

Usage of R programs for statistical powers of multistage designs

1. At first, before you run this R program, it is necessary to install "**mvtnorm**" package from download site of R project (<http://www.r-project.org/>). If you are not R users, you can install the R software freely from the same site.
2. Two kinds of programs are prepared for each input parameter. One is for disease model parameters, and the other is for practical 2×3 contingency table data of genotype AA, Aa and aa (as for example) as follows:

Input parameters for disease model parameters

Dim:	number of stages
MDL:	kinds of genotype models as follows; 1: allelic model (multiplicative model), 2: additive model, 3: dominant model, 4: recessive model
Alphagenome:	genome-wide false positive rate
N:	number of samples in cases and controls
M:	number of candidate loci
prev:	population prevalence of disease
freq:	population disease-associated allele rate
grr:	population genotype relative risk (GRR)
tm:	number of true loci in M candidate loci
pisamples k:	proportion of the sample size at k th stage
pimarkers k:	proportion of loci to be selected at k th stage

Input parameters for contingency table data

Dim:	number of stages
MDL:	kinds of genotype models as follows; 1: allelic model (multiplicative model), 2: additive model, 3: dominant model, 4: recessive model
alphagenome:	genome-wide false positive rate
M:	number of candidate loci
tm:	number of true loci in M candidate loci
case k:	number of each genotype, for example, AA, Aa and aa of cases at k th stage
cont k:	number of each genotype, for example, AA, Aa and aa of controls at k th stage

3. When you input above-mentioned parameters directly into a code, select the program sheet, and copy and paste it on the R console screen. As soon as you do so, the program will start running automatically, and calculate the following output parameters.

Output parameters

Output parameters for one-stage design

Co:	critical value of one-stage design
Power:	statistical power of one-stage design
PPV:	positive predictive value (PPV) of one-stage design
Typings:	number of genotypings

Output parameters for more than one-stage design

C of RBA:	critical value of replication-based analysis (RBA)
C of JA:	critical value of joint analysis (JA)
Power of RBA:	statistical power of RBA
Power of JA:	statistical power of JA
PPV of RBA:	positive predictive value (PPV) of RBA
PPV of JA:	positive predictive value (PPV) of JA
Typings:	number of genotypings

4. You can also use our R package (add-on package) to install the package from our website to your local R program files location. The package name is "nstage" and the zip file is "nstage_1.0.zip". After installing the package, load it by inputting "library(nstage)" on the R. Regarding the usage of various functions in the package, please refer the R documentation by inputting "help(nstage)" or "?nstage" on the R. For example, using "nfun" function of the package, it allows the user to specify an arbitrary n (the number of stages) as one of the input parameters.

5. Furthermore, in our website, we developed the web user interface (WUI) for calculating statistical powers, positive predictive values, numbers of typings and other output parameters of the multistage design. By using it, even if a reader cannot use R program at all, only by inputting the parameters or changing existing default values in cells and by pushing "Enter" key of your computer or by clicking "SUBMIT" button at the foot of the WUI, the results by R can be calculated (see <http://www.med.niigata-u.ac.jp/eng/resources/informatics/gwa.html>).

Examples

R program for arbitrary n-stage designs for disease model parameters

For simulating by arbitrary n-stage designs using disease model, the program titled “Program for arbitrary n-stage designs for disease model parameters” is prepared. Assume that the number of stages = 5, allelic (multiplicative) model, genome-wide false positive rate (alphagenome) = 0.05, N = 1000, M = 500,000, population allele prevalence = 0.1, disease allele frequency = 0.4, GRR= 1.5, number of true loci in M candidate loci = 5, input the parameters in the programs as follows (see red circles):

```
#n-stage design
#Input parameters
Dim <- 5 # number of stages
MDL <- 1 # genotype model 1:multiplicative model, 2:additive model, 3:dominant
alphagenome <- 0.05 # genome-wide false positive rate
N <- 1000 # number of samples in cases and controls
Pisamples1 <- 1/(Dim) # proportion of the sample size at 1st stage
Pimarkers1 <- (0.0001)^(1/(Dim-1)) # proportion of alleles to be selected at 1st stage
M <- 500000 # number of candidate loci
prev <- 0.1 # population prevalence of disease
freq <- 0.3 # population disease allele rate
grr <- 1.5 # population genotype relative risk (GRR)
tm <- 5 # number of true loci in M candidate loci
```

And the output parameters are displayed as follows:

nStageDesign	Output
1 C of RBA:	3.090
2 C of JA:	5.108
3 Power of RBA:	0.344
4 Power of JA:	0.707
5 PPV of RBA:	0.996
6 PPV of JA:	0.986
7 Typings:	222220000

R program for arbitrary n-stage designs for contingency table data

For simulating by arbitrary n-stage designs using practical contingency table data, the program titled “Programs for arbitrary n-stage designs for contingency table data” is prepared. Assume that the number of stages = 5, allelic (multiplicative) model, the numbers of AA, Aa and aa of cases are and the number of AA, Aa and aa of controls are described as follows, respectively, genome-wide false positive rate (alphagenome) = 0.05, M = 500,000 and number of true loci in M candidate loci = 5, input the parameters in the programs as follows:

```
#n-stage design
#Input parameters
Dim <- 5 # number of stages
MDL <- 1 # genotype model 1:multiplicative model, 2:additive model, 3:dominant
alphagenome <- 0.05 # genome-wide false positive rate
tm <- 5 # number of true loci in M candidate loci
case[1,] <- c(50,30,20) # number of each genotype of cases at 1st stage
cont[1,] <- c(35,30,35) # number of each genotype of controls at 1st stage
case[2,] <- c(50,30,20) # number of each genotype of cases at 2nd stage
cont[2,] <- c(35,30,35) # number of each genotype of controls 2nd stage
case[3,] <- c(50,30,20) # number of each genotype of cases at 3rd stage
cont[3,] <- c(35,30,35) # number of each genotype of controls 3rd stage
case[4,] <- c(50,30,20) # number of each genotype of cases at 4th stage
cont[4,] <- c(35,30,35) # number of each genotype of controls 4th stage
case[5,] <- c(50,30,20) # number of each genotype of cases at 5th stage
cont[5,] <- c(35,30,35) # number of each genotype of controls 5th stage
```

And the output parameters are displayed as follows:

nStageDesign	Output
1 C of RBA:	3.090
2 C of JA:	5.108
3 Power of RBA:	0.882
4 Power of JA:	0.986
5 PPV of RBA:	0.999
6 PPV of JA:	0.990
7 Typings:	222220000