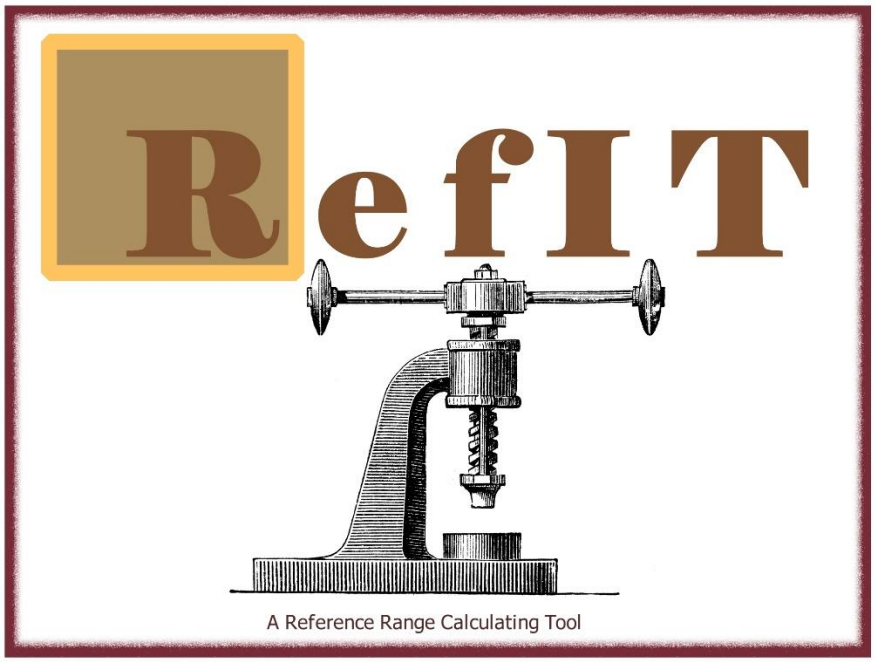
# RefIT

Reference Interval Tool




v. 1.0 documentation

**The GNU General Public License**



For reporting bugs or suggestions for further developments or modifications, please contact: jen7lar@gmail.com.

# Index

About RefIT ..... 5
List of functions ..... 5
1. User interface ..... 6
*Top menu* ..... 6
*File* ..... 6
Edit ..... 6
Help ..... 6
*Action Buttons* ..... 6
Single calc. ..... 6
Batch calc. ..... 7
Set Time Periode ..... 7
Co-medication ..... 7
Clear all calc ..... 7
*Parameters* ..... 7
Assay ID ..... 7
Model ..... 7
Gender ..... 7
Age from/to ..... 7
Low/High percentile ..... 7
Time between sample dates ..... 7
Percentile Compute Method ..... 7
Tukey Outlier Rule ..... 7
Data plot view ..... 7
*Data and Result* ..... 7
*Set Time Period* ..... 8
2. Importing datasets ..... 8
Analyte name, or analysis ID ..... 9
Patient ID ..... 9
Date of analysis ..... 9
Results ..... 9
Gender ..... 9
Age ..... 9
*Anonymization of Danish security numbers* ..... 9

*Automatically cleaning of the datasets* ..... 10
Result ..... 10
Analysis dates ..... 10
Age ..... 10
Gender ..... 10
Patient ID ..... 10
3. Analysis ..... 10
4. Exporting ..... 10
5. Known errors ..... 11

## About RefIT

RefIT – Reference Interval Tool - is a standalone program, designed to help visualize population based laboratory datasets and calculate reference ranges. Although primarily designed for use with therapeutic drugs, it can also be applied on datasets from endogenous substances as well. The program takes input in the form of excel files, which can be obtained from most Laboratory Information Systems (LIMS). It has build-in export function of processed data in the same format, and in addition provides export of graphical views. Any calculations can easily be validated, by opening the export file in Excel.

