard.

### B. Output path

The tool will check whether the specified output folder exists and if not, it will be created.

### C. Populations

The wizard will show the populations that were specified in the header rows of the input data files. A warning will be given if the number of groups or the names of the groups differ between any of the input files.

#### *Example*

The tool has sufficient input data files in order to be able to run both the Incidence Method and the London Method. The output folder for the results of the tool is correct. The wizard found four different risk groups or populations in the input data files.

## 6.3 Step 3 of 5: Incidence Method data availability

### NEW MODEL WIZARD

#### STEP 3 OF 5: INCIDENCE METHOD DATA AVAILABILITY

Do you have data from the start of the epidemic?

Yes  No

cancel previous **next** finish

In Step 3, the wizard asks if data for the Incidence Method are available from the start of the HIV epidemic.

Selecting **yes** is appropriate if:

- data on total annual number of HIV diagnoses (**HIV**) are available from the start of the epidemic
- *and* for each year there are data on total annual number of AIDS cases (**AIDS**) or data on HIV/AIDS cases (**HIVAIDS**).

Reasons for selecting **no** could be:

- reporting of HIV diagnoses only started a considerable number of years after the HIV epidemic started.
- data on HIV diagnoses before a certain calendar year are incomplete.

Selecting yes or no will only affect the way in which the annual number of HIV infections and other model outcomes are estimated. After the wizard has finished, the answer given here can still be changed in the advanced settings (see [Full/partial data](#) for more information).

## 6.4 Step 4 of 5: Incidence Method checks

### NEW MODEL WIZARD

#### STEP 4 OF 5: INCIDENCE METHOD CHECKS

##### WARNINGS

Due to the following warnings, data for the Incidence Method may be altered during the calculations performed:

- File HIVAIDS.csv does not provide data for 1 year(s).
- AIDS.csv : Data after 1995 are likely to be affected by the availability of combination treatment. Therefore, the end year for the use of this data will be 1995.

In Step 4, the wizard may give some Errors or Warnings about the input files for the Incidence Method.

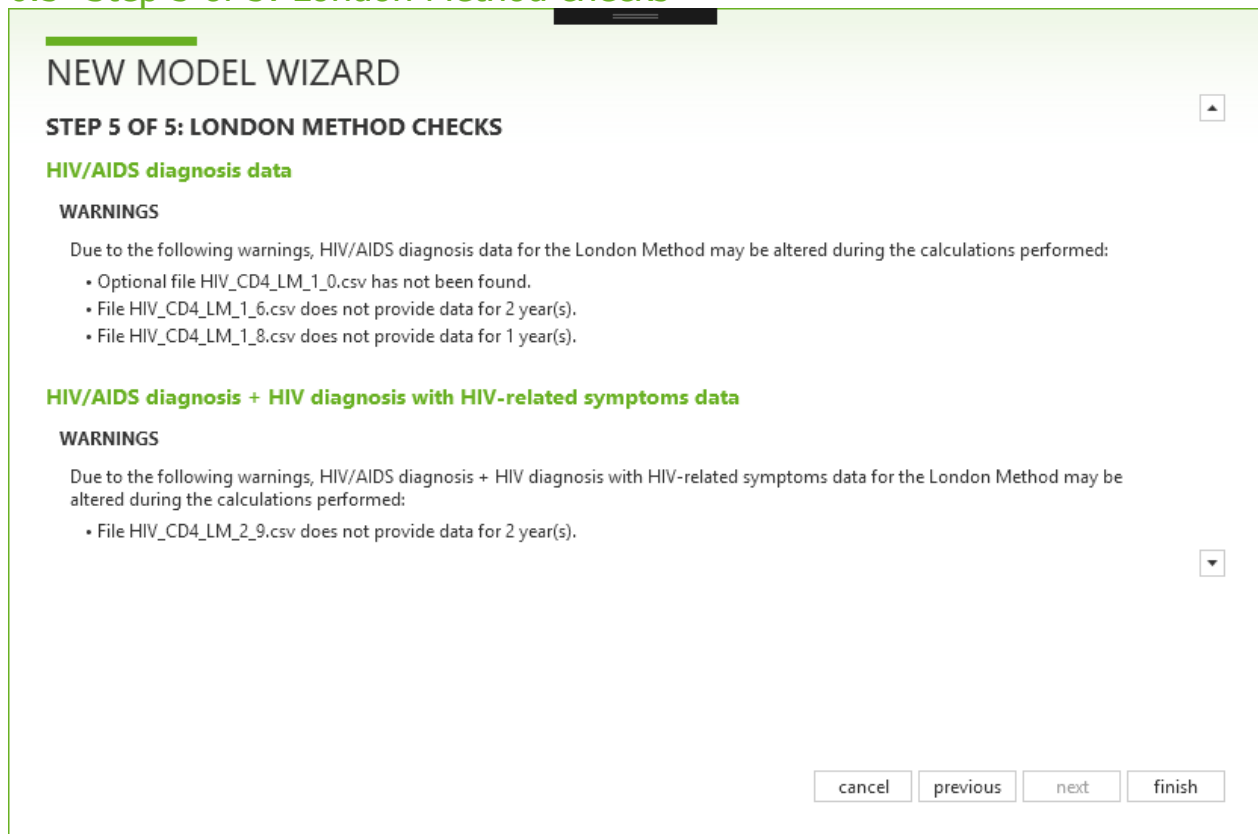
Errors will prevent the Incidence Method from running, while Warnings can be ignored although we encourage the user to check the decisions taken by the wizard.

### *Example*

In the example, the file **HIVAIDS.csv** has no record for one calendar year and the tool will assume there were 0 HIV/AIDS diagnoses in this year.

The input file with total number of AIDS cases contains data after 1995. These data will be ignored by the tool, because these data are likely to be affected by antiretroviral treatment.

## 6.5 Step 5 of 5: London Method checks



In Step 5, the wizard may give some Errors or Warnings about the input files for the London Method.

Errors will prevent the London Method from running, while Warnings can be ignored although we encourage the user to check the decisions taken by the wizard.

### Example

In the example, there is no data in input file **HIV\_CD4\_LM\_1\_0.csv**, which contains number of cases without a CD4 count measurement, and the tool will assume that all cases have a CD4 count. In the other input data files there appear to be missing records for some calendar years. For instance, in **HIV\_CD4\_LM\_1\_6.csv** data are missing for two years. The tool assumes that there were zero patients in these years.

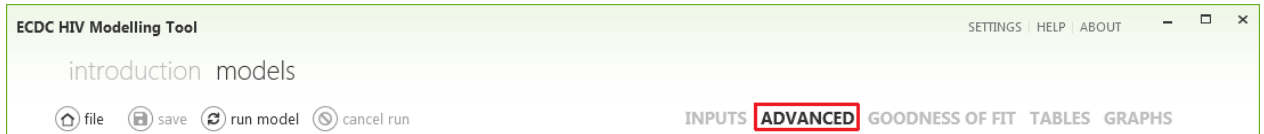
## 7 Parameter specification

The ECDC HIV Modelling Tool has two windows where parameters can be specified:

- **inputs**: basic input parameters and model settings.



- **Meta Information:** general information on the selected model with location of the data and the model output.
- **Incidence Method:** basic parameters for the Incidence Method.
- **London Method:** basic parameters for the London Method.
- **advanced:** advanced input parameters and model settings.



- **Incidence Method:** advanced parameters for the Incidence Method.
- **London Method:** advanced parameters for the London Method.

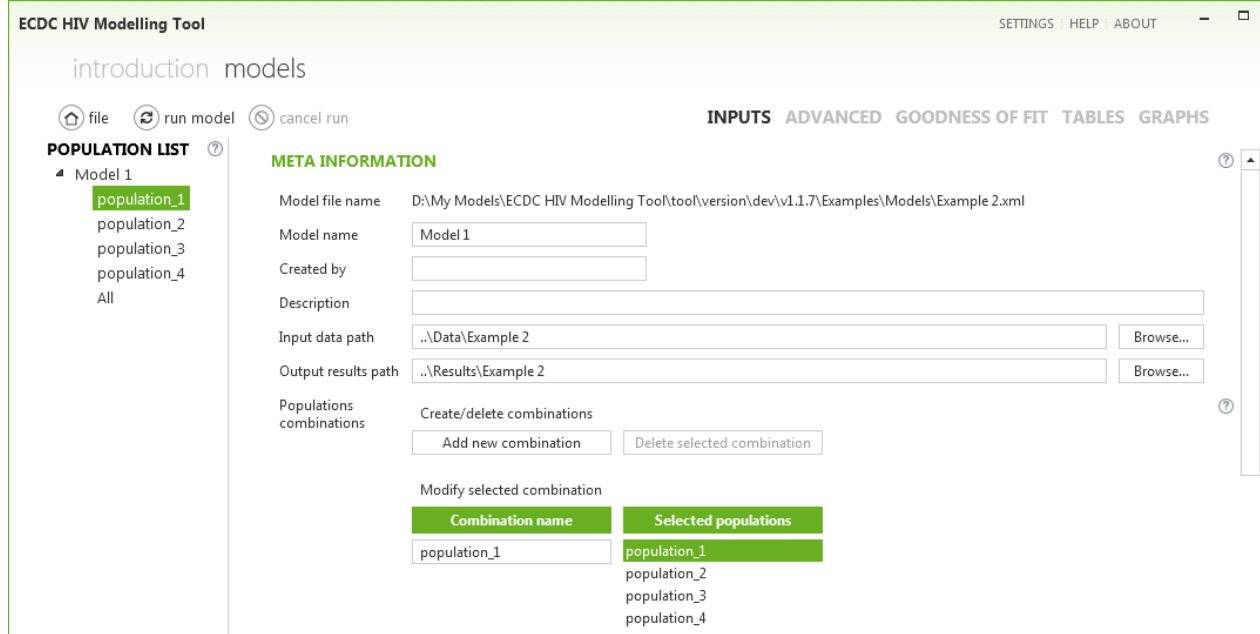
All parameters are pre-specified when initializing a population in the **POPULATION LIST** using the wizard.

### Notes

- The parameters specified in the **inputs** and **advanced** window will be the same for all populations within a model.

## 7.1 Meta Information

**Meta information** shows general information for the selected model.



### Model file name

Information on each population in the **population list** is stored in a XML file in the folder Models that was created when installing the tool.

## Model name

This is the name of the model that appears in the `population list`. The name can be changed by the user.

## Created by

Here the user can enter information on who created this particular model.

## Description

The user can provide an additional description of the model and the population.

## Input data path

This shows the location of the folder containing the datasets as specified in the wizard.

## Output results path

This shows the location of the folder for the results of the tool as specified in the wizard.

## Populations combinations

Here the user can manage combinations of populations within a population.

### 7.1.1 Combinations of populations

By default, populations within a model in the `population list` are the risk groups that were specified in the header row of the input data files. The user can easily add combinations of these risk groups as new populations.

The combination `All` is automatically created by the tool and includes all populations in the model. Hence, `All` will typically contain all diagnosed HIV infections in a country.

## Combination name

This is the name of the population or combination of populations that is currently highlighted in green in the `population list`.

## Selected populations

This shows which populations specified in the input data, highlighted in green, form the currently selected combination in the `population list`.

The following options are available for managing combinations of populations:

- **Add new combination:** Adds a new combination to the `population list`. By default, the new combination includes all populations specified in the input data.
- **Delete selected combination:** Deletes a combination from the `population list`. The combination `All` and the populations specified in the input data cannot be deleted.

## Example

Creating a new combination of populations `population_1`, `population_2`, and `population_4`:

- Click **Add new combination**. This will create a new combination `new name 1` that consists of all risk groups.
- Change **Combination name** to `combi_1+2+4`, for instance.
- Deselect `population_3` under **Selected populations**.

Populations combinations

Create/delete combinations

Add new combination      Delete selected combination

Modify selected combination

Combination name      Selected populations

combi\_1+2+4      population\_1  
 population\_2  
 population\_3  
 population\_4

The combination `combi_1+2+4` has now been added to the `population list`.

## 7.2 Incidence Method

**Incidence Method** contains basic parameter settings for the Incidence Method, which currently only includes the specification of the presumed shape of the diagnosis probability. This diagnosis probability is pre-specified by the wizard, but the user is *strongly encouraged* to change this specification.

