

# Package ‘RPower’

October 13, 2015

**Type** Package

**Title** RPower: Power Estimation and Sample Size Determination for Replication Studies of GWAS

**Version** 1.0

**Date** 2015-10-11

**Author** Wei Jiang and Weichuan Yu

**Maintainer** Wei Jiang <wjiaa@connect.ust.hk>

**Description** An Empirical Bayes (EB) based method to estimate the power of replication study in Genome-wide association studies. We can use this method to determine the sample size needed in the replication study.

**License** GPL-3

**Depends** splines

## R topics documented:

RPower-package	1
RPower-functions	4
<b>Index</b>	<b>7</b>

---

RPower-package	<i>RPower: Power Estimation and Sample Size Determination for Replication Studies of Genome-Wide Association Studies</i>
----------------	--

---

## Description

RPower is an R package to estimate power and determine the sample size for replication studies of genome-wide association studies. The power estimation method is based on Empirical Bayes, which is used for reducing bias of the "Winner's curse" in the primary study. The credible interval of the power can also be estimated from the package.

## Details

Package: RPower  
 Type: Package  
 Version: 1.0  
 Date: 2015-10-11  
 License: GPL-3

Replication study is a commonly used verification method to filter out false positives in genome-wide association studies (GWAS). If an association can be confirmed in a replication study, it will have a high confidence to be true positive. To design a replication study, traditional approaches calculate power by treating replication study as another independent primary study. These approaches do not use the information given by primary study. Besides, they need to specify a minimum detectable effect size, which may be subjective. One may think to replace the minimum effect size with the observed effect sizes in the power calculation. However, this approach will make the designed replication study underpowered since we are only interested in the positive associations from the primary study and the problem of the "winner's curse" will occur.

Here we provided RPower as an R package to estimate power and determine the sample size for replication studies of genome-wide association studies. The power estimation method is based on Empirical Bayes, which is used for reducing bias of the winner's curse in the primary study. Details about the method can be seen in our reference paper below.

The principal components of RPower are `repPowerEB`, `repSampleSize` and `repSampleSize2`. Also there is a simple function `SEest` in the package.

1. To estimate the power of replication study in GWAS, we need to obtain the observed log-odds ratios and their corresponding standard errors of every genotyped SNPs first. We have put an example summary statistics data (`gwasSmryEx`) in our package. You can use `data(gwasSmryEx)` to load the example data.

2. Then the standard errors of  $\log(\text{OR})$  in the replication study need to be anticipated, which can be calculated using Woolf's method with function `SEest`:

```
SEest(n0, n1, fU, fA)
```

Details about the function can be seen using [help\(SEest\)](#).

3. You can use `repPowerEB` to estimate the powers of each identified primary positive associations (i.e. associations identified from primary study) in the replication study. Their credible intervals and average power are also presented in the results.

```
repPowerEB(MUhat, SE, SE2, zalpha2, zalphaR, boot=100, num=100, output=T, dir=output, info=T)
```

Details about the function can be seen using [help\(repPowerEB\)](#).

4. To determine the sample size of the replication study with a certain power, `repSampleSize` or `repSampleSize2` can be used.

```
repSampleSize(power, n, MUhat, SE, zalpha2, zalphaR)
```

```
repSampleSize2(power, CCR2, MUhat, SE, fU, fA, zalpha2, zalphaR)
```

`repSampleSize` is used when the control to case ratio (CCR) of the replication study is the same with the primary study. Otherwise, for a designed replication study with specific CCR, `repSampleSize2` can be used.

Details about these function can be seen using [help\(repSampleSize\)](#) or [help\(repSampleSize2\)](#).

**Author(s)**

Wei Jiang and Weichuan Yu

Maintainer: Wei Jiang <wjiangaa@connect.ust.hk>

**References**

Jiang, W. and Yu, W. *Power Estimation and Sample Size Determination for Replication Studies of Genome-Wide Association Studies*. Submitted to APBC2016.

