NHS
Royal Papworth Hospital
NHS Foundation Trust

GD015

GD015 Monitoring and Audit: Random Sample Selection

R can be used to select a random sample from a set of numbers. This can be used to select a sample of participants from a study to be monitored. The following script can be run in R to produce the random sample selection:

```
#STUDY NAME P09999
set.seed(9999DDMM)
set=c()
pick=
sample(set,pick,replace=FALSE)
```

The set.seed() function in R is used to reproduce results i.e. it produces the same sample again and again. When we generate randoms numbers without set.seed() function it will produce different samples at different time of execution. To demonstrate that the script has not been rerun multiple times with different seeds to produce a desired result, the seed should have an obvious meaning. When sampling participants for a study, the study P0 number should be used in conjunction with the date in DDMM format. This is to prevent repeat selections of the same subject when the sample size remains constant.

`set` defines the sample from which the selection will be made. Numeric values are allowed as-is e.g. set=c(1,2,3….; non-numeric values must each be enclosed in single quotes e.g. set=c('17-001','17-002','17-003'.
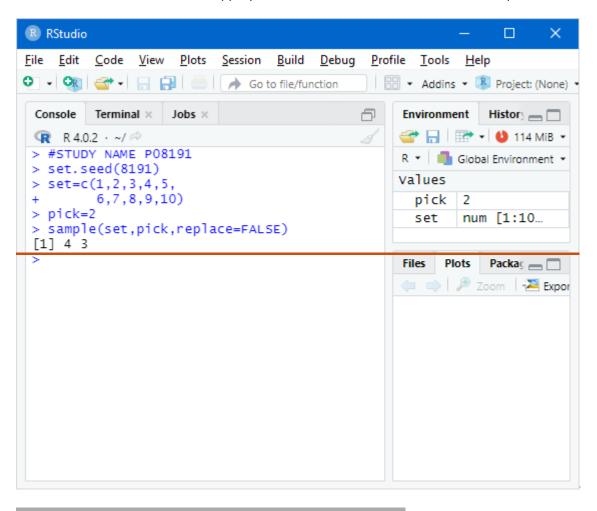
`pick` defines how many to sample.

These values are defined and then included in the following argument:
`sample(set,pick,replace=FALSE)`

`replace=FALSE` ensures each number can only be selected once.

An example script selecting 2 participants from a possible 10:

```
#STUDY NAME P08191
set.seed(8191DDMM)
set=c(1,2,3,4,5,
      6,7,8,9,10)
pick=2
sample(set,pick,replace=FALSE)
```

Once the script has been run and the sample selection has been generated, a screen shot of the R console should be saved and E-mailed to the person who requested the random sample selection. The E-mail should be saved in an appropriate location so it can be referred to if required.