### INCIDENCE METHOD

Time Intervals and Diagnosis Rate Modelling	Description	Start Year	End Year	Start from new baseline	Different by CD4 count categories	Changing during time interval
Delete	Time interval 1	1980	1984	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Delete	Time interval 2	1984	1996	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Delete	Time interval 3	1996	2000	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Delete	Time interval 4	2000	2005	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Delete	Time interval 5	2005	2012	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Add						

If the tick box is checked the tool will run the Incidence Method for the selected population. Unchecking the tick box will prevent the tool from running the Incidence Method and the table specifying the diagnosis probability will be faded out.

### 7.2.1 Diagnosis probability

Apart from estimating the annual number of new HIV infections, the Incidence Method also estimates the probability that HIV-positive individuals are diagnosed with HIV when their (unobserved) CD4 count is in one of the four CD4 intervals. This probability is usually unknown and needs to be estimated from the



observed input data. The tool uses the diagnosis probabilities to calculate the expected time between infection and diagnosis by year of infection.

Before using the Incidence Method, the user needs to specify:

- **Time intervals:** indicate *when* the probability of being diagnosed may change.
- **Presumed shape:** indicate *how* the probability of being diagnosed may change.

*Notes*

- It is only necessary to specify the *presumed shape* of the diagnosis probability, as shown in the examples. The tool will determine the best-matching diagnosis probability over calendar time.
- The specification of the diagnosis probability applies to *all* populations within a model.

**Intermezzo: diagnosis probability**

The tool assumes that the probability that an HIV-positive individual is diagnosed with HIV within  $t$  years after infection or entering a certain CD4 count interval is given by the expression  $1 - e^{-\delta t}$ . The parameter  $\delta$  is a rate parameter, or *diagnosis rate*, per time unit. For instance, when  $\delta$  is 0.2 per year, the probability of being diagnosed within 0.5 years is  $1 - e^{-0.1 \times 0.5}$ , which is approximately 0.05 or 5%. The diagnosis probability can thus be seen in terms of proportions: suppose there are 100 undiagnosed HIV-positive individuals at time 0, then 5% of them will be diagnosed after 0.5 years.

An increase in the rate parameter  $\delta$  will result in an increase in diagnosis probability. Within the tool, the user can specify the presumed shape of  $\delta$  and how it may change over calendar time.

7.2.2 Specifying time intervals

Time intervals in which the diagnosis probabilities may change are given in the left part of the table.

**INCIDENCE METHOD**

Time Intervals and Diagnosis Rate Modelling		Description	Start Year	End Year	Start from new baseline	Different by CD4 count categories	Changing during time interval
Delete		Time interval 1	1980	1984	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Delete		Time interval 2	1984	1996	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Delete		Time interval 3	1996	2000	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Delete		Time interval 4	2000	2005	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Delete		Time interval 5	2005	2012	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Add							

**Deleting and adding time intervals**

The following options are available for managing time intervals:

- **Add:** Add a new time interval. Time intervals will be automatically sorted according to Start Year.
- **Delete:** Delete a time interval.

**Description**

Here you can give a short description of the time interval or why this specific time interval was chosen. For instance, the wizard specifies a time interval 1980 to 1984 in which the probability of being diagnosed with HIV is zero due to the absence of serological testing. Other examples of a time interval could be the

time interval corresponding to the pre-cART era, or a new time interval could start in the year when there was a change in HIV testing policy, triggered, for instance, by an outbreak of HIV in a specific population.

### Start and End Year

By clicking on **Start Year** you can change the start year of an interval. This can be done by using the up and down arrows.

#### Notes

- It is not necessary to specify the end year of a time interval, because it will be automatically set to the start year of the next time interval.
- The start year of the first time interval will always be 1980.
- The end year of the last time interval will be the last year for which data are available.
- A higher number of time intervals will result in a more flexible shape of the diagnosis probability curve. However, this also means that a higher number of unknown parameters need to be estimated from the data.

### 7.2.3 Specifying presumed shape

In the right part of the table you can specify how the diagnosis probability may change in a time interval.

#### INCIDENCE METHOD

Time Intervals and Diagnosis Rate Modelling				Start from new baseline	Different by CD4 count categories	Changing during time interval	
	Delete	Description	Start Year	End Year	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
	Delete	Time interval 1	1980	1984	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
	Delete	Time interval 2	1984	1996	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
	Delete	Time interval 3	1996	2000	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
	Delete	Time interval 4	2000	2005	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
	Delete	Time interval 5	2005	2012	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
	Add						

#### Start from new baseline

If this box is not ticked, the diagnosis probability at the start of the interval will be the same as at the end of the previous time interval.

If the tick box is checked, the diagnosis probability will start at a new value.

An example of when this box should be ticked is when diagnosis of HIV by means of serologic tests became possible in 1984.

#### Different by CD4 count categories

Tick this box if you want to diagnosis probabilities to be different for each of the four CD4 count strata. This should only be done if there are data on HIV diagnoses by CD4 count.

#### Changing during time interval

If the tick box is unchecked, the tool will assume that the probability of being diagnosed will not change during the time interval.

If the tick box is checked, the diagnosis probability can increase or, less likely, decrease during the time interval.

An increase would be expected when, for instance, people are more frequently tested for HIV due to increasing awareness. The pace at which the diagnosis probability increases (or decreases) is determined from the input data.

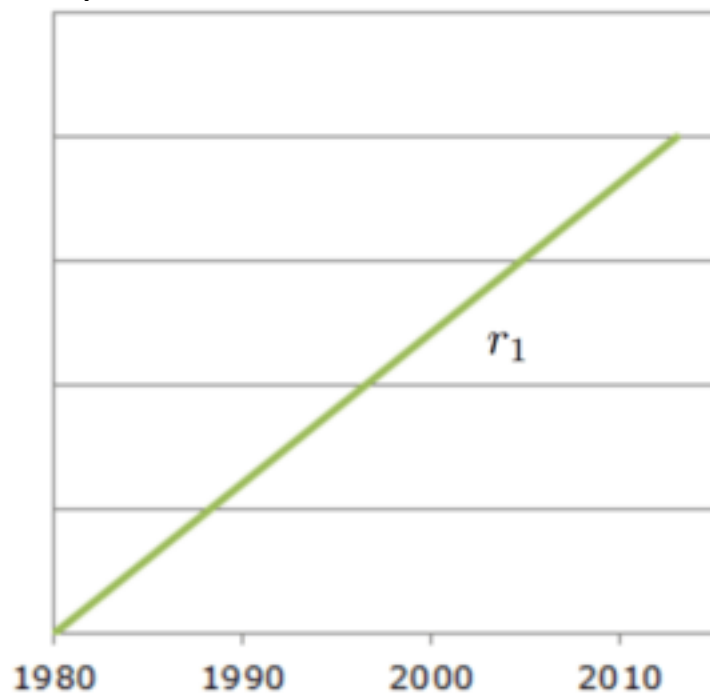
#### Notes

- The more boxes are ticked, the higher the number of parameters that need to be estimated from the data and the longer the Incidence Method will run.
- Be careful with ticking **Start from new baseline** and **Changing during time interval** at the same time as this may lead to negative diagnosis probabilities.
- Parameters describing the diagnosis probability curve are largely determined by the observed number of HIV diagnoses by CD4 count. Therefore, if data on CD4 counts are missing or sparse, it may not be possible to accurately determine the diagnosis probability curve.
- In the time interval 1980 to 1984, there was no testing for HIV and all three boxes should be unchecked. HIV could only be diagnosed when AIDS symptoms appeared. However, it is not necessary to specify the probability of being diagnosed with HIV when AIDS symptoms appear, because this is taken into account by the tool.

### 7.2.4 Examples

The examples below show a few possible shapes of the diagnosis probability curve and how these shapes are specified in the tool.

#### Example I



#### Presumed shape

The diagnosis probability is zero in 1980 and then gradually increases with calendar time at a rate  $r_1$ .

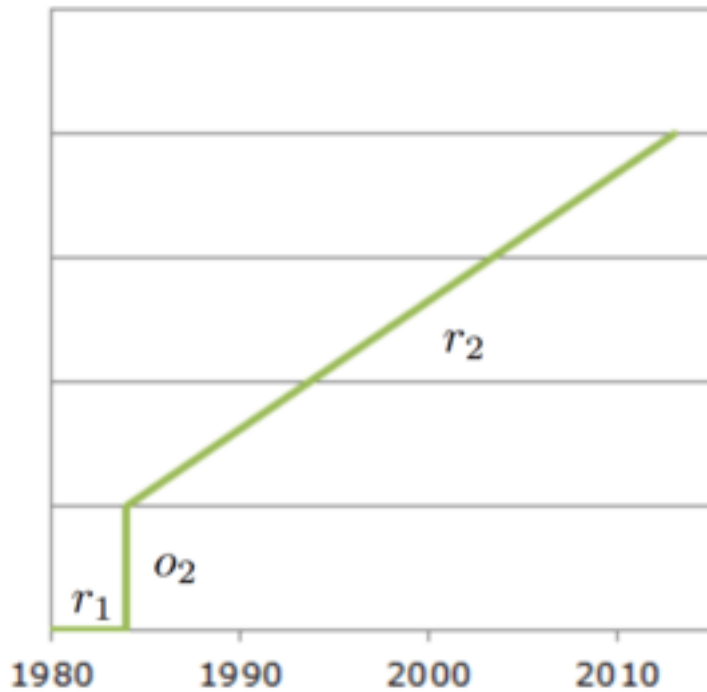
## Time intervals and diagnosis probabilities

INCIDENCE METHOD

Time Intervals and  
Diagnosis Rate  
Modelling

Description	Start Year	End Year	Start from new baseline	Different by CD4 count categories	Changing during time interval
Delete Time interval 1	1980	2012	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>
Add					

### Example II



### Presumed shape

The diagnosis probability is zero from 1980 until 1984. In 1984, the diagnosis probability is  $o_2$  and afterwards increases linearly.

## Time intervals and diagnosis probabilities

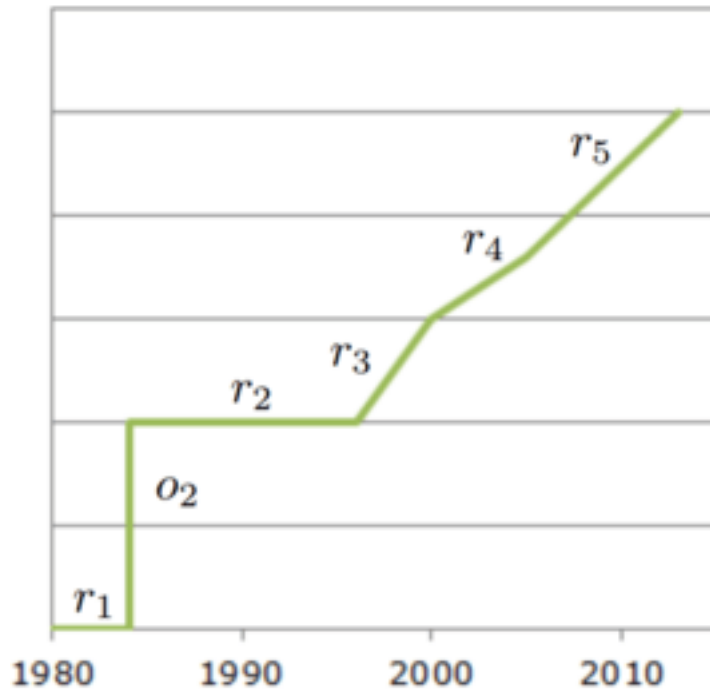
INCIDENCE METHOD

Time Intervals and  
Diagnosis Rate  
Modelling

Description	Start Year	End Year	Start from new baseline	Different by CD4 count categories	Changing during time interval
Delete Time interval 1	1980	1984	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Delete Time interval 2	1984	2012	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>
Add					

### Example III

#### Presumed shape



This is the presumed shape used in<sup>1</sup>. The diagnosis probability is zero from 1980 until 1984. Between 1984 and 1996, the diagnosis probability is at a constant value  $o_2$ . From 1996 onwards, the probability increases linearly, but the rate at which the probability increases with calendar time changes in 2000 from  $r_3$  to  $r_4$  and in 2005 from  $r_4$  to  $r_5$ . Diagnosis rates can be different for each of the four CD4 count categories.

