Package 'BayesMetaPenetrance'

March 29, 2023

Type Package

Title Bayesian Meta Analysis to estimate age-specific penetrance of getting cancer due to pathogenic variants of a given gene

Version 1.0

Date 2022-07-20

Author Thanthirige Lakshika Maduwanthi Ruberu

Maintainer Thanthirige Lakshika Maduwanthi Ruberu <lakshikaruberu@gmail.com>

Description

Estimate meta-analytic age-specific risk of getting cancer (penetrance) due to pathogenic variants of a given gene by integrating information from studies reporting different types of risk measures on that particular gene. These risk measures include age-specific penetrance, relative risk (RR), standard incidence ratio (SIR), and odds ratio (OR).

Depends R (>= 3.6.0)

License GPL-2

LinkingTo Rcpp

Imports dplyr (>= 1.0.9), mvtnorm (>= 1.1-3) , rlist (>= 0.4.6.2) , Rcpp (>= 1.0.9)

R topics documented:

`BayesMetaPenetrance` *Estimates consensus age-specific penetrance of cancer for carriers of a pathogenic gene mutation*

Description

`BayesMetaPenetrance` is used to estimate meta-analytic age-specific risk of getting cancer (penetrance) for carriers of pathogenic variants of a specific gene. Information from studies reporting different types of risk measuers on that gene (e.g., age-specific penetrance, relative risk(RR), standard incidence ratio (SIR), and odds ratio (OR)) are integrated to give the age-specific penetrance for carriers at each specified age up to age 85.

Usage

```
BayesMeta(penet,RR_studies=T,RR,OR_studies=T,OR,ages=20:85,zero_studies=T,zero_OR,pl=F ,
          ylim=c(0,1),xlim=c(20,85))
```

Arguments

`penet` A data frame containing information for studies reporting age-specific penetrance with following columns:

study_label	label to identify the set of penetrance values from a single study.
penetrance	age-specific penetrance value.
penet_ci_lower	lower 95% confidence limit.
penet_ci_upper	upper 95% confidence limit.
ages_penet	age at which penetrance is reported.

`RR_studies` A logical variable to indicate whether studies reporting RR or SIR are included in the meta analysis. Default is TRUE.

`RR` If `RR_studies` is TRUE, a data frame containing information for all RR and SIR studies with following columns where each row corresponds to a single study: for studies that do not provide age-related summaries or if age related column not provided, default values will be used

R.est	reported RR/SIR value.
RR.ci.lower	lower 95% confidence limit.
RR.ci.upper	upper 95% confidence limit.
A	mean age of onset for carriers. Default is 63.
V	sd of age of onset for carriers. Default is 14.00726.
A.lo	minimum age of among for carriers. Default is 20.
A.hi	maximum age of among for carriers. Default is 95.
A0	mean age of onset for non-carriers. Default is 63.
V0	sd of age of onset for non-carriers. Default is 14.00726.
A0.lo	minimum age of onset for non-carriers. Default is 20.
A0.hi	maximum age of onset for non-carriers. Default is 95.

`OR_studies` A logical variable to indicate whether studies reporting OR are included in the meta analysis. Default is TRUE.

`OR` If `OR_studies` is TRUE, a data frame containing information for all OR studies with following columns where each row corresponds to a single study: for studies that do not provide age-related summaries or if age related column not provided, default values will be used

OR.est	reported OR value.
OR.ci.lower	lower 95% confidence limit.
OR.ci.upper	upper 95% confidence limit.
A	mean age of onset for cases. Default is 63.
V	sd of age of onset for cases. Default is 14.00726.
A.lo	minimum age of onset among cases. Default is 20.
A.hi	maximum age of onset among cases. Default is 95.
A0	mean age of controls (at study inclusion). Default is value A.
V0	sd of age of controls (at study inclusion). Default is value V.
A0.lo	minimum age of controls (at study inclusion). Default is value A.lo.
A0.hi	maximum age of controls (at study inclusion). Default is value A.hi.

`zero_studies` A logical variable to indicate whether information from case control studies where no mutations were detected in controls are included in the meta analysis. Default if TRUE.

