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

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

Version 1, 23/04/2023

**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>).

The input required to run the code is a data 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.

In the codefile (*“Analyse alternative donor deferral strategy for SanquinStats v1.R”*) a number of parameters need to be specified by the user. Each of these parameters are to be stored in a variable:

1. The name of the datafile to use (variable name “FileToUse”)
2. Minimum acceptable Hb levels for males and females (variable names “dtm” and “dtf”
3. Whether the Hb levels in the datafile and acceptability levels are expressed in g/L (Hb\_in\_gpl=T) or in mmol/L (Hb\_in\_gpl=F)

One additional file with general code (*“General\_functions.R”*) needs to be stored in the same folder as the analysis codefile and will be accessed when running the analysis codefile.

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.

The analysis codefile will save an intermediate datafile (*"donations\_analysis\_data.RDS"*) on disk. Generation of this datafile will for 2 mln donations by 450,000 donors take about 20 minutes. Saving this file on disk will speed up the analyses on subsequent runs, e.g., for making additional graphs of donor profiles. Make sure to remove this file whenever the original data file is changed or updated!

During the analyses various input and output parameters will be stored for post processing. These will only contain information at an aggregated level. However, donor IDs will be included, so these should be anonymized. In the code there is an option to change donor IDs with random numbers if so required (if required, set the changeIDs variable to T).

Various population statistics and an output summary table is stored in an output file called “” *“SavedDeferralData\_DD-MM-YYYY.RDS”* (where DD-MM-YYYY is the day of file creation).

In the output file the following data will be stored:

|  |  |
| --- | --- |
| **Variable name** | **Description** |
| FileToUse | Name of the data file that is used for the analyses |
| Cutoffperc | Cutoff percentile that is used for deferral |
| Daterange | Range of donations in the dataset |
| Nrrecs | Number of records at various stages of exclusion of records |
| meanHb | Mean Hb level per individual donor |
| SdHb | Standard deviation of Hb levels per individual donor |
| nHb | Number of donations per individual donor |
| Defm | Table of donations/deferrals per year for male donors |
| Deff | Table of donations/deferrals per year for female donors |
| Coeff | Coefficient of change in Hb over time for females |
| Sdf | Various standard deviations for female donors |
| Coefm | Coefficient of change in Hb over time for males |
| Sdm | Various standard deviations for male donors |
| outputsummarytable | Table with a statistics of deferral/non-deferral outputs per subsequent donation |

At the end of the analysis codefile donation profiles of some individual donors with specific characteristics will be drawn. As the relevance and interest of these profiles will depend highly on the selection criteria, which again depend on the donor population, the selection criteria for donors to include have to be specified by the user.

Also, by setting the plot\_to\_pdf variable to True (plot\_to\_pdf=T), the graphs can be exported to a PDF file.

**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. Copy the analysis code (3 R-codefiles) 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 (FileToUse, dtm, dtf, Hb\_in\_gpl and changeIDs) in the “Analyse alternative donor deferral strategy for SanquinStats” codefile as required.
5. Run the *“Analyse alternative donor deferral strategy for SanquinStats v1.R”* codefile.
6. Send the “SavedDeferralData\_DD-MM-YYYY.RDS” to Mart or Amber for further processing.
7. Adjust the settings in the *“plot some individual donor profiles”* section of the analysis codefile to explore some individual donor profiles.