#### Time intervals and diagnosis probabilities

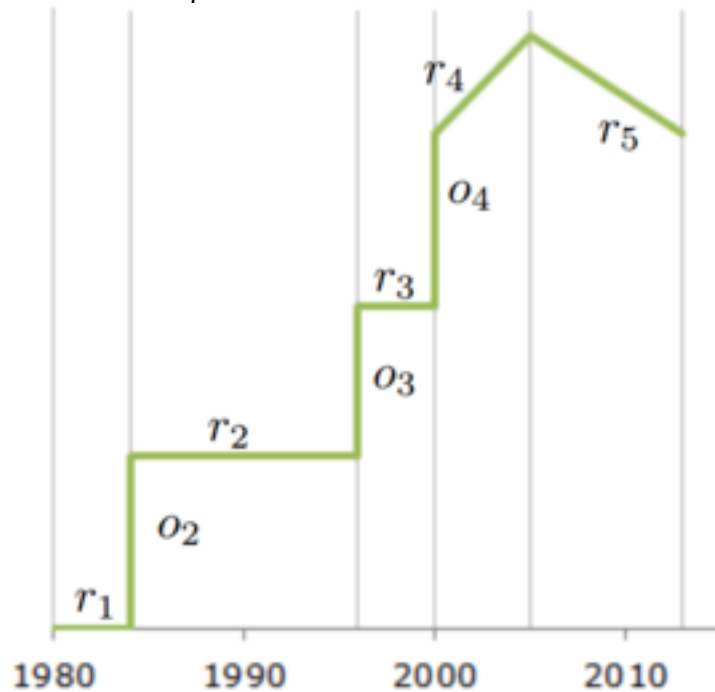
INCIDENCE METHOD

Time Intervals and  
Diagnosis Rate  
Modelling

Description	Start Year	End Year	Start from new baseline	Different by CD4 count categories	Changing during time interval
Delete Time interval 1	1980	1984	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Delete Time interval 2	1984	1996	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>
Delete Time interval 3	1996	2000	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
Delete Time interval 4	2000	2005	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
Delete Time interval 5	2005	2012	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
Add					

## Example IV

Presumed shape



The diagnosis probability is zero from 1980 until 1984,  $o_2$  from 1984 until 1996,  $o_3$  from 1996 until 2000. In 2000, the diagnosis probability is  $o_4$  and then changes linearly. The rate at which the probability increases with calendar changes in 2005 from  $r_4$  to  $r_5$ .

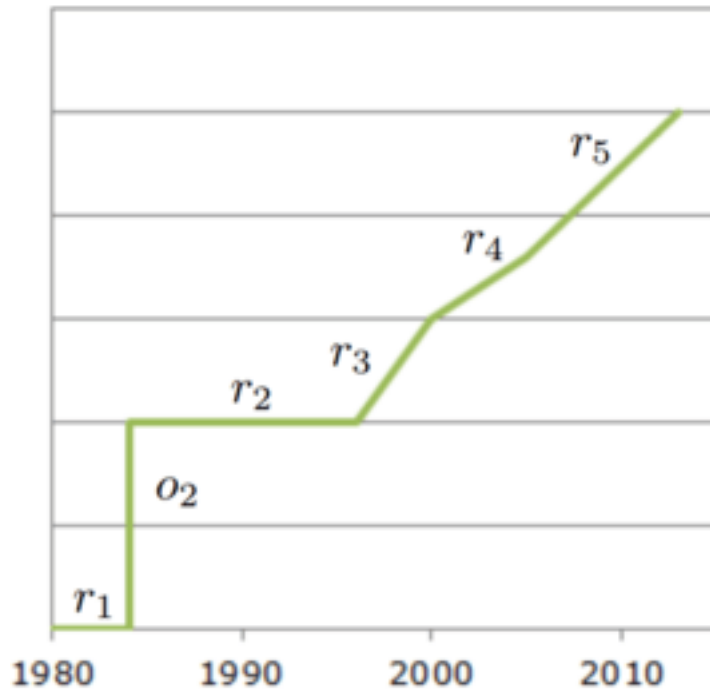
Time intervals and diagnosis probabilities

INCIDENCE METHOD

Time Intervals and Diagnosis Rate Modelling	Description	Start Year	End Year	Start from new baseline	Different by CD4 count categories	Changing during time interval
Delete	Time interval 1	1980	1984	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Delete	Time interval 2	1984	1996	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Delete	Time interval 3	1996	2000	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Delete	Time interval 4	2000	2005	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>
Delete	Time interval 5	2005	2012	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>
Add						

## Example V

Presumed shape



Same as in Example III. The diagnosis probability is zero from 1980 until 1984. Between 1984 and 1995, the diagnosis probability is a constant value  $o_2$ . After 1996, the probability increases linearly, but the rate at which the probability increase with calendar changes in 2000 from  $r_3$  to  $r_4$  and in 2005 from  $r_4$  to  $r_5$ . Diagnosis rates are the same for all CD4 count intervals.

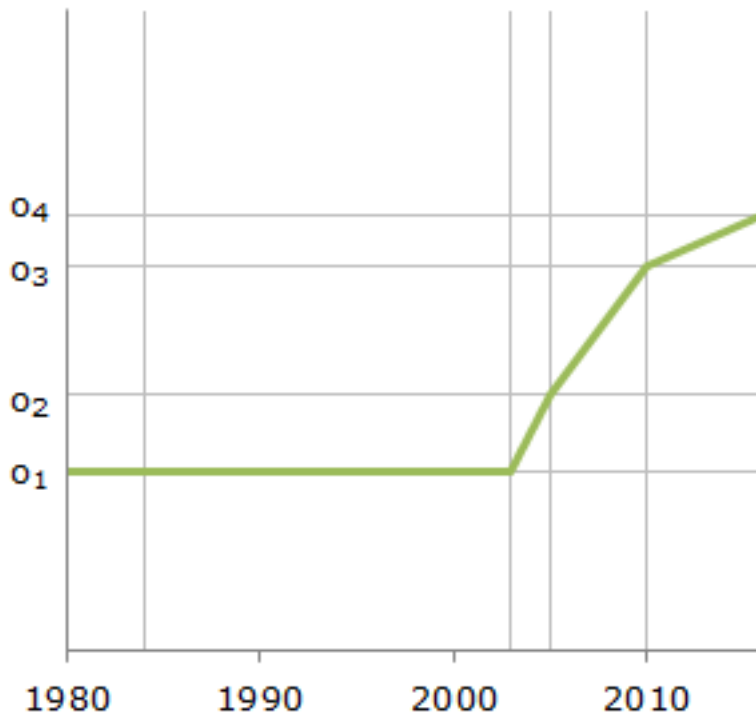
Time intervals and diagnosis probabilities

INCIDENCE METHOD

Time Intervals and Diagnosis Rate Modelling	Description	Start Year	End Year	Start from new baseline	Different by CD4 count categories	Changing during time interval
Delete	Time interval 1	1980	1984	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Delete	Time interval 2	1984	1996	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Delete	Time interval 3	1996	2000	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>
Delete	Time interval 4	2000	2005	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>
Delete	Time interval 5	2005	2012	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>
Add						

## Example VI

*Presumed shape*



This is the presumed shape used in Van Sighem et al.<sup>4</sup> where the diagnosis probability is  $o_1$  between 1980 and 2003. After 2003, the probability increases linearly at a rate that changes in 2005 and 2010. Diagnosis rates can be different for each of the four CD4 count categories.

*Time intervals and diagnosis probabilities*

INCIDENCE METHOD

Time Intervals and Diagnosis Rate Modelling		Description	Start Year	End Year	Start from new baseline	Different by CD4 count categories	Changing during time interval
Delete	Time interval 1	1980	2003	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	
Delete	Time interval 2	2003	2005	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	
Delete	Time interval 3	2005	2010	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	
Delete	Time interval 4	2010	2016	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	
Add							

## 7.3 London Method

**London Method** contains basic parameter settings for the London Method.



## LONDON METHOD

### Calculation based on

- HIV/AIDS diagnoses
- HIV/AIDS diagnoses + HIV diagnoses with HIV-related symptoms

Multiplication factor for rate of HIV symptoms from rate of AIDS

2.000

If the tick box is checked the tool will run the London Method for the selected population. Unchecking the tick box will prevent the tool from running the London Method.

## Intermezzo: London Method

The calculations by the London Method are based on data about HIV/AIDS diagnoses or on data about symptomatic HIV diagnoses, i.e., HIV/AIDS diagnoses plus HIV diagnoses with HIV-related symptoms. For both types of input data, the tool will run the London Method using data from the selected population, highlighted in green in the **population list**. The tool will also do calculations using data from combination **All** and then multiply the result with the proportion of HIV/AIDS diagnoses observed in the highlighted population. This latter method is less sensitive to fluctuations as a result of small number of observations in CD4 count intervals.

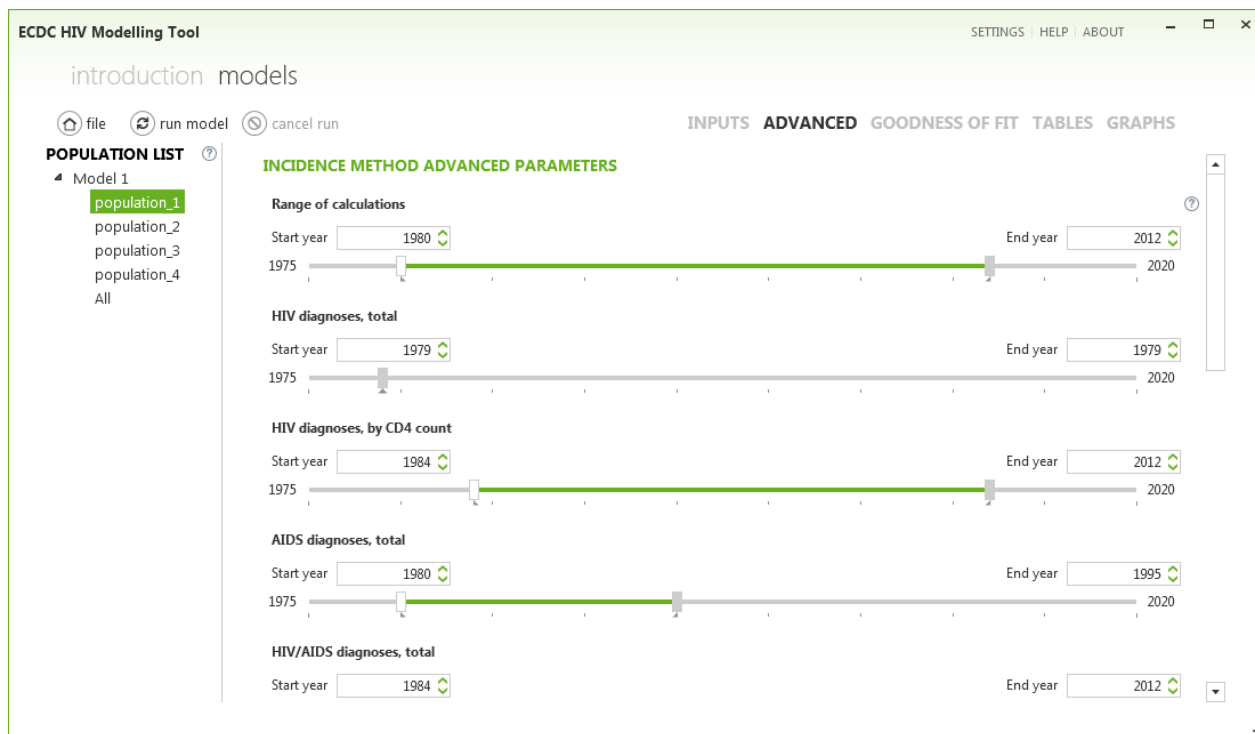
### 7.3.1 Calculation based on

A green tick mark indicates whether the London Method will be based on HIV/AIDS diagnoses only and/or on HIV/AIDS diagnoses plus HIV diagnoses with HIV-related symptoms.

When calculations are based on HIV/AIDS diagnoses plus HIV diagnoses with HIV-related symptoms, the London Method requires knowledge of the CD4 count-specific rate of HIV-related symptoms. While the CD4 count-specific rate of occurrence of AIDS is known, the CD4 count-specific rate of occurrence of such HIV-related symptoms is less well described. There is evidence to suggest that the rate of developing HIV-related symptoms is approximately two- to four-fold higher compared to the rate of AIDS<sup>2</sup>. By default, the CD4 count-specific rates of HIV-related symptoms are assumed to be two-fold the CD4 count-specific rates of AIDS, but this can be changed by the user.

