**Statistical Monitoring Program Instructions.**

**8 SAE checker**

The function **SAE\_check** is used to monitor the number of SAEs in each site. It calculates a rate based on the number of patients with an SAE, the total number of patients and the time the trial has been recruiting. It displays this graphically and highlights the patients with extreme (high or low) rates. These are also output into a text file.

Parameters to give the function:

1. data.rand:

This is the randomisation data. It should be in the form of a data frame with the patient id in the first column, the site ID in the second column (can be numeric or string), and the date of randomisation (in the form dd/mm/yyyy) in the third column. All patients should be included.

Data frames can be read in with the following code:

**options(stringsAsFactors = FALSE)**

**data.rand<-data.frame(read.table("ABC02\_REG.txt", row.names=NULL, header=TRUE, sep="\t"))**

(This would read in a text file called *ABC02\_REG.txt* and store it in the data frame *data*.*rand*)

1. data.saes

This is a vector of the ID numbers for all patients with an SAE. Patients should appear in the list for each SAE they have, i.e. multiple times for multiple SAEs. For example, if a trial had 5 patients with IDs 1-5 and patient 1 had 2 SAEs, patient 2 has none, patient 3 has 1, patient 4 has none and patient 5 has 4 this would be stored as:

**data.saes<-c(1,1,3,5,5,5,5)**

(This could be taken as the ID column from any spreadsheet which stores a complete list of SAES)

1. xmonths

This is the length of time after randomisation/registration when SAEs are likely to be recorded, i.e. if treatment lasted 6 months and SAEs were recorded for one month after the end of treatment this would be set as 7.

1. trial.name

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

**trial.name<- “ABC02”**

1. p

This is the percentage of low/high rates to output. For 10% set this as 10 for 20% as 20 etc.

(Sites with zero rates will be output on top of this percentage).

1. end.date

This is the date the data was extracted. This will be used to calculate how long each site has been recruiting.

**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:

**SAE\_check(data.rand, data.saes, xmonths, trial.name, p, end.date)**

Where each parameter is stored as in 1-6

**The output:**

The plot

The program outputs a plot that shows the rate against the number of patients in each site. The rate is the number of patients in each site with an SAE, divided by the total number of patients in the site, further divided by the time the site has been open. The time the site has been open, is calculated as the time between the first randomisation and the date the last SAE would be seen. This is the date of the last randomisation plus xmonths (or the date the data was dumped if this is sooner).

Sites with low rates (p percent plus all sites with zero rates) are shown as solid black squares and those with high rates are shown as solid black circles. The horizontal line shows the median rate. Sites shown as black squares on the right-hand side may need investigation as they have large numbers of patients but low rates.

The plot above has the name: *“SAE\_rate\_VS\_site\_size\_ABC02\_10percent\_2014-07-17”*

Where ABC02 is the name of the trial (this is whatever has been specified as trial.name when calling the function) and the date would be replaced with the date the program was run. 10% was the percentage chosen.

The program also outputs more detailed information about the sites with low and high rates. It has a name of the form:

 *“SAE\_CHECK\_details\_10\_percent\_ABC02\_2014-07-17.txt”* and contains details of all of the information which goes into calculating the rate as well as the total number of SAEs recorded for each site.



**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.