`zero_OR` If `zero_studies` is TRUE, a data frame containing information for such studies with following columns where each row corresponds to a single study: for studies that do not provide age-related summaries or if age related column not provided, default values will be used

carrier.cases	number of carrier cases.
non_carrier.cases	number of non-carrier cases.
non_carrier.controls	number of non-carrier controls.
A.lo	minimum age of onset among cases. Default is 20.
A.hi	maximum age of onset among cases. Default is 95.
A0	mean age of controls (at study inclusion). Default is value A.
V0	sd of age of controls (at study inclusion). Default is value V.
A0.lo	minimum age of controls (at study inclusion). Default is value A.lo.
A0.hi	maximum age of controls (at study inclusion). Default is value A.hi.

`ages` Ages at which penetrance values are required. Default is 20:85 at increments of 1. Maximum possible age is 85

`pl` If pl=TRUE, returns a plot of the estimated age specific penetrance value vs age with confidence intervals. Default is FALSE.

`ylim` If pl=TRUE, numeric vector of length 2, giving the y coordinates. Default is c(0,1).

`xlim` If pl=TRUE, numeric vector of length 2, giving the x coordinates. Default is c(20,85).

Details

The `BayesMeta` function estimates consensus age-specific risks of developing cancer associated with mutations in a specific gene, i.e., penetrance estimation. A meta-analysis approach based on a Bayesian hierarchical random-effects model is used to obtain penetrance estimates integrating studies reporting different types of risk measures (e.g., penetrance, relative risk, odds ratio, and standard incidence ratio) while accounting for associated uncertainties.

The cumulative penetrance$F_s(t|\kappa_s, \lambda_s)$ at age t for study s is assumed to be given by the c.d.f. of a Weibull distribution with shape parameter κ_s and scale parameter λ_s.

The prior distributions are $\pi(\kappa_s|a, b) = \text{Gamma}(a, b),\ \pi(\lambda_s|c, d) = \text{Gamma}(c, d),$ where a and c are shape parameters and b and d are scale parameters.

Continuous uniform distributions are assumed for all the hyper-parameters. Specifically, $\pi(a|l_a, u_a) = \text{U}(l_a, u_a),\ \pi(b|l_b, u_b) = \text{U}(l_b, u_b),\ \pi(c|l_c, u_c) = \text{U}(l_c, u_c),$ and $\pi(d|l_d, u_d) = \text{U}(l_d, u_d)$ with $l_a, u_a, l_a, u_a, l_a, u_a, l_a, u_a$ pre-specified.

Posterior distributions are obtained via a Markov chain Monte Carlo algorithm, that employ a Metropolis-Hastings algorithm within Gibbs sampling. The default values for age distributions are based on the distribution of age of onset of breast cancer for US general population (https://seer.cancer.gov/statfacts/html/breast.html). Credible interval at a given age is the 0.025^{th} and 0.0975^{th} quantiles of the posterior dstribution of penetrance estimate at that age.

Value

`BayesMeta` returns a list of two objects— the estimated penetrance values and 95% credible interval.

`penetrance` Estimated age-specif penetrance values

`penetrance_CI` Credible interval

Author(s)

Thanthirige Lakshika Maduwanthi Ruberu

References

1. Ruberu, T. L. M, Braun, D, Parmigiani, G. and Biswas, S. Bayesian Meta-Analysis of Penetrance and its Application to Breast Cancer Risk among ATM Mutation Carriers. (Manuscript)
2. Marabelli, M., Cheng, S. C., and Parmigiani, G. (2016). Penetrance of ATM Gene Mutations in Breast Cancer: A Meta-Analysis of Different Measures of Risk. Genetic Epidemiology, 40, 425-431.