## 7.4 Incidence Method advanced parameters

The **advanced** tab displays shows advanced parameters settings for the selected population in the **population list**. Parameters in this window are pre-specified by the wizard.



### 7.4.1 Calendar year ranges

The five slider bars show the range of calendar years used in the calculations of the Incidence Method. These ranges are specified by the wizard and are based on the minimum and maximum calendar year in the input datasets.

Ranges can be made wider or narrower by dragging the rectangles at either end of the green bar. Alternatively, the lower boundary (**Start year**) and upper boundary (**End year**) of each calendar year range can be selected by clicking the up or down button in the white boxes above the slider bars.

#### Range of calculations

Specifies the range of the model calculations. Calculations start on 1 January of **Start year** and end on 31 December of **End year**.

**Start year** should be the approximate year in which the HIV epidemic started in a country (default value 1980).

Note that it is **not** possible to make future projections with the tool. Therefore, **End year** should not be larger than the most recent calendar year for which data are available.

#### HIV diagnoses, total

Range of calendar years for which data on the total annual number of HIV diagnoses are used in the model fit. This range should correspond to years for which there is no or insufficient information on CD4 counts at the time of diagnosis.

## HIV diagnoses, by CD4 count

Range of calendar years for which data on the annual number of HIV diagnoses by CD4 count interval are used by the Incidence Method. For years with no or insufficient data on CD4 counts at the time of diagnosis (see below), total number of diagnoses should be used (see [HIV diagnoses, total](#))

## AIDS diagnoses, total

Range of calendar years for which data on the annual number of AIDS diagnoses are used. The upper boundary of this range should be earlier than the year in which combination antiretroviral treatment (cART) became widely available.

After the introduction of cART, the annual total number of AIDS diagnoses will strongly depend on how many individuals are treated. The effect of treatment on the time to developing AIDS is difficult to quantify. Therefore, and also because treatment is not taken into account in the tool, total number of AIDS diagnoses should not be used in the era of cART.

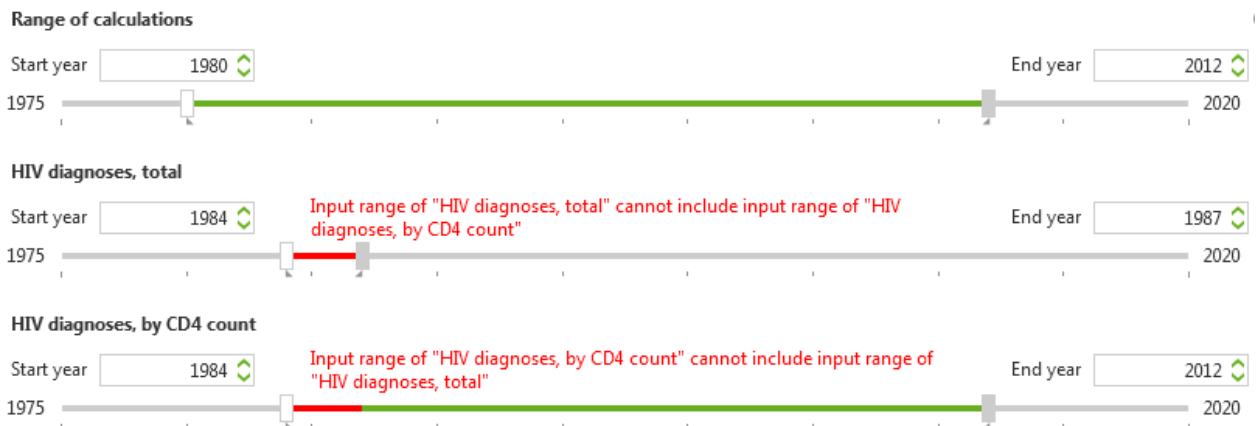
## HIV/AIDS diagnoses, total

Range of calendar years for which data are available on the number of HIV diagnoses with a concurrent AIDS diagnosis.

Since HIV diagnosis generally precedes start of treatment, HIV/AIDS diagnoses can be used during the entire course of the epidemic. There is no limitation on the upper boundary as there is for total number of AIDS diagnoses (see [AIDS diagnoses, total](#)).

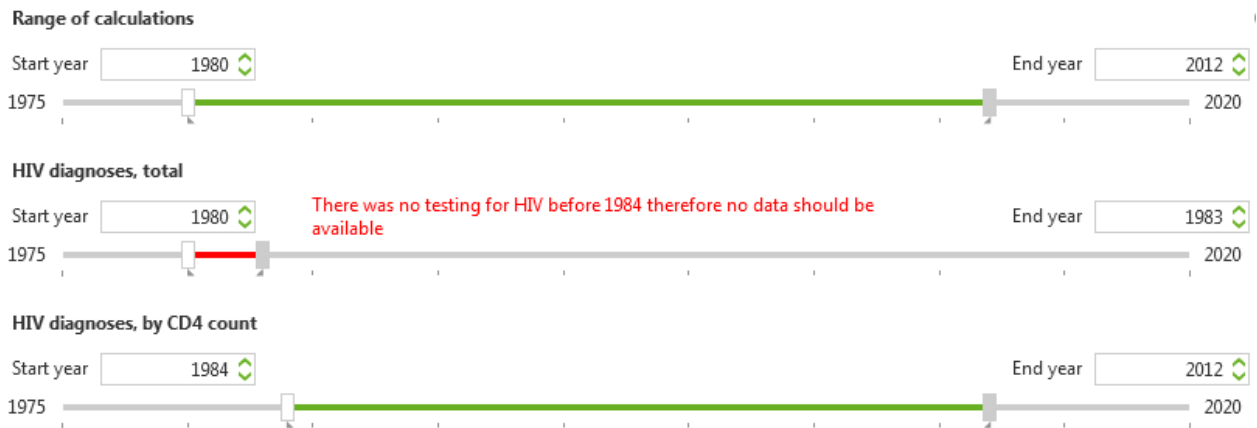
### Notes

- The calendar year ranges for [HIV diagnoses, total](#) and [HIV diagnoses, by CD4 counts](#) *should not overlap*. The tool will issue a warning when the ranges overlap.



The tool automatically calculates the yearly proportion of HIV diagnoses with and without a CD4 count and uses this proportion in the model fit. Therefore, the two data items are not independent and cannot be used simultaneously in the model fit.

- The tool will also issue a warning when specifying a calendar year range that includes years before 1984 for [HIV diagnoses, total](#) or [HIV diagnoses, by CD4 count](#).



- Data items in calendar years outside **Range of calculations** will not be taken into account in the model fit. A data item can be omitted from the model fit by setting both **Start year** and **End year** before **Range of calculations**.
- For calendar years for which no data are provided for one or more data items the tool assumes that the number of observations is 0. This assumption may not be correct if the data item only started to be collected from a specific calendar year onwards. Users are advised to be cautious when extending calendar ranges beyond the pre-specified setting.

## 7.4.2 Full/partial data

If (complete) HIV surveillance data are not available from the start of the HIV epidemic but only from a certain year  $Y$  onwards, **no** should be selected here. In this case it is still possible to estimate the number of newly acquired HIV infections and the undiagnosed population. However, these estimates will only be reliable for the most recent calendar years, i.e., for years where the undiagnosed proportion of people who acquired their HIV infection before year  $Y$  is small. Since data before year  $Y$  are not available, the tool cannot estimate the total population living with HIV.

Please note that model outcomes for recent calendar years may be similar if **yes** is selected despite data not being available from the start of the HIV epidemic. However, the estimation process will be less efficient, especially if a large number of parameters needs to be estimated or if confidence intervals are calculated.


The following settings are recommended if data are only available from year  $Y$  onwards:

- **Start year** of **Range of calculations** is the approximate year in which the HIV epidemic started in a country (default value 1980). **Stop year** is the most recent year for which data are available.
- The slider bars for the data items should start in the year from which surveillance data are available. In the figure below, data on HIV diagnoses, stratified by the presence of a concurrent AIDS event or CD4 cell count, are available from 2003 onwards.

## INCIDENCE METHOD ADVANCED PARAMETERS


**Range of calculations**

Start year  End year

1975 


**HIV diagnoses, total**

Start year  End year

1975 


**HIV diagnoses, by CD4 count**

Start year  End year

1975 


**AIDS diagnoses, total**

Start year  End year

1975 

**HIV/AIDS diagnoses, total**

Start year  End year

1975 

**Full/partial data**

Do you have data from the start of the epidemic?  Yes  No

- In the diagnosis matrix, make sure **Start Year** for the first time interval equals the start year of the range of calculations and **End Year** equals  $Y$ . For this first time interval, select **Start from new baseline** and, if data on CD4 counts are available in year  $Y$ , also **Different by CD4 count categories**, but do not select **Changing during time interval**. Since there are no data before year  $Y$ , it is not possible to estimate diagnosis probabilities in more than one time interval. An example of a diagnosis matrix can be found in [Example VI](#).
- Do not specify too many knots for the incidence curve. The default value of 4 is likely to be enough.

### 7.4.3 Confidence intervals

The tool has the option to determine confidence intervals on estimated parameters and model outcomes via a so-called bootstrap analysis (see below [Intermezzo: bootstrap analysis](#)).

A bootstrap analysis can be time-consuming because it involves running the Incidence Method multiple times on *bootstrap replicates* of the data. Confidence intervals should, therefore, only be determined when the main model gives a satisfactory description of the observed data.

#### Intermezzo: bootstrap analysis

The Incidence Method calculates 95% confidence intervals on estimated parameters and model outcomes by doing a bootstrap analysis. In brief, a bootstrap analysis works as follows. Assuming that the data are distributed according to a certain probability distribution, in this case either a Poisson or a negative binomial distribution, with a mean defined by the best-fitting model, the tool generates a new dataset by sampling from this distribution for every year for each of the relevant data items. The model is then

refitted to this new dataset starting from the parameter values found in the main fit. This procedure of sampling and refitting is repeated many times. From these many fits, 95% confidence intervals around the estimated model parameters and model outcomes can then be determined as the 2.5-th and 97.5-th percentile.

