**Twin Research and Human Genetics**

Two SNPs Associated With Spontaneous DZ Twinning: Effect Sizes and How We Communicate Them

H. Mbarek, C.V. Dolan, & D.I. Boomsma

Supplemental material R script (Tables 1 and 3).

# R code 1

# to obtain Table 1 and Table 3. The required input is shown in bold type face

rm(list=ls(all=TRUE))

f1=function(par, N=N, pa=pa, pdz=pdz, ab=ab, cd=cd, ef=ef, ace=ace, bdf=bdf, relrisk=relrisk) {

#

a=par[1]; b=par[2]; c=par[3]; d=par[4]; e=par[5]; f=par[6]

p=rep(0,8)

p[1]=(ab-(a+b))

p[2]= (cd-(c+d) )

p[3]= (ef-(e+f) )

p[4]= (ace-(a+c+e))

p[5]= (bdf-(b+d+f))

p[6]= ( (c/(c+d))/(e/(e+f))-relrisk)

p[7]= ( (a/(a+b))/(c/(c+d))-relrisk)

p[8]=sum(par)-N

f=sum(p^2) # least squares

f

}

**# input --------------------------------------------------------------- input**

**raf=pa=.85 # risk allele probability FB; .24 for S3**

**relrisk=1.18 # allelic relative risk associated with risk increasing allele FB; 1.09 for S3**

**pdz=.0107 # spontaneous DZ twinning prevalence NL 1911 (Glasner et al)**

**labels=c("GG","GA","AA") # genotype labels; G = risk increasing allele FB; (CC, CT, TT for S3)**

**# end input --------------------------------------------------------- end input**

#

#

N=10000 # arbitrary

qa=1-pa #

qdz=1-pdz #

ab= N\*pa^2

cd= N\*pa\*2\*qa

ef= N\*qa^2

ace=N\*pdz

bdf=N\*qdz

## starting values based on independence

a=N\*pa^2\*pdz # starting values

c=N\*2\*pa\*qa\*pdz

e=N\*qa^2\*pdz

b=N\*pa^2\*qdz

d=N\*2\*pa\*qa\*qdz

f=N\*qa^2\*qdz

par=c(a,b,c,d,e,f)

#

res=optim(par,f1, gr=NULL, method = c("L-BFGS-B"),lower =.00001,upper=N, hessian=TRUE,

N,pa, pdz, ab, cd, ef, ace, bdf, relrisk)

#

ch1=sum(eigen(res$hessian)$value>0)

ch2=res$convergence

mess=" optim not OK: check optim output ...."

if (ch1==6 & ch2==0) mess= ' optim OK'

# table of probabilities 3 (genotypes) x2 (outcome)

ptab=matrix(0,4,3)

ii=0

for (i in 1:3) { for (j in 1:2) {ii=ii+1; ptab[i,j]=res$par[ii]/N}}

# marginals

ptab[4,1:2]=apply(ptab[1:3,1:2],2,sum)

ptab[1:3,3]= apply(ptab[1:3,1:2],1,sum)

ptab[4,3]=sum(res$par)/N

#

or=matrix(0,3,1) # odds ratios

rr=matrix(0,3,1) # relative risk

cp=matrix(0,3,1) # conditional probs outcome|genotype

refgen=3

for (i in 1:3) { cp[i] =ptab[i,1]/ptab[i,refgen] }

for (i in 1:3) {

or[i]=(ptab[i,1]/ptab[i,2]) / (ptab[refgen,1]/ptab[refgen,2])

rr[i]= (ptab[i,1]/(ptab[i,1]+ptab[i,2])) / (ptab[refgen,1]/(ptab[refgen,2]+ptab[refgen,1]))