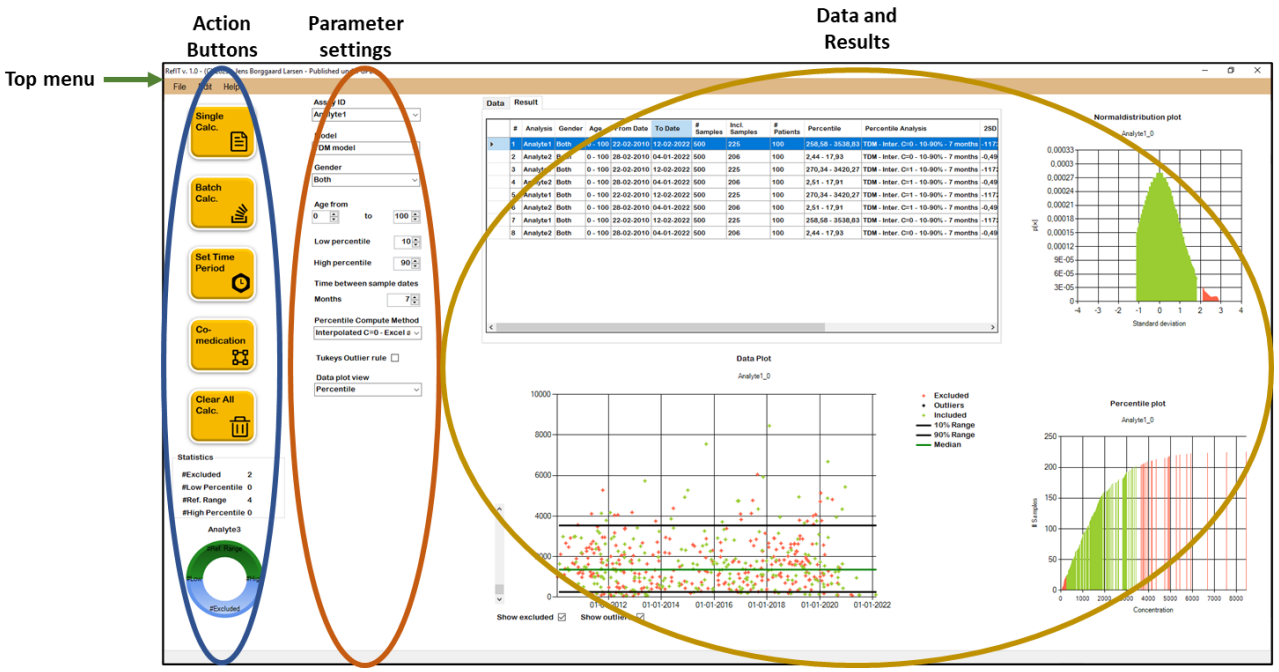
## List of functions

- Batch import and calculation of multiple analytes from a single dataset file.
- Automatically cleans data, and removes results “lower/higher than” on import.
- Automatically verifies date of analysis on import.
- Automatically identifies and anonymize patient ID’s when Danish social security numbers are used.
- Automatically calculates age and gender based on Danish social security number.
- Checks for missing age and gender data when importing.
- Provides four different models for selecting data when calculating reference ranges.
- Provides settings for selection of age group and gender when calculating reference ranges.
- Provides settings for selection of time range of samples.
- Provides selection of percentile range for calculating therapeutic reference ranges, and three modes of calculating the percentile.
- Provides calculation of reference ranges based on -/+2 standard deviations.
- Provides basic statistics like average, median, min/max.
- Provides removal of outliers based on Tukey’s fences.
- Provides visualization of all sample results by data plot, percentile plot and normal distribution plot.
- Export of reports as excel file, or by printing.
- Export of processed data as excel file.
- Export of all graphical views as high resolution images.

# 1. User interface

RefIT is best viewed at a large resolution on a desktop screen, as smaller laptops may have problems with proper sizing. The main user interface is divided into three areas. Action buttons for settings and performing analysis is located in the left side of the screen, along with import/analysis statistics. General parameter settings are located next to it, while data tables and graphs for analysis results are located on the right side of the screen. In addition a top menu provides access to import/export functions and further settings (Figure 1).

**Fig.1**




## *Top menu*

The top menu provides access to *File*, *Edit* and *Help* sub menu.

*File* contains the functions:

- Import : Imports a dataset
- Export selected : Export highlighted analysis as excel file.
- Export all : Export all performed analysis as excel file.
- Print report : Prints the result table (see below).
- Save Charts as image : Saves either of the charts as a an image file for later use.
- Close : Exit the program.

Edit contains a setup function. This is used for debugging the program. If an email address is set, an error report is automatically send to the specified address using outlook. This is especially useful for finding errors in the import file, where wrong datatypes can lead to errors while loading.

Help contains the software About Box.

## *Action Buttons*

Single calc. – Performs calculation on the selected analyte, based on the parameter settings.

Batch calc. – Performs calculations on all the loaded analytes, based on the parameter settings.