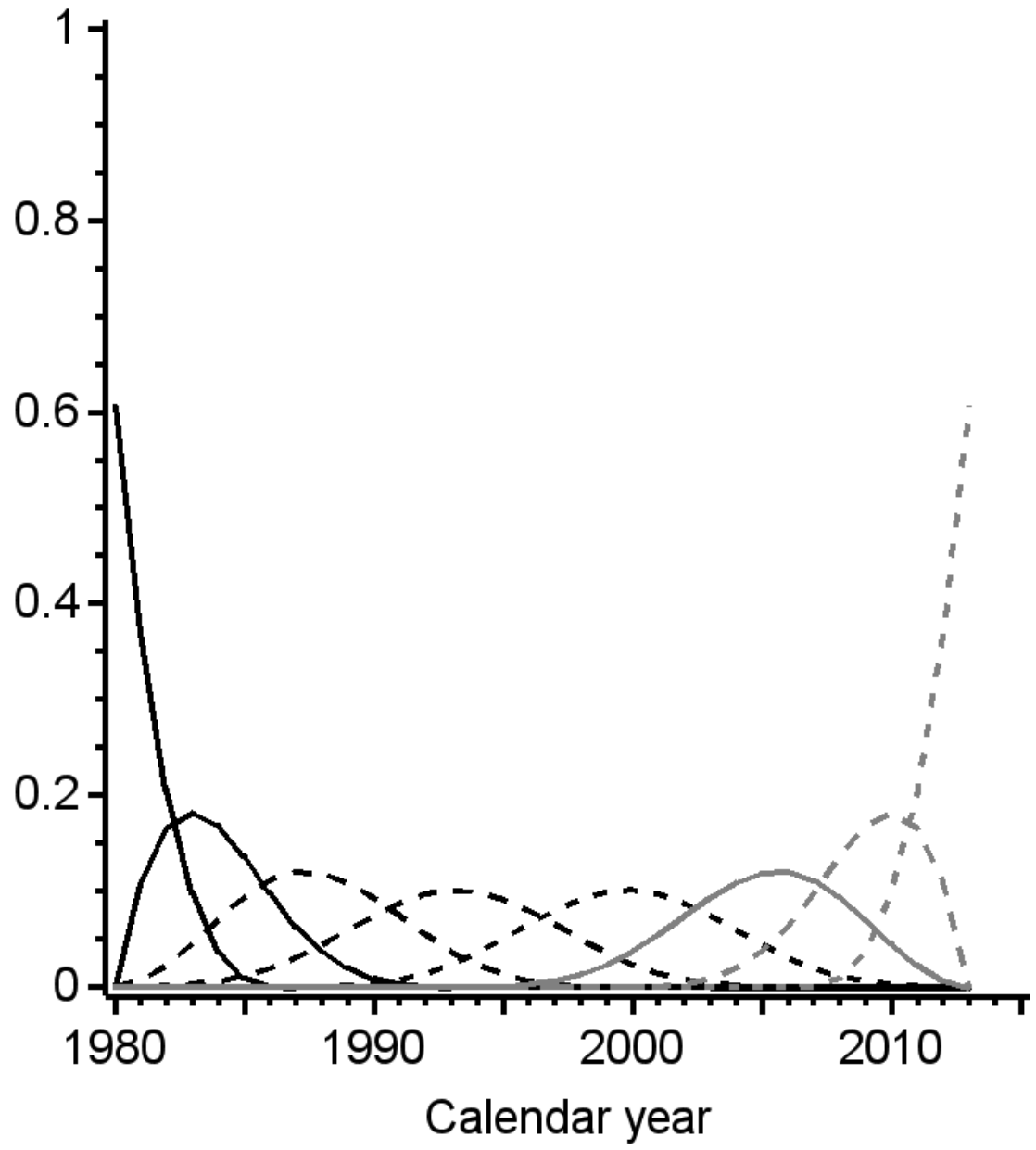
### **Number of iterations**

The user can specify the number of iterations in the bootstrap analysis, ranging from 0 to 1000. A value of 0 means that no bootstrap analysis is done. To get a feeling for the variation in the estimated model fits a value of 20 would suffice. For a full calculation of confidence intervals at least 100 to 200 iterations are recommended.

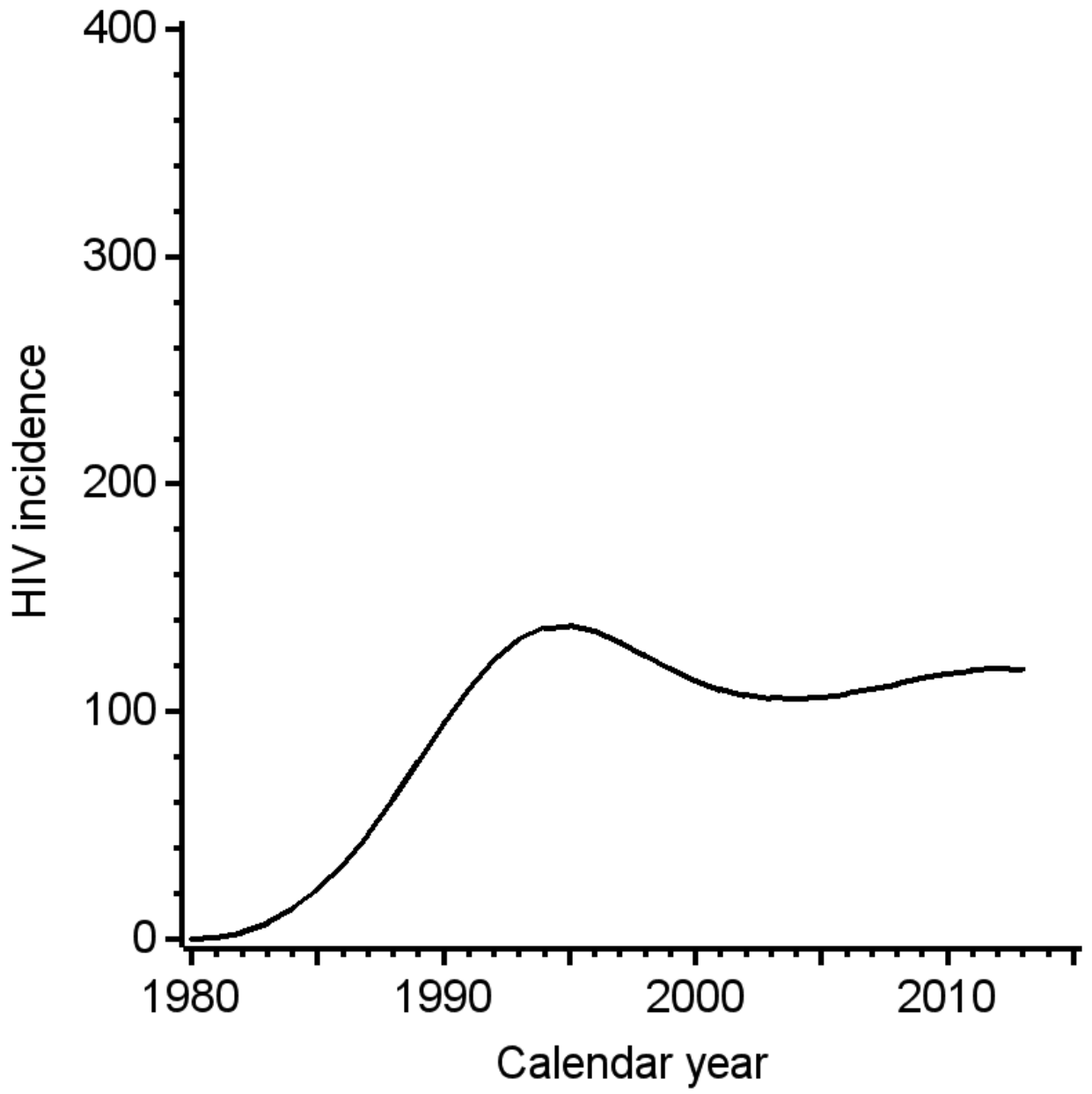
### **7.4.4 Incidence curve**

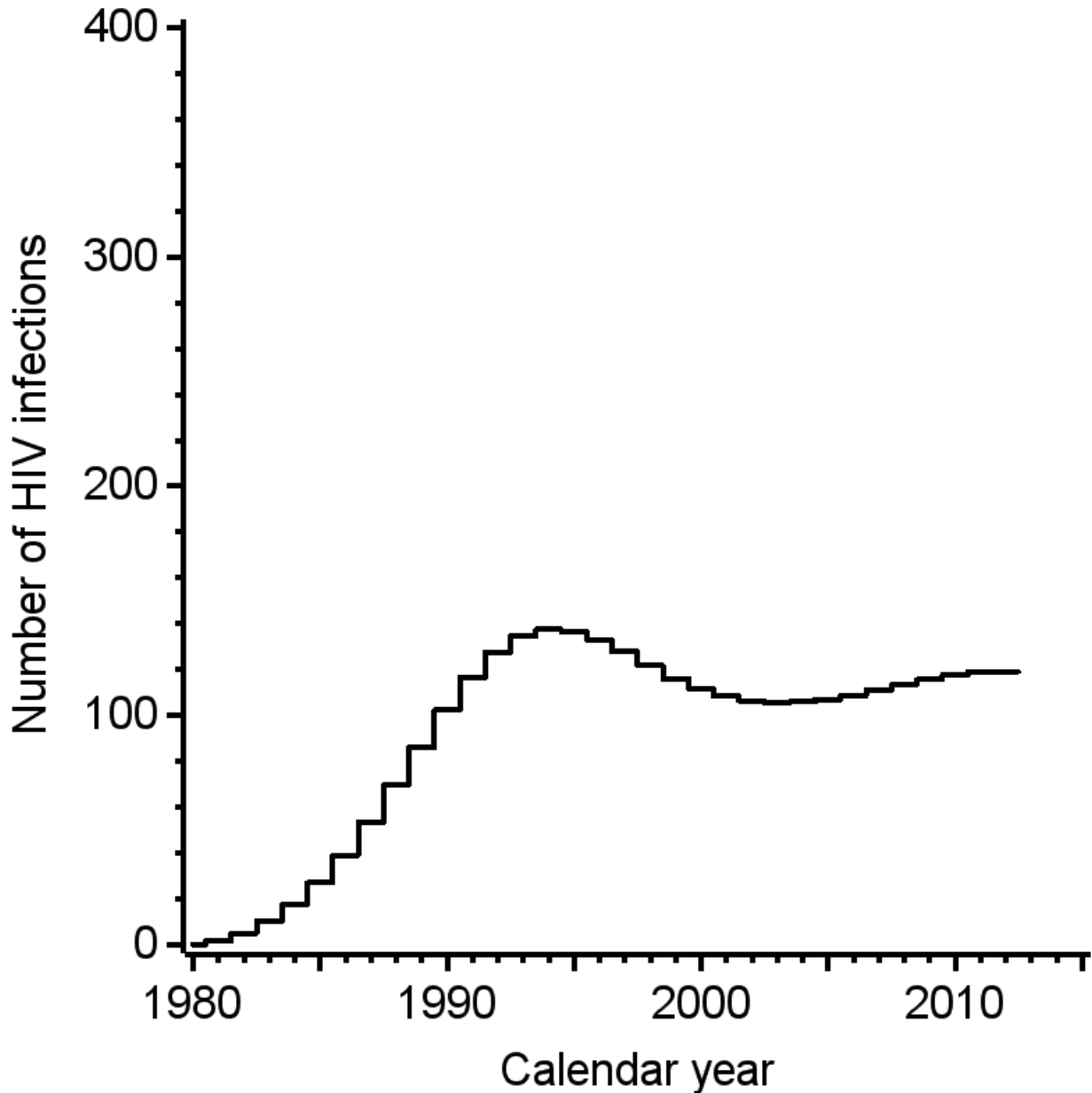
For each calendar year, the Incidence Method determines the annual number of HIV infections from the estimated HIV incidence curve. In the tool, this incidence curve is fully described by a limited number of parameters that is determined by the number of so-called *knots*. A higher number of knots will result in more flexibility for the incidence curve, but also involves estimating a higher number of parameters.

**Intermezzo: HIV incidence**









The annual number of HIV infections (right figure) is determined by the tool from the input data on diagnosed HIV cases. Usually it is not possible to estimate this number separately for each year, because this would involve estimating a too large number of unknown parameters, one for each calendar year. Therefore, the number of infections is derived from an HIV incidence curve (middle figure) that is approximated using so-called *cubic B-splines*.

Cubic splines are polynomial functions of the form  $at^3 + bt^2 + ct + d$  where  $t$  is calendar time and  $a$ ,  $b$ ,  $c$  and  $d$  are constant numbers. By definition, the spline has a non-zero value within a certain time interval that is specified by knots. Outside this time interval the function is zero. In the figure on the left, there are 4 knots equally spaced between 1980 and 2013 and each curve is a different spline.

Adding up all the different splines, each with a different weight, determines the shape of the incidence curve (middle figure). The number of cubic splines needed is always the number of knots plus 4, so 8 splines in this case. For each of the splines the tool needs to estimate a weight parameter, which

determines its contribution to the incidence curve. These weight parameters are estimated from the input data. In case of 4 knots, 8 parameters are necessary, one for each spline. Even though only 8 splines are used, the resulting incidence curve is very flexible. These 8 parameters are much easier to estimate than one parameter for each calendar year.

The tool will do an initial run to determine if there are any splines with a very small weight parameter. Since such splines will not contribute significantly to the incidence curve their weight parameter is then set to zero. If complete surveillance data are not available from the start of the HIV epidemic (see [Full/partial data](#)), the tool will do additional runs to determine the number of weight parameters that can be set to zero before the model fit to the data gets worse.

A property of cubic B-splines is that if all weight parameters are equal, the incidence curve will be a horizontal line. Another property, which is used by the tool to control the incidence curve at the end of the calendar year range, is that straight lines can be obtained by setting  $\theta_i = 2\theta_{i-1} - \theta_{i-2}$ , where  $\theta_{i-2}$ ,  $\theta_{i-1}$ , and  $\theta_i$  are the weight parameters associated with the  $i - 2^{\text{th}}$ ,  $i - 1^{\text{th}}$ , and  $i^{\text{th}}$  spline function.