**See Also**

[SEest](#), [repPowerEB](#), [repSampleSize](#), [repSampleSize2](#)

**Examples**

```
##Load data
data(gwasSmryEx)      #Example of summary statistics in GWAS
MU<-log(gwasSmryEx$OR) #log(Odds Ratio)
SE<-gwasSmryEx$SE     #SE of log(OR)
fU<-gwasSmryEx$F_U   #Reference allele frequencies in controls
fA<-gwasSmryEx$F_A   #Reference allele frequencies in cases

##Some parameters used for power estimation of the example data
alpha<-5e-5          #The significance level of the primary study
alphaR<-5e-3         #The significance level of the replication study
n2_0<-500           # The control number in the replication study
n2_1<-500           # The case number in the replication study
zalpha2<-qnorm(1-alpha/2) #Cutoff value for primary study
zalphaR<-qnorm(1-alphaR) #cutoff value for replication study

##Estimate standard error in the replication study
SE2<-SEest(n2_0,n2_1,fU,fA)

##Estimate power
###The average power for replication study is powerResult$avgPower
###CI: [powerResult$bootAvgCIlow,powerResult$bootAvgCIhigh]
powerResult<-repPowerEB(MU,SE, SE2,zalpha2,zalphaR,output=TRUE,info=TRUE, dir=)

##Some parameters used for determining the sample size of the replication study
power<-0.8          #Power we needed in the replication study
n<-2000             #Example data are generated based on 2000 individuals
# (#controls:#cases=1:1)

##The sample size we needed to obtain a certain power in the replication study
##The same case-to-control ratio used as the primary study
n2.1<-repSampleSize(power,n,MU,SE,zalpha2,zalphaR)

CCR<-2              #Designed control-to-case ratio in the replication study
##The sample size we needed to obtain a certain power in the replication study
##Different control-to-case ratio used as the primary study
n2.2<-repSampleSize2(power,CCR,MU,SE,fU,fA,zalpha2,zalphaR)
```

**Description**

RPower is an R package to estimate power and determine the sample size for replication studies of genome-wide association studies. The principal functions include `repPowerEB`, `repSampleSize` and `repSampleSize2`. `repPowerEB` is an Empirical Bayes (EB) based method to estimate the power of replication studies in GWAS. `repSampleSize` and `repSampleSize2` are used for determining the sample size of the replication study with desired power. There is a simple function `SEest` in the package, which can be used to estimate the standard errors of the observed log-odds ratios using Woolf's method.

**Usage**

```
repPowerEB(MUhat, SE, SE2, zalpha2, zalphaR, boot = 100, num = 100,
  output = T, dir = "output", info = T)
```

```
repSampleSize(power, n, MUhat, SE, zalpha2, zalphaR)
```

```
repSampleSize2(power, CCR2, MUhat, SE, fU, fA, zalpha2, zalphaR)
```

```
SEest(n0, n1, fU, fA)
```

**Arguments**

<code>MUhat</code>	The observed log-odds ratio in the primary study.
<code>SE, SE2</code>	The standard error of log(OR) in the primary study and replication study, respectively.
<code>zalpha2</code>	The critical value of z-values used in the primary study, i.e. $z_{\alpha/2}$ .
<code>zalphaR</code>	The critical value of z-values used in the replication study, i.e. $z_{\alpha R}$ .
<code>boot</code>	The resampling number of bootstrap used for estimating the credible interval of the power.
<code>num</code>	The sampling number of Monte carlo method used for estimating the credible interval of the power.
<code>output</code>	Bool value. To determine whether to output the estimated results in the <code>dir</code> or not.
<code>dir</code>	The directory to save the estimated results. It has effect when <code>output=T</code>
<code>info</code>	Bool value. To determine whether to show the hyperparameters inference results in the terminal or not.
<code>power</code>	The desired power to design a replication study.
<code>n</code>	The original sample size used in the primary study.
<code>CCR2</code>	The control-to-case ratio used in the replication study.
<code>fU, fA</code>	The reference allele frequencies in the control and case group, respectively.
<code>n0, n1</code>	The individual number in the control and case group, respectively.