Examples

```
#penetrance value
set.seed(123)

penet1=c(0.134,0.230,0.400,0.571,0.706) #penetrance values from study 1
penet2=c(0.092,0.146, 0.260,0.403,0.556) #penetrance values from study 2
penet3=c(0.034,0.072,0.131, 0.243,0.370) #penetrance values from study 3
penet4=c(0.082,0.160,0.256, 0.422 ,0.588) # #penetrance values from study 4

penet1_ci_low=c( 0.113, 0.203,0.369,0.539,0.675) #lower 95% confidence limit of values in penet1
penet1_ci_hi=c(0.155, 0.256,0.430,0.601,0.734) #upper 95% confidence limit of values in penet1

penet2_ci_low=c( 0.066,0.114, 0.221 , 0.359,0.508) #lower 95% confidence limit of values in penet2
penet2_ci_hi=c(0.117,0.176,0.298,0.446,0.600) #upper 95% confidence limit of values in penet2

penet3_ci_low=c( 0.023,0.056,0.110,0.216,0.338)#lower 95% confidence limit of values in penet3
penet3_ci_hi=c(0.045,0.088,0.152,0.270,0.402) #upper 95% confidence limit of values in penet3

penet4_ci_low=c(0.058,0.127,0.217,0.377,0.540)#lower 95% confidence limit of values in penet4
penet4_ci_hi=c(0.106 ,0.192,0.294,0.465,0.632) #upper 95% confidence limit of values in penet4


ages_penet1=c(35,40,50,60,70) # ages corresponding to values in penet 1
ages_penet2=c(40,50,60,70,80) # ages corresponding to values in penet 2
ages_penet3=c(35,40,50,60,70) # ages corresponding to values in penet 3
ages_penet4=c(40,50,60,70,80) # ages corresponding to values in penet 4


study_number=c(rep(1,5),rep(2,5),rep(3,5),rep(4,5)) # 1 for 5 records of study 1, 2 for study 2 ...

penet=data.frame('penetrance'=c(penet1,penet2,penet3,penet4),
            'penet_ci_lower'=c(penet1_ci_low,penet2_ci_low,penet3_ci_low,penet4_ci_low),
                'penet_ci_upper'=c(penet1_ci_hi,penet2_ci_hi,penet3_ci_hi,penet4_ci_hi),
                'ages_penet'=c(ages_penet1,ages_penet2,ages_penet3,ages_penet4),
                'study_label'=study_number)


## RR values
R.est=c(4.71,4.31,3.31, 3.27) #RR or SIR values reported by each study
RR.ci.lower=c(3.80,3.69,2.87,2.97) #corresponding 95% lower limit
RR.ci.upper=c(5.84,5.04,3.82,3.60) #corresponding 95% upper limit


#Age related summaries from each study
```

```
A=c(60.21,64.88,65.81,63)
V=c(17.34, 15.07, 15.18, 14.01)
A.lo=c(20, 20, 20, 20)
A.hi=c(95,95,95,95)
A0=c(67.61, 67.34, 68.17, 67.18)
V0=c( 14.33, 16.30, 15.18, 14.01)
A0.lo=c(20, 20, 20, 20)
A0.hi=c(95,95,95,95)



RR=data.frame(R.est,RR.ci.lower,RR.ci.upper,A,V,A.lo,A.hi,A0,V0,A0.lo,A0.hi)
colnames(RR)=c("R.est","RR.ci.lower","RR.ci.upper","A","V","A.lo","A.hi","A0","V0","A0.lo","A0.hi")


###OR values
OR.est=c(6.25, 8.91, 22.97) #OR value reported by each study
OR.ci.lower=c( 1.38,4.89, 3.08 ) #corresponding 95% lower limit
OR.ci.upper=c(28.23,16.24, 171.05) #corresponding 95% lower limit

#Age related summaries from each study
A=c(66.69, 66.58, 63.00)
V=c(15.34, 16.32, 14.01)
A.lo=c(20, 20, 20)
A.hi=c(95,95,95)
A0=c(66.69, 66.58, 63.00)
V0=c(15.34, 16.32, 14.01)
A0.lo=c(20, 20, 20)
A0.hi=c(95,95,95)

OR=data.frame(OR.est,OR.ci.lower,OR.ci.upper,A,V,A.lo,A.hi,A0,V0,A0.lo,A0.hi)
colnames(OR)=c("OR.est","OR.ci.lower","OR.ci.upper","A","V","A.lo","A.hi","A0","V0","A0.lo","A0.hi")


#######studies with no mutations in controls
carrier.cases=c(1)

non_carrier.cases=c(99)

non_carrier.controls=c(100)

#Age related summaries from each study
A=c(48)
V=c(13.57)
A.lo=c(25)
A.hi=c(78)
A0=c(48)
V0=c(13.57)
A0.lo=c(25)
A0.hi=c(78)

zero_OR=data.frame(carrier.cases,non_carrier.cases,non_carrier.controls,
          A,V,A.lo,A.hi,A0,V0,A0.lo,A0.hi)
colnames(zero_OR)=c("carrier.cases","non_carrier.cases","non_carrier.controls",
          "A","V","A.lo","A.hi","A0","V0","A0.lo","A0.hi")
```

```
BayesMeta(penet,RR_studies=TRUE,RR=RR,OR_studies=TRUE,OR=OR,ages=20:85,
        zero_studies=TRUE,zero_OR=zero_OR,pl=TRUE ,ylim=c(0,1),xlim=c(20,85))
```