In earlier versions of the ECDC HIV Modelling Tool (version 1.2.2 and before), there was also the option to use cubic M-splines to approximate the incidence curve. Since M-splines do not have a clear benefit over B-splines this option has been discontinued as of version 1.3.0.

### **Knots count**

This is the number of knots used in the incidence curve (default value 4).

We recommend using 4 to 6 knots. Models with less internal knots are preferred if the fit to the data does not become worse.

### **Start at zero**

If this box is ticked the tool assumes that the HIV incidence curve is zero on January 1st of the year in which the model calculations start (**Start year of Range of calculations**). This means that one of the parameters necessary to specify the HIV incidence curve is fixed at zero and does not need to be estimated.

### **Prevent sudden changes at end of observation interval**

If this box is ticked the tool will prevent sudden increases or decreases in the estimated HIV incidence curve at the end of the calendar year range.

At the end of the calendar year range, the incidence curve is only constrained by individuals who have been diagnosed relatively shortly after becoming infected. The majority of these individuals will have been diagnosed with a CD4 count  $\geq 500$  cells/mm<sup>3</sup> and fluctuations in this number will have a large impact on the behaviour of the incidence curve. The impact of such fluctuations can be attenuated by requiring that sudden increases or decreases in the incidence curve should be prevented. Generally, this does not give a worse fit to the data. However, confidence intervals will become narrower and the estimated incidence curve may appear more precise than it really is.

## **7.4.5 Maximum likelihood**

Maximum likelihood methods are used to find the set of parameters that best fit the observed data. To define the likelihood, it is assumed that all data items are distributed according to a certain probability density function around a mean defined by the model. For convenience, instead of maximising the likelihood, the tool minimises the equivalent deviance measure.

## Distribution

The default distribution is a Poisson distribution in which the mean is equal to the variance. In practice this distribution works well enough.

The other option is a negative binomial distribution, which is a generalisation of the Poisson distribution such that the variance can be larger than in a Poisson distribution. The increase in variance is determined by a so-called *dispersion parameter*. If the negative binomial distribution is selected, the main model fit will be identical to the one obtained with the Poisson distribution. In addition, the tool will estimate two dispersion parameters, one for AIDS cases and one for HIV diagnoses.

### 7.4.6 Diagnosis rate

#### Extra rate due to non-AIDS symptoms

When this value is larger than zero, the tool will add an extra contribution to the probability of being diagnosed when (unobserved) CD4 counts are below 200 cells/mm<sup>3</sup>. This contribution takes into account that HIV-positive individuals may be diagnosed because of HIV-related non-AIDS symptoms. The publication on the Incidence Method uses a value of 0.4 per year<sup>1</sup>. Note that the extra rate will be added to the parameter  $\delta$  (see [Intermezzo: diagnosis probability](#)) for CD4 counts below 200 cells/mm<sup>3</sup>.

Setting this contribution to a non-zero value is particularly useful if there are no CD4 counts available and/or if for one or more time interval the box **Different by CD4 count categories** in the diagnosis probability matrix in the **inputs** tab is unchecked. In this case, diagnosis probabilities will be the same for the three highest CD4 categories and higher for CD4 counts below 200 cells/mm<sup>3</sup>.

If CD4 counts are used in the fit and all boxes **Different by CD4 count categories** are checked (except for the interval 1980-1984), the specified value of the extra rate will not have a large impact. This is because the extra contribution to the diagnosis probability is almost entirely be offset by the estimated probabilities.

### 7.4.7 Country-specific settings

The tool has the possibility to incorporate county-specific settings.

Currently, this option is only implemented for the Netherlands where a correction needs to be done for truncation of HIV diagnoses before 1996. National surveillance of HIV diagnoses only started in 1996 and patients diagnosed before that time were recorded if they survived up to 1996.


## 7.5 London Method advanced parameters

This section contains advanced parameter settings for the London Method.

### LONDON METHOD ADVANCED PARAMETERS

Range of calculations

Start year  End year



Bootstrap

Number of iterations

## 7.5.1 Calendar year ranges

The slider bar shows the range of calendar years used in the calculations of the London Method.

The range can be made wider or narrower by dragging the rectangles at either end of the green bar. Alternatively, the lower boundary (**Start year**) and upper boundary (**End year**) of each calendar year range can be selected by clicking the up or down button in the white boxes above the slider bars.

### Range of calculations

Specifies the range of the model calculations. Calculations are done for all the years from **Start year** to **End year**.

## 7.5.2 95% confidence intervals

The tool has the option to determine 95% confidence intervals using a simple simulation method<sup>2</sup>. These confidence intervals include two sources of uncertainty for the estimated number of people living with undiagnosed HIV: the stochastic uncertainty concerning the CD4 count-specific rate of symptoms and the stochastic uncertainty associated with the possibility that the observed number of HIV diagnoses with HIV-related symptoms may not correspond to the expected number based on the CD4 count-specific rate of symptoms.

### Number of iterations

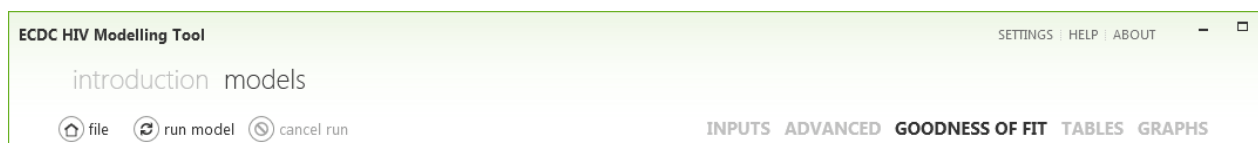
The user can specify the number of iterations in the simulation method, ranging from 0 to 100,000 (default 50,000). A value of 0 means that confidence intervals are not determined. Note that calculation of 95% confidence intervals for the London Method is much faster than for the Incidence Method and can, therefore, always be done.

# 8 Outputs

## 8.1 Goodness-of-fit

In the **goodness of fit** window you can find:

- Tables and graphs showing input data and model fits for the Incidence Method (panels to **A** to **G**).
- Summary statistics of goodness-of-fit for the Incidence Method (section **H**).



Each of the panels labelled to shows a table with the following columns

- **year**: calendar year.
- **N<sub>xxx</sub><sub>D</sub>**: observed number of patients in the input data.
- **N<sub>xxx</sub><sub>Obs</sub><sub>M</sub>**: 'observed' number of patients according to the best-fitting model.

The names of the columns are the same as those used in the output CSV files.

On the right of each panel is a graph showing

- observed number of patients in the input data (green line).
- best-fitting model (dashed grey line).
- 95% confidence intervals (grey band, labelled **Min-max**), only available when **Number of iterations** under **Bootstrap** in the **advanced** window is set to a value larger than zero during running of the Incidence Method.

Results are shown for the selected population in the **population list**.

### A. HIV diagnoses, total

Total annual number of HIV diagnoses. This number will be the total of the four CD4 count categories (**B**, **C**, **D**, and **E**), the number of HIV/AIDS cases (**G**), and the number of HIV diagnoses without AIDS and no CD4 count available.

### B. HIV diagnoses, CD4 $\geq$ 500

Annual number of HIV diagnoses with CD4 counts  $\geq$  500 cells/mm<sup>3</sup>.

### C. HIV diagnoses, CD4 350-499

Annual number of HIV diagnoses with CD4 counts 350-499 cells/mm<sup>3</sup>.

### D. HIV diagnoses, CD4 200-349

Annual number of HIV diagnoses with CD4 counts 200-349 cells/mm<sup>3</sup>.

### E. HIV diagnoses, CD4 < 200

Annual number of HIV diagnoses with CD4 counts <200 cells/mm<sup>3</sup>.

### F. HIV/AIDS diagnoses

Annual number of HIV/AIDS diagnoses (HIV diagnoses with a concurrent AIDS diagnosis).

### G. AIDS diagnoses, total

Total annual number of AIDS diagnoses. The model prediction is only shown for calendar years up to **End year** as specified in the slider bar for **AIDS diagnoses, total**. Data on total number of AIDS diagnoses after **End year** are not used in the model fit, because they are likely to be effected antiretroviral treatment.

### H. Goodness-of-fit statistics

This table shows goodness-of-fit statistics in terms of the deviance of the best-fitting model:

- **Data item:** data item for which the deviance is calculated.
  - **HIV diagnoses, total:** total number of HIV diagnoses
  - **HIV diagnoses, CD4  $\geq$ 500:** HIV diagnoses with CD4 counts  $\geq$  500 cells/mm<sup>3</sup>.
  - **HIV diagnoses, CD4 350-499:** HIV diagnoses with CD4 counts 350-499 cells/mm<sup>3</sup>.

- **HIV diagnoses, CD4 200-349**: HIV diagnoses with CD4 counts 200-349 cells/mm<sup>3</sup>.
  - **HIV diagnoses, CD4 <200**: HIV diagnoses with CD4 counts <200 cells/mm<sup>3</sup>.
  - **HIV/AIDS diagnoses**: HIV diagnoses with a concurrent AIDS diagnosis.
  - **AIDS diagnoses, total**: total number of AIDS diagnoses.
  - **Total**: all data items together.
- **Deviance**: value of the deviance function.
  - **Observations**: number of observations used in the model fit.

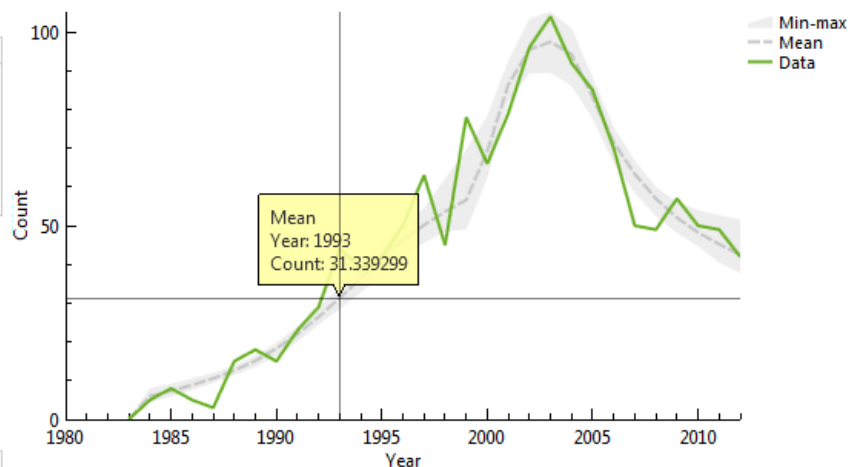
The best-fitting model is the model that minimises the sum of the deviances for all data items. As a rule of thumb a model gives an adequate fit to the data if the deviance is approximately equal to the number of observations.

### Displaying values

By clicking with the left mouse button near or on top of the lines in the graph will show the value in the nearest calendar year.

E. HIV diagnoses, CD4 < 200

year	N_CD4_4_D	N_CD4_4_Obs_M
1980	0	0.00
1981	0	0.00
1982	0	0.00
1983	0	0.00
1984	5	6.06
1985	8	7.35
1986	5	8.87
1987	3	10.59
1988	15	12.63
1989	18	15.16
1990	15	18.29
1991	23	22.07
1992	29	26.46



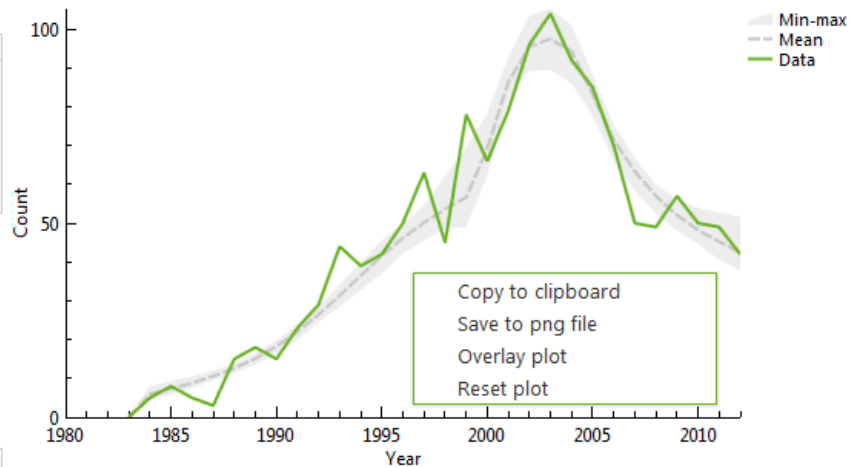
### Copying, saving, and overlaying plots

By clicking with the right mouse button in a graph a menu will be opened:

- **Copy to clipboard**: This will copy the graph to the clipboard so that it can be pasted in e.g. Microsoft Word.
- **Save to png file**: Save the graph as a PNG file on your computer.
- **Overlay plot**: Overlay the current plot with the corresponding plot (data and best-fitting model) from another population in the **population list**.
- **Reset plot**: Clear the overlay plot and return to the plot for the currently selected population.