## Details

These functions are the implementation of the Empirical Bayes based method to estimate power and determine the sample size for replication studies of genome-wide association studies. The method has good performance in terms of overcoming winner's curse and providing higher estimation accuracy. Details about the method can be seen in our reference paper below.

repPowerEB can be used to estimate the powers of each identified primary positive associations (i.e. associations identified from primary study) in the replication study. Their credible intervals and average power are also presented in the results.

repSampleSize or repSampleSize2 can be used to determine the sample size of the replication study with a certain power. repSampleSize is used when the control to case ratio (CCR) of the replication study is the same with the primary study. Otherwise, for a designed replication study with specific CCR, repSampleSize2 can be used.

SEest can be used to estimate the standard errors of log(OR) in the replication study.

## Value

repPowerEB gives the estimated power of the replication study. The returned value is a LIST:

idx	The index of the identified associations from the primary study.
power	The estimated replication study's powers for each identified primary positive associations.
bootCIlow, bootCIhigh	The lower and upper bound in the estimated credible interval for the power of each primary positive association.
avgPower	The average power of all identified associations from the primary study.
bootAvgCIlow, bootAvgCIhigh	The lower and upper bound in the credible interval of avgPower
repSampleSize and repSampleSize2	return the sample size needed in the replication study.

## Author(s)

Wei Jiang and Weichuan Yu

Maintainer: Wei Jiang <wjiaa@connect.ust.hk>

## References

Jiang, W. and Yu, W. *Power Estimation and Sample Size Determination for Replication Studies of Genome-Wide Association Studies*. Submitted to APBC2016.

## See Also

[RPower](#)

## Examples

```
##Load data
data(gwasSmryEx)           #Example of summary statistics in GWAS
MU<-log(gwasSmryEx$OR)    #log(Odds Ratio)
SE<-gwasSmryEx$SE        #SE of log(OR)
fU<-gwasSmryEx$F_U       #Reference allele frequencies in controls
fA<-gwasSmryEx$F_A       #Reference allele frequencies in cases
```

```

##Some parameters used for power estimation of the example data
alpha<-5e-5           #The significance level of the primary study
alphaR<-5e-3          #The significance level of the replication study
n2_0<-500             # The control number in the replication study
n2_1<-500             # The case number in the replication study
zalpha2<-qnorm(1-alpha/2) #Cutoff value for primary study
zalphaR<-qnorm(1-alphaR) #cutoff value for replication study

##Estimate standard error in the replication study
SE2<-SEest(n2_0,n2_1,fU,fA)

##Estimate power
###The average power for replication study is powerResult$avgPower
###CI: [powerResult$bootAvgCIlow,powerResult$bootAvgCIhigh]
powerResult<-repPowerEB(MU,SE, SE2,zalpha2,zalphaR,output=TRUE,info=TRUE, dir=)

##Some parameters used for determining the sample size of the replication study
power<-0.8           #Power we needed in the replication study
n<-2000              #Example data are generated based on 2000 individuals
                    # (#controls:#cases=1:1)
##The sample size we needed to obtain a certain power in the replication study
##The same case-to-control ratio used as the primary study
n2.1<-repSampleSize(power,n,MU,SE,zalpha2,zalphaR)

CCR<-2               #Designed control-to-case ratio in the replication study
##The sample size we needed to obtain a certain power in the replication study
##Different control-to-case ratio used as the primary study
n2.2<-repSampleSize2(power,CCR,MU,SE,fU,fA,zalpha2,zalphaR)

```

# Index

## \*Topic **design**

RPower-functions, 4

RPower-package, 1

## \*Topic **package**

RPower-package, 1

gwasSmryEx (RPower-package), 1

help(repPowerEB), 2

help(repSampleSize), 2

help(repSampleSize2), 2

help(SEest), 2

repPowerEB (RPower-functions), 4

repSampleSize (RPower-functions), 4

repSampleSize2 (RPower-functions), 4

RPower, 5

RPower (RPower-package), 1

RPower-functions, 4

RPower-package, 1

SEest (RPower-functions), 4