Set Time Periode – Provides access to menu for setting the period of time to analyse data from.

Co-medication – Feature to be added in future version.

Clear all calc. – Deletes all previous analysis and resets the parameters. The imported data are maintained in memory, and no reload is necessary.

### *Parameters*

Parameters contain settings for selecting the algorithm, and population to perform the calculations on.

Assay ID – To be used for selection of single analyte to perform calculations on. The selection is active when using the *Single Calc.* action button.

Model – Provides choice of model for selecting data to perform calculations on. Either all samples from each patient, TDM model (see Larsen et al. 2023), First or last sample from each patient.

Gender – Select gender to perform calculation for.

Age from/to – Setting age range to perform calculations on.

Low/High percentile – Select the percentile to calculate.

Time between sample dates – Applies to the TDM model only. Sets the minimum time in month for approval of two samples (see Larsen et al. 2023).

Percentile Compute Method – Allows the selection of algorithm applied for calculating percentiles. For further information see: https://en.wikipedia.org/wiki/Percentile .

Tukey Outlier Rule – Turns on/off Tukey's Outlier rule. This removes outliers from the analysis by calculating Tukey's fences. The program will carry out multiple rounds until zero outliers are each. For more information see : https://en.wikipedia.org/wiki/Outlier .

Data plot view – Allows the selection of the view in the data plot. Either percentiles with median (used for therapeutic drug monitoring) or 2xStandard Deviations with average (used for endogenous substances). The selection has no consequence on the calculations, as both ranges are calculated for each analyte.

### *Data and Result*

The Data and Result view is located to the right of the screen (Figure 1). This area contains visualization of datasets and calculated results. The table contains a view for showing either *Result* or *Data*. The *Result* view shows the analysis report, along with the settings used. For each calculation performed using the action buttons, analysis are added as rows to this table. The *Data* table shows the dataset of the currently selected analysis. By double-clicking on a analysis row in the *Result* tabel, this dataset is loaded into the *data* table and charts.

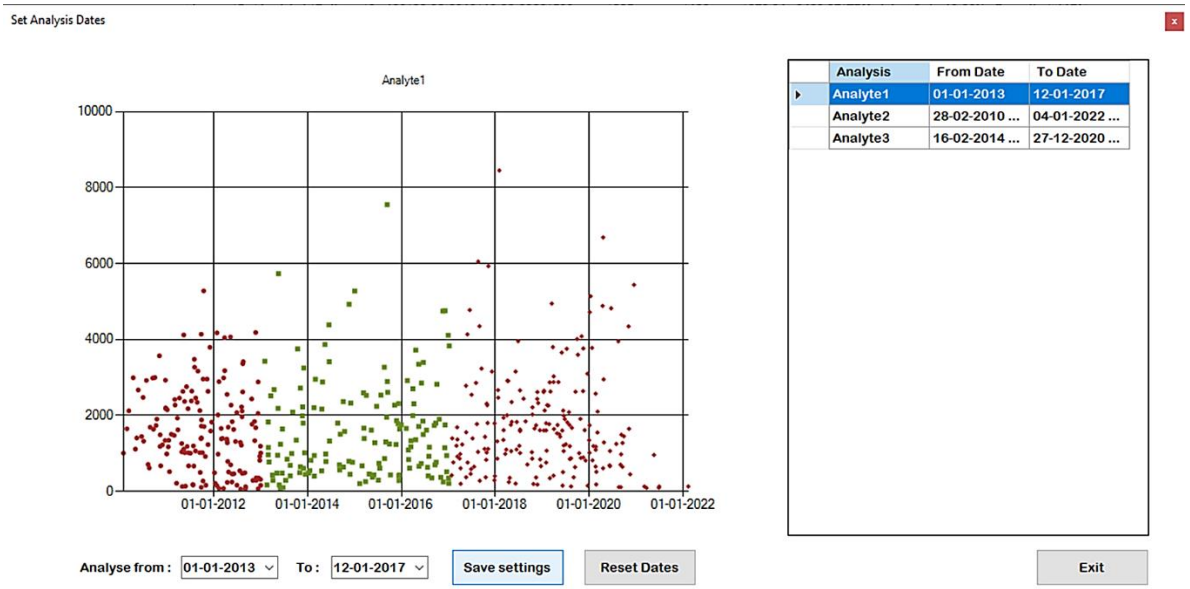
Three charts are used to visualize the data from an analysis. A normal distribution chart, a percentile chart, and a data chart. The normal distribution chart shows the distribution around the average value, and visualize the +/-2SD range primarily used for endogen substances. The percentile chart shows the distribution of number of samples, as a function of the concentrations in the dataset. The Data chart plots the concentration as a function of the date of analysis, and visualizes samples included or excluded. This chart has a slide to the

left for zooming in on the y-axis, while additional selection for showing outliers are available as checkboxes underneath (figure 1).

