Instructions on running the R-code for analysing an alternative donor deferral strategy FOR COUNTRIES WITH A POST-DONATION TESTING STRATEGY

By Mart Janssen & Amber Meulenbeld (m.janssen@sanquin.nl, a.meulenbeld@sanquin.nl)
Version 1, 26/02/2024

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 initial analysis code was developed for countries that perform a pre-donation Hb screening (like the Netherlands). However, some countries perform post-donation screening and defer donors post-donation. In that case most countries perform a pre-donation (re-entry) screening for low-Hb donors only. As this requires a different analysis, R-code was developed to analyse unwarranted deferrals for those settings.

The input required to run the code is a data file containing 5 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 measured in the post-donation screening	Number
pre_Hb	Donor Hb measured using a pre-donation screening (if available)	Number

In case there is sufficient data is available it is recommended to include only donors for which the full donation history is available.

The analysis consists of two parts, one for the analysis of the post-donation measurements (PART1) and one for the analysis of the pre-donation measurements (PART2). The analysis code files are named as follows:

1) *Analyse alternative donor deferral strategy for SanquinStats with post donation measurements_PART1 v1.R*
2) *Analyse alternative donor deferral strategy for SanquinStats with pre donation measurements_PART2 v1.R*

There is no need to create separate datasets for these two analyses. The analysis includes selection of the correct subset of donations (with post-donation or pre-donation measurements available).

Part 1: Analysis of post-donation measurements

In the codefile for post-donation analysis (Part1), 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) Whether plots produced should be saved as PDF files ("plot_to_pdf")
3) Minimum acceptable Hb levels for males and females (variable names "dtm" and "dtf")
4) 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 codefiles 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 referenced 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 objects 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 will 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 (<Control+F>) and check whether all settings are in order.

Part 2: Analysis of pre-donation measurements

In the codefile for pre-donation analysis (Part 2), again 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) The name of the output file created during **the previous step** (variable name "PostDonationResultsFile"). Note that this filename required here will be something like *"SavedDeferralData_XX_YYYY-MM-DD.RDS",* where XX is the cutoff percentage used and YYYY-MM-DD is the date of file creation.
3) Whether plots produced should be saved as PDF files ("plot_to_pdf")
4) Minimum acceptable Hb levels for males and females (variable names "dtm" and "dtf")
5) 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)

Various population statistics and analysis results will be stored in an output file called *"SavedDeferralData_PREDONATION_XX_YYYY-MM-DD.RDS",* where again XX is the cutoff percentage used and YYYY-MM-DD is the date of file creation.

All outputs, that will contain only aggregated results, will be stored in a sub-folder referenced 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 objects will be stored:

Variable name	Description
FileToUse	Name of the data file that was used for the analyses
PostDonationResultsFile	Name of the file that was used to access results of the post-donation analysis (Part 1)
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
pre_defm	Table of pre-donation screenings/deferrals per year for male donors
pre_deff	Table of pre-donation screenings/deferrals per year for female donors
coeff_notdeferred	Coefficient of change in Hb over time for female donors that were not deferred
coeff_deferred_capillary	Coefficient of change in capillary Hb over time for female donors that were deferred
coeff_deferred_venous	Coefficient of change in venous Hb over time for female donors that were deferred
coefm_notdeferred	Coefficient of change in Hb over time for male donors that were not deferred
coefm_deferred_capillary	Coefficient of change in capillary Hb over time for male donors that were deferred
coefm_deferred_venous	Coefficient of change in venous Hb over time for male donors that were deferred
sd_cap	Various standard deviations for the capillary Hb measurement
mean_cap	Various mean differences between subsequent measurements for the capillary Hb measurement
sd_cap2	Various standard deviations for the capillary Hb measurement based on the difference with the last measurement
analysisresults	Summary statistics of pre-donation screening
analysisresults_f	Summary statistics of pre-donation screening for females only
analysisresults_m	Summary statistics of pre-donation screening for males only

Where to find the code

The code can be downloaded from Github at https://github.com/Sanquin/SanquinStats---Donor-Deferral-project/tree/main/Code%20for%20a%20Post-donation%20screening%20setting

How to perform the analysis step-by-step

1) Create a data set containing donor ID, pre-donation and post-donation 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 (4 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 with pre donation measurements_PART1 v1.R"* file as required.

5) Run this analysis with the cutoff percentage set to 0.99 (cutoffperc<-0.99).

6) Run this analysis with the cutoff percentage set to 0.999 (cutoffperc<-0.999).

7) Adjust the settings in the "plot some individual donor profiles" section of the R-code file to explore individual donor profiles.

8) Change various input parameters (FileToUse, dtm, dtf, Hb_in_gpl and changeIDs) in the *"Analyse alternative donor deferral strategy for SanquinStats with pre donation measurements_PART2 v1.R"* file as required. Pay special attention to the "PostDonationResultsFile" parameter which has to reference the output file from the previous analysis.

9) Run this analysis code with the cutoff percentage set to 0.99 (cutoffperc<-0.99). Use the "PostDonationResultsFile" that was produced using the cutoff percentage 0.99.

10) Run this analysis code with the cutoff percentage set to 0.999 (cutoffperc<-0.999). Use the "PostDonationResultsFile" that was produced using the cutoff percentage 0.999.

11) Send the two *"SavedDeferralData_XX_YYY-MM-DD.RDS"* and the two *"SavedDeferralData_PREDONATION_XX_YYYY-MM-DD.RDS"* files (and if possible, also the pdf files in the sub-folders) to Mart or Amber for further processing.