

Software Documentation

Iuliana Ionita-Laza (ii2135@columbia.edu)

This document describes the use of the accompanying R code that implements the methods described in Ionita-Laza and Ottman (2011). The code is useful in comparing the performance of various affected-relative pair designs as well as designs based on unrelated individuals in the context of association studies with rare variants. There are two versions of the code, depending on the measure used to assess the performance:

1. Expected P-value: `software_EPV.R`
2. Power: `software_Power.R`

To run the code type: `source(software_EPV.R)` or `source(software_Power.R)`. There are several main functions that one could use:

1. Function `'EPV_pairs(f,r,k,nnn,fi,lambda_s)'` returns the expected P-value for a study with
 - `nnn` = # pairs of affected relatives
 - `fi`=1/4 for sibs, `fi`=1/16 for first cousins, `fi`=1/64 for second cousins, `fi`=0 for unrelateds
 - variant frequency `f`
 - genotype relative risk `r`
 - population prevalence `k`
 - sibling recurrence risk ratio `lambda_s`.
2. Function `'EPV_one_pairs(f,r,k,nnn,fi,lambda_s)'` is similar with the only difference that `nnn` now represents the number of affected individuals with a known affected relative.
3. Function `'EPV_pairs_aggregate(f,r,nd,k,nnn,fi,lambda_s)'` calculates the expected P-value at a locus with `nd` variants whose frequencies and GRRs are given in vectors `f` and `r`.
4. Function `'EPV_one_pairs_aggregate(f,r,nd,k,nnn,fi,lambda_s)'` is as in [3.] only that `nnn` represents the number of affected individuals with a known affected relative.

Similar functions are defined if Power (rather than EPV) is desired. For power an alpha level needs to be specified. Note that, since we assume that the frequency in the population is known, the resulting power levels need to be regarded as achievable when the number of controls grows large. The corresponding functions are:

1. Power_pairs(f,r,k,nnn,fi,lambda_s,alpha_level)
2. Power_one_pairs(f,r,k,nnn,fi,lambda_s,alpha_level)
3. Power_pairs_aggregate(f,r,nd,k,nnn,fi,lambda_s,alpha_level)
4. Power_one_pairs_aggregate(f,r,nd,k,nnn,fi,lambda_s,alpha_level)

Examples

EPV

To calculate the expected P-value at a variant with frequency $f = 0.01$ and $GRR = 1.2$ for a study with 1000 affected sib-pairs, for a disease with population prevalence $k = 0.03$ and sib recurrence risk ratio $\lambda_s = 2$ type: EPV_pairs(0.01, 1.2, 0.03, 1000, 1/4, 2).

Similarly, for a study with 2000 affected individuals with an affected sibling, type: EPV_one_pairs(0.01, 1.2, 0.03, 2000, 1/4, 2)

For a locus with 3 disease variants and 3 random variants with $f = (0.001, 0.005, 0.01, 0.001, 0.001, 0.007)$ and $GRR = (2, 2, 2, 1, 1, 1)$ and for a study with 1000 affected sib-pairs, for a disease with population prevalence $k = 0.03$ and sib recurrence risk ratio $\lambda_s = 2$ type:

$f = c(0.001, 0.005, 0.01, 0.001, 0.001, 0.007)$
 $GRR = c(2, 2, 2, 1, 1, 1)$
 EPV_pairs_aggregate(f , GRR , 6, 0.03, 1000, 1/4, 2)

Power

To calculate the power for $\alpha = 1.6 \cdot 10^{-6}$ at a variant with frequency $f = 0.01$ and $GRR = 1.2$ for a study with 1000 affected sib-pairs, for a disease with population prevalence $k = 0.03$ and sib recurrence risk ratio $\lambda_s = 2$ type: Power_pairs(0.01, 1.2, 0.03, 1000, 1/4, 2, $1.6 * 10^{-6}$).

Similarly, for a study with 2000 affected individuals with an affected sibling, type: Power_one_pairs(0.01, 1.2, 0.03, 2000, 1/4, 2, $1.6 * 10^{-6}$).

For a locus with 3 disease variants and 3 random variants with $f = (0.001, 0.005, 0.01, 0.001, 0.001, 0.007)$ and $GRR = (2, 2, 2, 1, 1, 1)$ and for a study with 1000 affected sib-pairs, for a disease with population prevalence $k = 0.03$ and sib recurrence risk ratio $\lambda_s = 2$ type:

$f=c(0.001,0.005,0.01,0.001,0.001,0.007)$

$GRR=c(2,2,2,1,1,1)$

$\text{Power_pairs_aggregate}(f, GRR, 6, 0.03, 1000, 1/4, 2, 1.6 * 10^{-6})$