### *Set Time Period*

The action button *Set time period* provides access to a menu for selecting a specific time period to be analysed from each analyte (figure 2).

**Fig. 2**

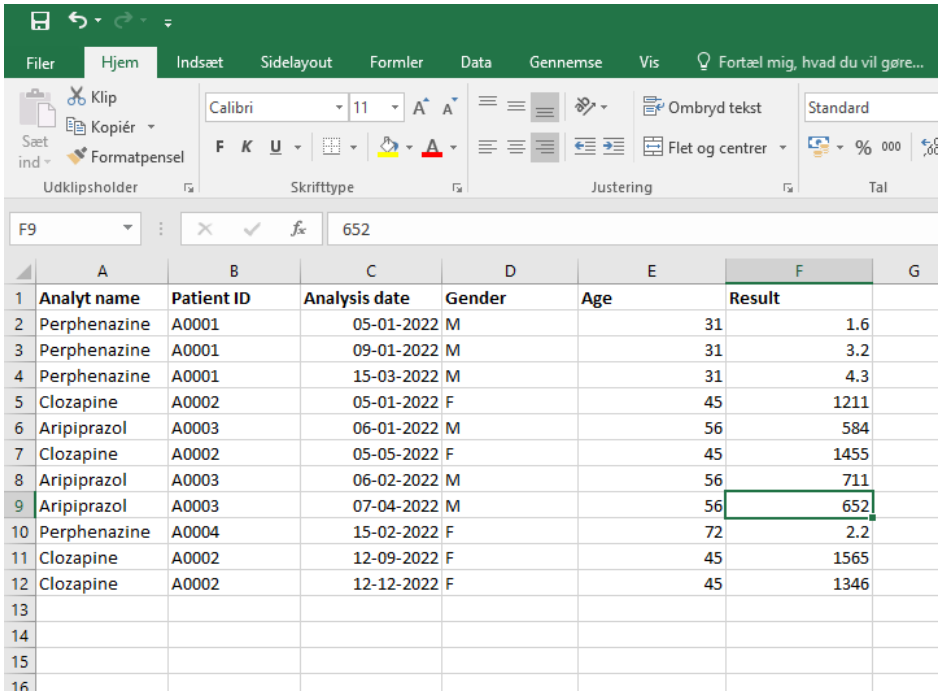



The analyte is selected by clicking on the row in the table to the right. The start and end date of the period is then set using the date settings. By pressing the *Save Settings* button, the dates set for the analyte is saved. The plot updates showing the selected data (period) in green. Pressing the *Reset Dates* removes all changes made to the dates. *Exit* returns to the main menu.

## 2. Importing datasets

RefIT takes input as datafiles in the format .xlsx, .xlsm, .xltx, and .xltm. When importing excel file with multiple workbooks, a prompt will occur for selecting which one data should be loaded from. The data in the import file should be ordered in columns with the first row containing headers (figure 3 – and supplied demo file). Multiple analytes can be loaded on the same time, by placing them in the same column. The datafile should as a minimum include 5 columns containing the data of all analytes, with a single row per result, as in the following example:

**Fig. 3**



| | A | B | C | D | E | F |
|---|---|---|---|---|---|---|
| 1 | **Analyt name** | **Patient ID** | **Analysis date** | **Gender** | **Age** | **Result** |
| 2 | Perphenazine | A0001 | 05-01-2022 | M | 31 | 1.6 |
| 3 | Perphenazine | A0001 | 09-01-2022 | M | 31 | 3.2 |
| 4 | Perphenazine | A0001 | 15-03-2022 | M | 31 | 4.3 |
| 5 | Clozapine | A0002 | 05-01-2022 | F | 45 | 1211 |
| 6 | Aripiprazol | A0003 | 06-01-2022 | M | 56 | 584 |
| 7 | Clozapine | A0002 | 05-05-2022 | F | 45 | 1455 |
| 8 | Aripiprazol | A0003 | 06-02-2022 | M | 56 | 711 |
| 9 | Aripiprazol | A0003 | 07-04-2022 | M | 56 | 652 |
| 10 | Perphenazine | A0004 | 15-02-2022 | F | 72 | 2.2 |
| 11 | Clozapine | A0002 | 12-09-2022 | F | 45 | 1565 |
| 12 | Clozapine | A0002 | 12-12-2022 | F | 45 | 1346 |

The columns are:

Analyte name, or analysis ID if calculating analytes in batches these are put in the same column with one result per row (see example figure 3).