E. HIV diagnoses, CD4 < 200

year	N_CD4_4_D	N_CD4_4_Obs_M
1980	0	0.00
1981	0	0.00
1982	0	0.00
1983	0	0.00
1984	5	6.06
1985	8	7.35
1986	5	8.87
1987	3	10.59
1988	15	12.63
1989	18	15.16
1990	15	18.29
1991	23	22.07
1992	29	26.46

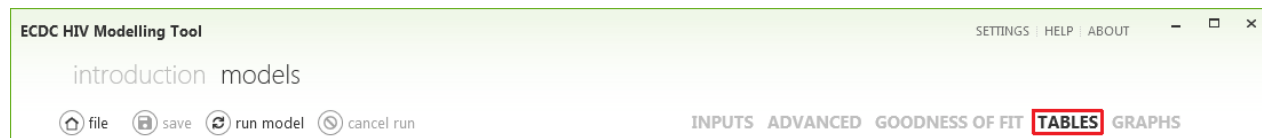


### Notes

- Graphs will include confidence intervals if during running of the Incidence Method **number of iterations** is set to a value larger than zero under **Bootstrap** in the **advanced** window.
- After running the Incidence Method, all graphs will be available as PNG files in the output folder that was specified in **Step 1** of the wizard.

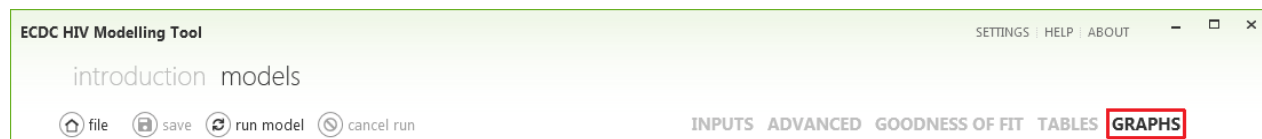
## 8.2 Tables

The **tables** window shows a table with all the data items and model outcomes for the Incidence Method. The names of the columns are explained in the **Appendix**.



## 8.3 Graphs

The **graphs** tab displays outcomes of the Incidence Method and the London Method.



Each of the graphs labelled **A**, **B** etc. shows a table with the following columns

- **year**: calendar year.
- **model outcomes**: estimated model outcome according to the best-fitting model, e.g. annual number of infections or time between infection and diagnosis.

The names of the columns are the same as those used in the output CSV files. A further explanation of the names is given in the **Appendix**.

On the right of each section is a figure showing



- estimated model outcome according to the best-fitting model (line).
- 95% confidence interval (band, labelled **Min-max**).

Results are shown for the selected population in the **population list**.

Graphs can be copied and saved as explained [here](#).

### 8.3.1 Incidence Method

The following graphs are available for the Incidence Method:

#### A. HIV infections per year

Estimated number of HIV infections in each calendar year.

#### B. Time to diagnosis, by year of infection

Estimated time between infection and diagnosis by year of infection if diagnosis probabilities would remain the same as in the year of infection. By default this graph shows the average time to diagnosis. Using the right mouse button the user can select to show the median and interquartile range instead (**show->Median with quartiles**).

#### C. Time to diagnosis, by year of diagnosis

Estimated average time between infection and diagnosis by year of diagnosis. This is the average duration patients have been infected by the time they are diagnosed.

#### D. Total number living with HIV

Estimated number of individuals living with HIV by the end of each calendar year. The three lines include the total number living with HIV (green), the number of diagnosed individuals living with HIV (grey), and the number living with undiagnosed HIV (blue).

Please note that if **Full/partial data** is set to **no**, the graph will only show the number of individuals living with undiagnosed HIV.

#### E. Proportion undiagnosed of all those alive

Percentage of individuals with undiagnosed HIV among those living with HIV. This percentage is equal to the ratio of the blue and the green line in graph.

Please note that if data are not available from the start of the HIV epidemic onwards, it is not possible to determine the total number of people living with HIV. Therefore, this graph is not shown if **Full/partial data** set to **no**.

### 8.3.2 London Method

The following graphs are available for the London Method:

#### A. Number of undiagnosed, CD4 <200

Estimated number of individuals living with undiagnosed HIV and a CD4 count <200 cells/mm<sup>3</sup>.

## B. Number of undiagnosed, CD4 <350

Estimated number of individuals living with undiagnosed HIV and a CD4 count <350 cells/mm<sup>3</sup>.

## 8.4 CSV and Excel files

Results of the Incidence Method and the London Method are written to CSV and Excel files. These files are stored in the output folder that was specified in Step 1 of the wizard.

For each population in the `population list` a separate set of CSV and Excel files are created.

A detailed description of the variables in each file is given in the Appendix.

### Incidence Method

The following CSV and Excel files are created for the Incidence Method:

- **name\_Result\_main**: data and model outcomes for the best-fitting model.
- **name\_Result\_main\_ConfIntervals**: estimates of the confidence intervals for a selection of model outcomes.
- **name\_Result\_BS**: data and model outcomes for each bootstrap run.
- **name\_Param\_BS**: internal and estimated parameters and goodness-of-fit statistics for the best-fitting model and each bootstrap run.

Here, *name* is the name of the selected population in the `population list`.

### London Method

The following CSV and Excel files are created for the London Method:

- **name\_Result\_LM\_1**: estimates of the number living with undiagnosed HIV and CD4 count below 200 or below 350 cells/mm<sup>3</sup>, based on HIV/AIDS diagnoses.
- **name\_Result\_LM\_2**: estimates of the number living with undiagnosed HIV and CD4 count below 200 or below 350 cells/mm<sup>3</sup>, based on HIV/AIDS diagnoses and HIV diagnoses with HIV-related symptoms.

Here, **name** is the name of the selected population in the `population list`.

## 9 Appendix

### 9.1 `Result_main`

The files `name_Result_main` contain the data, model fits to these data, and model outcomes for the Incidence Method. Model fit and outcomes distinguish between true and observed. True refers to the total number of diagnoses in a calendar year, while observed is the observed number of diagnoses taking into account, for instance, missing CD4 counts at the time of diagnosis.

Items with an asterisk (\*) are the outcomes that are compared with the observed data to find the best-fitting model.

## General

---

**run** ID of model fit (0: main model; 1: bootstrap)

**year** calendar year

### HIV diagnoses, total

---

**N\_HIV\_M** true annual number of HIV diagnoses (model)

**Cum\_HIV\_M** true annual number of HIV diagnoses, cumulative (model)

**N\_HIV\_Obs\_M\*** observed annual number of HIV diagnoses (model)

**N\_HIV\_D** observed annual number of HIV diagnoses (data)

### HIV diagnoses by CD4 count

---

**N\_CD4\_1\_M** true annual number of HIV diagnoses CD4  $\geq$  500 (model)

**N\_CD4\_1\_Obs\_M\_NoW** observed annual number of HIV diagnoses CD4  $\geq$  500 if no missing CD4 counts (model)

**N\_CD4\_1\_Obs\_M\*** observed annual number of HIV diagnoses CD4  $\geq$  500 (model)

**N\_CD4\_1\_D** observed annual number of HIV diagnoses CD4  $\geq$  500 (data)

**N\_CD4\_2\_M** true annual number of HIV diagnoses with CD4 350-499 (model)

**N\_CD4\_2\_Obs\_M\_NoW** observed annual number of HIV diagnoses CD4 350-499 if no missing CD4 counts (model)

**N\_CD4\_2\_Obs\_M\*** observed annual number of HIV diagnoses CD4 350-499 (model)

**N\_CD4\_2\_D** observed annual number of HIV diagnoses CD4 350-499 (data)

**N\_CD4\_3\_M** true annual number of HIV diagnoses with CD4 200-349 (model)

**N\_CD4\_3\_Obs\_M\_NoW** observed annual number of HIV diagnoses CD4 200-349 if no missing CD4 counts (model)

**N\_CD4\_3\_Obs\_M\*** observed annual number of HIV diagnoses CD4 200-349 (model)

**N\_CD4\_3\_D** observed annual number of HIV diagnoses CD4 200-349 (data)

**N\_CD4\_4\_M** true annual number of HIV diagnoses with CD4  $<$ 200 (model)

**N\_CD4\_4\_Obs\_M\_NoW** observed annual number of HIV diagnoses CD4  $<$ 200 if no missing CD4 counts (model)

**N\_CD4\_4\_Obs\_M\*** observed annual number of HIV diagnoses CD4  $<$ 200 (model)

**N\_CD4\_4\_D** observed annual number of HIV diagnoses CD4  $<$ 200 (data)

### AIDS diagnoses, total

---

**N\_HIVAIDS\_M** true annual number of HIV/AIDS diagnoses (model)

**N\_HIVAIDS\_Obs\_M\*** observed annual number of HIV/AIDS diagnoses (model)

**N\_HIVAIDS\_D** observed annual number of HIV/AIDS diagnoses (data)

### AIDS diagnoses, total

---

**N\_AIDS\_M\*** true annual number of AIDS diagnoses (model)

**N\_AIDS\_D** observed annual number of AIDS diagnoses (data)

### Death

---

**N\_Death\_D** observed annual number of deaths among diagnosed individuals (data)

**Cum\_Death\_D** observed annual number of deaths among diagnosed individuals,

	cumulative (data)
<b>N_Diag_Dead_M</b>	observed annual number of deaths among diagnosed individuals (model)
<b>Cum_Diag_Dead_M</b>	observed annual number of deaths among diagnosed individuals, cumulative (model)
<b>N_Und_Dead_D</b>	observed annual number of deaths among undiagnosed individuals (data)
<b>Cum_Und_Dead_D</b>	observed annual number of deaths among undiagnosed individuals, cumulative (data)
<b>N_Und_Dead_M</b>	number of AIDS-related deaths among undiagnosed individuals by the end of the year (model)
<b>Cum_Und_Dead_M</b>	number of AIDS-related deaths among undiagnosed individuals by the end of the year, cumulative (model)
Migration	
<b>N_Emig_D</b>	observed annual number of diagnosed individuals who migrated out of the country (data)
<b>Cum_Emig_D</b>	observed annual number of diagnosed individuals who migrated out of the country, cumulative (data)
HIV infections	
<b>Cum_Inf_M</b>	number of newly-acquired HIV infections by the end of year, cumulative (model)
<b>Cum_Inf_D</b>	observed number of newly-acquired HIV infections by the end of year, cumulative (data)
<b>N_Inf_M</b>	annual number of newly-acquired HIV infections (model)
<b>N_Inf_D</b>	observed annual number of newly-acquired HIV infections (data)
Diagnosis probabilities and time to diagnosis	
<b>delta1</b>	diagnosis rate parameter, CD4 $\geq$ 500
<b>delta2</b>	diagnosis rate parameter, CD4 350-499
<b>delta3</b>	diagnosis rate parameter, CD4 200-349
<b>delta4</b>	diagnosis rate parameter, CD4 $<$ 200
<b>t_diag</b>	estimated average time between infection and diagnosis by year of infection <i>when diagnosis probabilities remain the same as in the year of infection</i>
<b>t_diag_p50</b>	median time between infection and diagnosis by year of infection <i>when diagnosis probabilities remain the same as in the year of infection</i>
<b>t_diag_p25</b>	lower quartile of the distribution of time between infection and diagnosis by year of infection <i>when diagnosis probabilities remain the same as in the year of infection</i>
<b>t_diag_p75</b>	upper quartile of the distribution of time between infection and diagnosis by year of infection <i>when diagnosis probabilities remain the same as in the year of infection</i>
<b>D_Avg_Time</b>	estimated average time between infection and diagnosis by year of diagnosis, i.e. the average duration patients have been infected by the time they are diagnosed
<b>t_diag_i</b>	percentage of individuals diagnosed in the i-th calendar year among those infected in calendar year year

