**Statistical Monitoring Program Instructions.**

1. **Date order checking.**

The function date\_order\_check checks whether dates appear to fall in the wrong order.

Parameters to give the function:

1. Date of randomisation/registration:

This should be in the form of a data frame with the participant id number in the first column and the date the participant entered the trial (date of registration/randomisation) in the second column. The participant ID can take any form (numerical/string) but the date must be given in the form dd/mm/yyyy.

Data frames can be read in with the following code:

**options(stringsAsFactors = FALSE)**

**reg.data<- data.frame(read.table("STUDY12\_REG.txt", row.names=NULL, header=TRUE, sep="\t" ))**

(This would read in a text file called *STUDY12\_REG.txt* and store it in the data frame *reg.data*.)

1. The other data sheets to check:

These should be given as data frames in a list. Each sheet should have the participant ID in the first column followed by all of the other dates to check. All non-date variables or dates you do not wish to check i.e. dates that do not relate directly to the participant (e.g. dates of date entry) should be removed. Dates should be given in the form dd/mm/yyyy.

Date frames for each sheet can be read in as above and then saved as a list using this code:

**sheets.to.check<- list(STUDY12\_QOL= qol, STUDY12\_FUP= fup, STUDY12\_DRUGSUMM = drugsumm, STUDY12\_CHEMO= chemo, STUDY12\_ADDTREAT= addtreat)**

(This would create a list with 5 data sheets. The first is quality of life data from the data frame “qol”, given the name “STUDY12\_QOL”. The data frames should be given recognisable names which will be output if any errors are found in them.)

1. Dates of death:

The death data should be given in a data frame. The first column should be the participant ID, followed by the date of death (in the form dd/mm/yyyy). Any other dates which need to be checked from the same data sheet may also be included but they must come AFTER the dates of death, i.e. columns 3+.

1. The start date:

This should be the date the trial opened or the date of the first randomisation. All dates should fall after this date. It should be given as a single date in the form dd/mm/yyyy. For example:

**start.date<-“06/06/2013”**

1. The end date:

This should be the end of the trial or the date the data was dumped. i.e. no date should fall after this point. As with the start date, it should be entered as a single date in the form dd/mm/yyyy. For example:

**end.date<-“06/06/2013”**

1. The trial name

The name of the trial. This will be used to label the output files. For example:

**trial.name<- “STUDY12”**

**Calling the function**

Once the program and the parameters above are stored in R’s memory the program can be run using the following command:

**date\_order\_check(reg.data, sheets.to.check, death.data,start.date,end.date,"STUDY12")**

Where each parameter is stored as in 1-6.

**The output:**

The program outputs 2 text files.

The first has a name in the form: *“TRIALNAME\_ERRORS\_2013-09-02.txt”*

Where TRIALNAME is the name of the trial and the date would be replaced with the date the program was run. This will store any errors detected in the form:

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **ID** | **Sheet with error** | **Variable with error** | **Problem** | **Error Date** | **Comparison date\*** |
| 162 | STUDY12\_FUP | DFUP | Date appears after the date of death | 10/11/2005 | 23/12/2004 |
| 439 | STUDY12\_DRUGSUMM | DDRUGSTART | Date appears before registration | 02/02/2005 | 07/02/2005 |
| 532 | STUDY12\_DRUGSUMM | DDRUGSTART | Date appears before registration | 25/05/2005 | 26/05/2005 |
| 81 | STUDY12\_CHEMO | DPREFBC | Date appears before registration | 25/11/2003 | 27/11/2003 |
| 250 | STUDY12\_CHEMO | DPREFBC | Date appears before registration | 28/04/2004 | 24/05/2004 |
| . | . | . | . | . | . |
| . | . | . | . | . | . |

\*Comparison date is the date used for comparison, for row 1 this will be patient 162’s date of death. For row 2 this will be patient 439’s date of registration.

If no errors are found the text file will contain the message “No errors found”.

The second text file has a name in the form: *“TRIALNAME\_vars\_checked\_2013-09-02”* and contains information about the numbers of sheets, variables, values etc checked:

**Warnings:**

There are no error messages coded into the function. If data is not read in as above the function may not work as it should, or possibly at all. Please take care when creating the parameters from your data.

Please ensure that the dates you are checking should be checked. For example a date of diagnosis would be expected before the date of registration or randomisation. If you wish to check dates like this (comparison with the dates of death and end date) the output files can be sorted by “variable” and “problem” to make the removal of the “before registration” results easier.