}

rownames(ptab[1:3,])=row.names(cp)=row.names(rr)=row.names(or)=labels

colnames(ptab)=c('y','no','marg')

rownames(ptab)=c(labels,'marg')

print(mess)

print(c('reference genotype',labels[refgen]))

print(' table genotype X outcome ')

print(ptab)

print(' relative risk ')

print(round(rr,4))

print('odd ratio')

print(round(or,4))

print('conditional DZ given genotype')

print(round(cp,4))

Supplemental material: R script (Table A2)

# R code 2: construct the approxiamate haplotype by twinning probability table using Bayes' theorem

# Used to calculate Table A2.

rm(list=ls(all=TRUE))

**# -----------------------------------input**

**pdz=.0107 # spontaneous twinning prob**

**pa1=.85; # FB raf**

**pa2=.24; # S3 raf**

**# # genotype x twinning table (FB)**

**gv1=matrix(c(**

**0.008097002, 0.7144030, 0.7225,**

**0.002421898, 0.2525781, 0.2550,**

**0.000181100, 0.0223189, 0.0225,**

**0.010700000, 0.9893000, 1.0000**

**),4,3,byrow=T)**

**# # genotype x twinning table (S3)**

**gv2=matrix(c(**

**0.0007016085, 0.05689839, 0.0576,**

**0.0040766056, 0.36072339, 0.3648,**

**0.0059217860, 0.57167821, 0.5776,**

**0.0107000001, 0.98930000, 1.0000**

**),4,3,byrow=T)**

**# ------------------------------------------------- end input**

**# prob(genotype | outcome)**

cgv1=matrix(0,3,3)

cgv1[1:3,1]=gv1[1:3,1]/gv1[4,1]

cgv1[1:3,2]=gv1[1:3,2]/gv1[4,2]

cgv2=matrix(0,3,3)

cgv2[1:3,1]=gv2[1:3,1]/gv2[4,1]

cgv2[1:3,2]=gv2[1:3,2]/gv2[4,2]

#

qdz=1-pdz

cgv=matrix(0,9,4)

ii=0

for (i in 1:3) { for (j in 1:3) { ii=ii+1

cgv[ii,1]=cgv1[i,1]\*cgv2[j,1] # approx prob(haplotype | outcome)

cgv[ii,2]=cgv1[i,2]\*cgv2[j,2]

}}

for (i in 1:9) {

cgv[i,3]=(cgv[i,1]\*pdz)/(cgv[i,1]\*pdz+cgv[i,2]\*qdz) # bayes prob(outcome | haplotype)

cgv[i,4]=(cgv[i,2]\*qdz)/(cgv[i,1]\*pdz+cgv[i,2]\*qdz)

}

qa1=1-pa1

qa2=1-pa2

happrob=c(

pa1^2\*(pa2^2),pa1^2\*(2\*pa2\*qa2),pa1^2\*(qa2^2),

(pa1\*qa1\*2)\*(pa2^2),(pa1\*qa1\*2)\*(2\*pa2\*qa2),(pa1\*qa1\*2)\*(qa2^2),

qa1^2\*(pa2^2),qa1^2\*(2\*pa2\*qa2),qa1^2\*(qa2^2))

ptab=matrix(0,10,3)

ptab[1:9,3]=happrob

ptab[1:9,1]=cgv[1:9,3]\*ptab[1:9,3] # haplotype x twinning probability table

ptab[1:9,2]=cgv[1:9,4]\*ptab[1:9,3]

ptab[10,1]=sum(ptab[1:9,1])

ptab[10,2]=sum(ptab[1:9,2])

ptab[10,3]=sum(ptab[1:9,3])

ptab=round(ptab,7)

lab1=c("GG","GA","AA") # add genotype designation

lab2=c("CC","CT","TT")

lab=rep(NA,9); ii=0

for (i in 1:3){ for (j in 1:3) {ii=ii+1; lab[ii]=paste(lab1[i],lab2[j],sep="\_")}}

rownames(ptab)=c(lab,'marg')

colnames(ptab)=c('yes','no','marg')

print(ptab)