**Instructions on running the R-code for analysing an alternative donor deferral strategy**

By Mart Janssen,Amber Meulenbeld & Mart Pothast ([m.janssen@sanquin.nl](mailto:m.janssen@sanquin.nl), [a.meulenbeld@sanquin.nl](mailto:a.meulenbeld@sanquin.nl), [m.pothast@sanquin.nl](mailto:m.pothast@sanquin.nl))

Version 5, 2025-01-21

**General description**

Generic R-code was developed that allows performing an initial analyses of potential benefits of a mean Hb-level based donor deferral policy that compensates for biological and measurement variability. A more complete discourse of this approach is provided in the paper *“Why the Majority of On-Site Repeat Donor Deferrals Are Completely Unwarranted…”* (Transfusion 2022, 62 (10), 2068–2075. <https://doi.org/10.1111/trf.17085>).

**Data**

The input required to run the code is an .rds file containing 4 variables:

|  |  |  |
| --- | --- | --- |
| **Variable name** | **Variable description** | **Variable data type** |
| KeyID | Unique identifier for each donor | integer |
| Sex | indicator for donor being Male (M) or Female (F) | Factor |
| DonDate | date of donation | Date |
| Hb | donor Hb at donation | number |

If sufficient data is available it is recommended to include only donors for which the full donation history is available. For blood establishments with pre-donation and post-donation screening, we recommend keeping ONLY the post-donation screening result.

**Code parameters**

In the analysis file (*“alt\_def.R”*) a number of parameters need to be specified by the user. Each of these parameters are to be stored in a variable:

1. FILE\_DIR: the directory in which the .rds file is located
2. DATAFILE\_NAME: the name of the .rds file
3. CUTOFF\_M/CUTOFF\_F: Minimum acceptable Hb levels for males and females in the units of measurement also used in the data file (e.g. g/dL, g/L, mmol/L, etc.)
4. UNITS: indicates the unit of measurement in the data file as a string (e.g. g/dL, g/L, mmol/L)
5. TEST: indicates whether a test version of the code is ran (using only 10,000 records). Set this to F for the full analysis.

In case the data are available as a text (or csv) file, there is some separate code (*“Convert CSV data to RDS.R”*) to help convert this data to an R datafile. If you need help with this conversion, contact Mart or Amber.

**Output**

During the analyses, which will be performed within a couple of minutes various input and output parameters will be stored for post-processing. These will only contain information at an aggregated level. No donor IDs will be included.

Various population statistics and an output summary table are stored in the output files called “*confusion\_matrix\_results.csv*” and *“results\_characteristics.csv”*. We also produce histograms of the Hb levels of certain groups of donors, these are saved in the file “*hist\_results.rds*”.

In the output file “*confusion\_matrix\_results.csv*” the following data will be stored:

|  |  |
| --- | --- |
| **Variable name** | **Description** |
| sex | Results are split by sex |
| alpha\_outlier | The alpha\_outlier used for this row of results |
| alpha\_mean | The alpha\_mean used for this row of results |
| type | The type of deferral calculated in this row (def\_OR, def\_AND, def\_mean, def\_outlier) |
| TN | The number of true negatives (deferral = positive) |
| FP | The number of false positives |
| FN | The number of false negatives |
| TP | The number of true positives |
| mean\_Hb\_wrt\_thres | The mean difference between Hb and the threshold for new donations with the alternative algorithm |
| std\_Hb\_wrt\_thres | The standard deviation of the difference between Hb and the threshold for new donations with the alternative algorithm |

In the output file “*results\_characteristics.csv*” the following data will be stored:

|  |  |
| --- | --- |
| **Variable name** | **Description** |
| daterange\_min | Minimum date in the dataset |
| daterange\_max | Maximum date in the dataset |
| donations\_M | Number of donations by males |
| donations\_F | Number of donations by females |
| donors\_M | Number of unique male donors |
| donors\_F | Number of unique female donors |
| mean\_std\_dev\_meas\_M | Measurement variability for males |
| mean\_std\_dev\_meas\_F | Measurement variability for females |
| InputFile | Name of the input file |
| codeversion | The version of the code that was used |

The output file “*hist\_results.rds*” is an R list, containing lists for every configuration of alpha mean and alpha outlier, for both sexes. The lists contain the parameters to plot histograms of the Hb of the false positive, false negative, true positive and true negative records. Only the counts per group are stored, no individual level data.

**Where to find the code**

The code can be downloaded from Github at: [Sanquin/SanguinStats---Donor-Deferral-project: Code for the SanguinStats collaboration (github.com)](https://github.com/Sanquin/SanguinStats---Donor-Deferral-project)

**How to perform an analysis step-by-step**

1. Create a data set containing donor ID, Hb measurements, donation dates and donor sex. Ideally the dataset should contain the full donation history per donor. Save the data either as an R-file or CSV textfile.
2. Download the analysis code for this project from GitHub.
3. If the data are available as a text- or CSV-file use the *“Convert CSV data to RDS.R”* code to convert the text data to an R-data object. This will require some manual adjustment of the code, depending on the format of the data in the textfile.
4. Change various input parameters as outlined under “Code parameters” above.
5. Run the *“alt\_def.R” code.*
6. Send the three output files “confusion\_matrix\_results.csv”, “results\_characteristics.csv” and “hist\_results.rds” to Mart or Amber for further processing.