## People living with HIV

---

<b>N_Alive</b>	total number of HIV-positive individuals who are still alive. This number is equal to the total number of infected individuals ( <b>Cum_Inf_M</b> ) minus the observed number of individuals who died ( <b>Cum_Dead_D</b> ) minus the estimated number of individuals who died before being diagnosed with HIV ( <b>Cum_Und_Dead_M</b> ).
<b>N_Alive_Diag_M</b> Undiagnosed population (at the end of the year)	total number of diagnosed HIV-positive individuals who are still alive
<b>N_Und</b>	number of undiagnosed individuals, total
<b>N_Und_1</b>	number of undiagnosed individuals who acquired their HIV infection before year <i>Y</i> , i.e., the earliest year for which (complete) surveillance data are available
<b>N_Und_2</b>	number of undiagnosed individuals who acquired their HIV infection in or after year <i>Y</i>
<b>N_Und_Inf_p</b>	proportion undiagnosed individuals of those ever infected
<b>N_Und_Alive_p</b>	proportion undiagnosed individuals of those still alive
<b>N_Und_PrimaryInf</b>	number of undiagnosed individuals, primary infection
<b>N_Und_CD4_1_M</b>	number of undiagnosed individuals, CD4 $\geq$ 500
<b>N_Und_CD4_2_M</b>	number of undiagnosed individuals, CD4 350-499
<b>N_Und_CD4_3_M</b>	number of undiagnosed individuals, CD4 200-349
<b>N_Und_CD4_4_M</b>	number of undiagnosed individuals, CD4 <200
<b>N_Und_HIVAIDS_M</b>	number of undiagnosed individuals, AIDS
<b>N_Und_500</b>	number of undiagnosed individuals, CD4 $\geq$ 500 or primary infection
<b>N_Und_350</b>	number of undiagnosed individuals, CD4 <350 or AIDS
Undiagnosed population (at the end of year, infected in the same year)	
<b>N_Und_T_1</b>	number of undiagnosed individuals, total
<b>N_Und_T_1_p</b>	proportion of N_Und
<b>N_Und_PrimaryInf_T_1</b>	number of undiagnosed individuals, primary infection
<b>N_Und_CD4_1_T_1</b>	number of undiagnosed individuals, CD4 $\geq$ 500
<b>N_Und_CD4_2_T_1</b>	number of undiagnosed individuals, CD4 350-499
<b>N_Und_CD4_3_T_1</b>	number of undiagnosed individuals, CD4 200-349
<b>N_Und_CD4_4_T_1</b>	number of undiagnosed individuals, CD4 <200
<b>N_Und_CD4_5_T_1</b>	number of undiagnosed individuals, AIDS
Undiagnosed population (at the end of year, infected for 1 to 4 years)	
<b>N_Und_T_2</b>	number of undiagnosed individuals, total
<b>N_Und_T_2_p</b>	number undiagnosed, proportion of N_Und
<b>N_Und_PrimaryInf_T_2</b>	number of undiagnosed individuals, primary infection
<b>N_Und_CD4_1_T_2</b>	number of undiagnosed individuals, CD4 $\geq$ 500
<b>N_Und_CD4_2_T_2</b>	number of undiagnosed individuals, CD4 350-499

<b>N_Und_CD4_3_T_2</b>	number of undiagnosed individuals, CD4 200-349
<b>N_Und_CD4_4_T_2</b>	number of undiagnosed individuals, CD4 <200
<b>N_Und_CD4_5_T_2</b>	number of undiagnosed individuals, AIDS

Undiagnosed population (at the end of year, infected for 5 years or more)

---

<b>N_Und_T_3</b>	number of undiagnosed individuals, total
<b>N_Und_T_3_p</b>	number undiagnosed, proportion of N_Und
<b>N_Und_PrimInf_T_3</b>	number of undiagnosed individuals, primary infection
<b>N_Und_CD4_1_T_3</b>	number of undiagnosed individuals, CD4 $\geq$ 500
<b>N_Und_CD4_2_T_3</b>	number of undiagnosed individuals, CD4 350-499
<b>N_Und_CD4_3_T_3</b>	number of undiagnosed individuals, CD4 200-349
<b>N_Und_CD4_4_T_3</b>	number of undiagnosed individuals, CD4 <200
<b>N_Und_CD4_5_T_3</b>	number of undiagnosed individuals, AIDS

Goodness of fit

---

<b>LL_HIV</b>	deviance function total HIV diagnoses
<b>LL_CD4_1</b>	deviance function CD4 $\geq$ 500
<b>LL_CD4_2</b>	deviance function CD4 350-499
<b>LL_CD4_3</b>	deviance function CD4 200-349
<b>LL_CD4_4</b>	deviance function CD4 <200
<b>LL_AIDS</b>	deviance function AIDS diagnoses
<b>LL_HIVAIDS</b>	deviance function HIV/AIDS diagnoses

ECDC tool

---

<b>Version</b>	version of the ECDC HIV Modelling Tool
<b>Timestamp</b>	date and time the model finished running

## 9.2 Result\_BS

The files **name\_Result\_BS** have the same structure as **name\_Result\_main**. These files contain bootstrap replicates of the data and model outcomes for each bootstrap run.

## 9.3 Result\_main\_ConfIntervals

The files **name\_Result\_main\_ConfIntervals** contain estimates of the confidence intervals for a selection of model outcomes. Confidence intervals are updated after every bootstrap run. Suffixes LB and UB represent the lower and upper boundary of the confidence interval.

## 9.4 Param\_BS

General

---

<b>run</b>	ID of model fit (0: main model; 1 and larger: bootstrap)
<b>runtime</b>	approximate time to complete the run

Internal parameters

---

<b>mu</b>	not used
<b>alphaP</b>	rate of progression from acute to chronic infection

q_1	rate of progression from CD4 $\geq$ 500 to CD4 350-499
q_2	rate of progression from CD4 350-499 to CD4 200-349
q_3	rate of progression from CD4 200-349 to CD4 <200
q_4	rate of progression from CD4 <200 to AIDS
q_5	rate of progression from AIDS to death
q_6	not used
f_1	proportion with CD4 $\geq$ 500 directly after primary infection
f_2	proportion with CD4 350-499 directly after primary infection
f_3	proportion with CD4 200-349 directly after primary infection
f_4	proportion with CD4 <200 directly after primary infection
f_5	proportion with AIDS directly after primary infection
f_6	not used
d4_fac	extra rate due to non-AIDS symptoms

Modifiable and estimated parameters

---

t_i	boundaries of the time intervals for diagnosis probabilities
theta_i	estimated parameters associated with the HIV incidence curve
beta_i	estimated parameters associated with the diagnosis probabilities

Negative binomial distribution

---

r_AIDS	dispersion parameter, AIDS diagnoses
r_AIDSPos	dispersion parameter, HIV/AIDS diagnoses
r_Pos	dispersion parameter, HIV diagnoses, total
r_PosCD4	dispersion parameter, HIV diagnoses, by CD4 count

Number of observations used in the model fit

---

N_LL_HIV	number of observations, HIV diagnoses total
N_LL_CD4_1	number of observations, HIV diagnoses CD4 $\geq$ 500
N_LL_CD4_2	number of observations, HIV diagnoses CD4 350-499
N_LL_CD4_3	number of observations, HIV diagnoses CD4 200-349
N_LL_CD4_4	number of observations, HIV diagnoses CD4 <200
N_LL_HIVAIDS	number of observations, HIV/AIDS diagnoses
N_LL_AIDS	number of observations, AIDS diagnoses

Goodness-of-fit

---

LL_HIV	deviance function, HIV diagnoses total
LL_CD4_1	deviance function, HIV diagnoses CD4 $\geq$ 500
LL_CD4_2	deviance function, HIV diagnoses CD4 350-499
LL_CD4_3	deviance function, HIV diagnoses CD4 200-349
LL_CD4_4	deviance function, HIV diagnoses CD4 <200
LL_HIVAIDS	deviance function, HIV/AIDS diagnoses
LL_AIDS	deviance function, AIDS diagnoses
LL_Smooth_1	not used

<b>LL_Smooth_2</b>	not used
<b>LL_Poisson</b>	deviance function, Poisson distribution
<b>LL_Total</b>	deviance function, total
<b>X2_Total</b>	$\chi^2$ statistic, calculated as <b>LL_Total</b> divided by the number of observations used in the fit minus the number of estimated parameters
<b>X2_Pearson</b>	$\chi^2$ statistic, calculated as the sum of the squared difference of each data point and the corresponding model estimate divided by the variance
<b>X2_Pearson_n</b>	$\chi^2$ statistic, calculated as <b>X2_Pearson</b> divided by the number of observations used in the fit minus the number of estimated parameters

#### ECDC tool

---

<b>Version</b>	version of the ECDC HIV Modelling Tool
<b>Timestamp</b>	date and time the model finished running

## 9.5 Result\_LM

The CSV file **name\_Result\_LM\_1** contains results of the London Method based HIV/AIDS diagnoses. The file **name\_Result\_LM\_2** contains results based on HIV/AIDS diagnoses and HIV diagnoses with HIV-related symptoms.

#### General

---

<b>year</b>	calendar year
Estimates based on selected population only	
<b>w_missing</b>	correction factor for symptomatic diagnoses with missing CD4 count
<b>N_200_D</b>	observed annual number of symptomatic diagnoses CD4 <200 (data)
<b>N_200_M</b>	estimated number of individuals CD4 <200 (model)
<b>N_200_M_lcl</b>	estimated number of individuals CD4 <200, lower bound (model)
<b>N_200_M_ucl</b>	estimated number of individuals CD4 <200, upper bound (model)
<b>N_350_D</b>	observed annual number of symptomatic diagnoses CD4 <350 (data)
<b>N_350_M</b>	estimated number of individuals CD4 <350 (model)
<b>N_350_M_lcl</b>	estimated number of individuals CD4 <350, lower bound (model)
<b>N_350_M_ucl</b>	estimated number of individuals CD4 <350, upper bound (model)
Estimates based on all populations in the model	
<b>w_ALL_missing</b>	correction factor for symptomatic diagnoses with missing CD4 count
<b>N_200_ALL_D</b>	observed annual number of symptomatic diagnoses CD4 <200 (data)
<b>N_200_ALL_M</b>	estimated number of individuals CD4 <200 (model)



<b>N_200_ALL_M_lc1</b>	estimated number of individuals CD4 <200, lower bound (model)
<b>N_200_ALL_M_uc1</b>	estimated number of individuals CD4 <200, upper bound (model)
<b>N_350_ALL_D</b>	observed annual number of symptomatic diagnoses CD4 <350 (data)
<b>N_350_ALL_M</b>	estimated number of individuals CD4 <350 (model)
<b>N_350_ALL_M_lc1</b>	estimated number of individuals CD4 <350, lower bound (model)
<b>N_350_ALL_M_uc1</b>	estimated number of individuals CD4 <350, upper bound (model)

## 10 References

1. van Sighem A, Nakagawa F, De Angelis D, et al. Estimating HIV incidence, time to diagnosis, and the undiagnosed HIV epidemic using routine surveillance data. *Epidemiology*. 2015;26(5):653-660. doi:[10.1097/EDE.0000000000000324](https://doi.org/10.1097/EDE.0000000000000324).
2. Lodwick RK, Nakagawa F, van Sighem A, Sabin CA, Phillips AN. Use of surveillance data on HIV diagnoses with HIV-related symptoms to estimate the number of people living with undiagnosed HIV in need of antiretroviral therapy. *PLoS ONE*. 2015;10(3). doi:[10.1371/journal.pone.0121992](https://doi.org/10.1371/journal.pone.0121992).
3. Centers for Disease Control and Prevention. 1993 revised classification system for HIV infection and expanded surveillance case definition for AIDS among adolescents and adults. *Clinical Infectious Diseases*. 1992;41(RR-17):802-810. doi:[10.1093/clinids/17.4.802](https://doi.org/10.1093/clinids/17.4.802).
4. van Sighem A, Pharris A, Quinten C, Noori T, Amato A. Reduction in undiagnosed HIV infection in the European Union/European Economic Area, 2012 to 2016. *Euro Surveill*. 2017;22(48). doi:[10.2807/1560-7917.ES.2017.22.48.17-00771](https://doi.org/10.2807/1560-7917.ES.2017.22.48.17-00771).