Patient ID A unique number or ID that identifies samples from the same patient.

Date of analysis The date the sample was analyzed or the result approved in LIMS. Present in a recognizable date format.

Results Result can be in any unit, however should be a number without unit suffix (see figure 3).

Gender (optional) As either ‘M’, ‘m’, ‘Male’, ‘male’, ‘Mand’, ‘mand’, ‘F’, ‘f’, ‘K’, ‘k’, ‘Female’, ‘female’, ‘Kvinde’, ‘kvinde’.

Age (optional) An integer of the patients age at analysis time.

Upon loading the datasheet, the software will provide a box for selecting from which column each parameter should be imported. Any excess column will be omitted, and it is not necessary to clean the spreadsheet further or use specified column names. Import statistics is shown in the table and the graph in the lower right corner of the screen (figure 1).

### *Anonymization of Danish security numbers*

On loading a dataset, the program will check if the provided patient ID is a Danish social security number in the form of ddmmyy-xxxx. If that is the case, gender and age is calculated based on this number, and a new anonymized patient ID is assigned to all samples from each patient. The anonymization provides no key for later identification, and the loaded table is deleted in the memory. No change is however made to the original

file while loading. PLEASE NOTE : When importing large datasets (>10.000) anonymization may take a long time. Estimated time for 100.000 samples is anything from 30 min to 1 hour depending on the system.

### *Automatically cleaning of the datasets*

While importing, the program performs a control of the data in each sample row.

Result Any sample with a missing value, or containing a string (e.g. "<", ">") that cannot be converted to a number, will be deleted.

Analysis dates Any sample with a missing or incorrect sample date is deleted.

Age if column or a single age of a patient is missing, "111" will be added instead, to indicate that the value is missing. Hereby these samples are omitted if the age setting is anything but 100 in the program.

Gender If gender is missing or in an incorrect format (see above), it will be substituted with a "N".

Patient ID Only applies when loading a sample set containing Danish social security numbers. If an incorrect number is identified, it will be deleted from the sample set. This includes numbers from Greenland or the Faroe Islands.

## 3. Analysis

Analysis can be performed either on a single analyte, or in batch on all loaded analytes. For single analysis, the analyte is selected from the combobox *Assay ID* (see *Parameter settings* – Figure 1), followed by pressing the action button *Single Calc.* The analysis is then performed based on the parameter settings, and the result is display in the data and result table and charts to the right.

When using the action button *Batch Calc*. the program will cycle though and compute on all analytes in the *Assay ID* combobox, based on the parameter settings. When finished, the tables and charts are updated with the data from the last computed analyte.

The action button *Clear All Calc.* erases all performed analysis from the tables and charts, and resets all parameters. The loaded Raw data will be retained in memory, and are not erased unless another dataset is loaded using the *File* menu.

## 4. Exporting

Data and results can be exported as excel .xlsx files. This can either be done using the function *Export selected* or *Export all* found in the *File* menu. Multiple analysis can be selected by holding down the CTRL key and clicking on the rows in the Result table, hereby marking them in blue. These can then be saved as a single .xlsx file by *Export selected*. In contrast all performed analysis are exported as one excel file when selecting *Export all*.

Each of the three charts can be saved as an image file, by selecting *Save Charts as image* in the *File* menu. When choosing this function, a combobox appears on screen for selecting which chart (Normal distribution, Percentile or Data plot) to save. Following selection the user is asked to type a header for the chart, and to input concentration unit used, after which the filedialog allows saving as either png, jpg, bmp, tiff, emf or wmf file.

# 5. Known errors

- The Statistical chart, at the lower left side of the UI does not update when performing a new analysis.
- The program imports analysis dates as both date and time, while selecting the population in the *Set period time* menu, is based on dates only. To ensure all samples from the “From” and “To” date is included, a day is subtracted from the minimum date and one is added to the maximum date (e.g. instead of the minimum date 12-01-2022 00:00:00 the date 31-11-2022 will be used for selecting the population).
- When calculating percentiles with <10 samples an error may occur, resulting in an error message.