**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 3, 09/02/2024

**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 v2.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.

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, but if you want to share plots on individual donor patterns you might want to change these. 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\_XX\_YYYY-MM-DD.RDS”* (where YYYY-MM-DD is the date of the day of file creation and XX is the cutoff percentage used).

All outputs, that will contain only aggregated results will be stored in a sub-folder indicated by the cut-off percentage used in the code (so e.g. “cutoff\_0.99” or “cutoff\_0.999”).

In the output file the following data will be stored:

|  |  |
| --- | --- |
| **Variable name** | **Description** |
| FileToUse | Name of the data file that was used for the analyses |
| maincodedatestamp | Reference to the R-code version that was used |
| generalfunctions  codedatestamp | Reference to the R-code version that was used |
| cutoffperc | Cutoff percentile that is used for deferral |
| daterange | Range of dates of donations in the dataset |
| nrrecs | Number of records at various stages of exclusion of records |
| numdons | Table with the frequency of donations made per donor and sex |
| malefits\_Hb | Kernel density and spline fits and quantiles of the distribution of Hb values and the standard deviation of these values per nr of donations performed for male donors |
| malemeanHb | Mean Hb level of male donors |
| femalefits\_Hb | Kernel density and spline fits and quantiles of the distribution of Hb values and the standard deviation of these values per nr of donations performed for female donors |
| femalemeanHb | Mean Hb level of female donors |
| malefits\_int | Kernel density and spline fits and quantiles of the distribution of donation intervals and the standard deviation of these values per nr of donations performed for male donors |
| malemeanint | Mean donation interval of male donors |
| femalefits\_int | Kernel density and spline fits and quantiles of the distribution of donation intervals and the standard deviation of these values per nr of donations performed for female donors |
| malemeanint | Mean donation interval of female donors |
| 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 female donors |
| sdf | Various standard deviations for female donors |
| coefm | Coefficient of change in Hb over time for male donors |
| sdm | Various standard deviations for male donors |
| analysisresults | Table with a statistics of deferral/non-deferral outputs per subsequent donation |
| analysisresults\_f | Table with a statistics of deferral/non-deferral outputs per subsequent donation for female donors only |
| analysisresults\_m | Table with a statistics of deferral/non-deferral outputs per subsequent donation for male donors only |

At the end of the analysis R-code file some code is provided that allows visualization of donation profiles of individual donors with specific characteristics. For instance, donors that should have been deferred at donation 6 but who in fact donated at least 9 times. 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. Play around and see which parameters provide interesting donation profiles for your donor population!

By setting the plot\_to\_pdf variable to True (plot\_to\_pdf=T), the graphs can be exported to a PDF file instead of to the screen.

All parts in the code that require setting by the user are indicated with the comment ***“# to be set by the USER”*** which will allow you to easily search the code and check whether all settings are in order.

**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 (3 R-code files) 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 v3.R”* file as required.
5. Run the *“Analyse alternative donor deferral strategy for SanquinStats v3.R”* file with the cutoffpercentage set to 0.99 (cutoffperc<-0.99).
6. Run the *“Analyse alternative donor deferral strategy for SanquinStats v3.R”* file with the cutoffpercentage set to 0.999 (cutoffperc<-0.999).
7. Send the two *“SavedDeferralData\_XX\_YYY-MM-DD.RDS”* files (and if possible, also the pdf files in the sub-folders) to Mart or Amber for further processing.
8. Adjust the settings in the *“plot some individual donor profiles”* section of the R-code file to explore individual donor profiles.