Risk in relatives, heritability, SNP-based heritability and genetic correlations in psychiatric disorders: a review

04 August, 2020

Introduction

This supplementary note provides more details on the content discussed in the paper "*Risk in relatives, heritability, SNP-based heritability and genetic correlations in psychiatric disorders: a review*", by Baselsman, Yengo, van Rheenen & Wray published in *Biological Psychiatry.* If you use this code, please cite that paper. In the .html version of this document click on "Code"" to expand to see the R code. "There is an accompanying ShinyApp, CHARRGE (Calculating Heritabilities and Relative Risks and GEnetic correlations). The Rmarkdown version of this file as well as the code for the ShinyApp can be found at github

links:
ShinyApp: https://shiny.cnsgenomics.com/CHARRGe/
Github: https://github.com/BartBaselmans/CHARRGe)

Heritability

Parameters for psychiatric disorders

In Figure 1 of the paper we show estimates of the population lifetime risk (K) and the risk in first degree relatives of those with one affected parent (K_1), and the risk ratio ($RR_1 = \lambda_1 = K_1/K$) for common psychiatric disorders. Although K and the risk in relatives of different types (K_R) are measurable, often only the heritability estimated from one or more types of relatives are reported. Heritability is defined as the proportion of variance in liability attributed to additive genetic factors. Here, we present the K and heritabilities for the major psychiatric disorders (see Supplementary Table 1 for references). The estimates using twin data are usually higher than those estimated from family data (different types of relatives) although the latter are often based on larger sample sizes. Here, we use "round" h^2 numbers (approximate average of h^2_{twin} and h^2_{family}) in subsequent calculations. We also provide SNP-based heritability (h^2_{SNP}) estimates (see section 3), where SNP-based heritability, by definition is smaller than the estimates based on family records, as it tracks only the proportion of variance attributable to additive genetic values tagged by common SNPs (or other measurable common DNA variants). In contrast, total heritability tracks the contribution to variance of genetic variants across the allelic frequency spectrum.

Phenotype	h^2_{twin} (%)	h^2_{family} (%)	h^2_{round} (%)	$h^2_{SNP(sBayesS)}$ (%)	$h^2_{SNP(ldsc)}$ (%)	K
Schizophrenia	81	64	70	30	26	0.01
Bipolar Disorder	75	59	65	22	20	0.01
Major Depressive Disorder	37	32	35	9	10	0.15
ADHD	75	-	75	24	21	0.05
Anorexia Nervosa	56	43	50	15	14	0.01

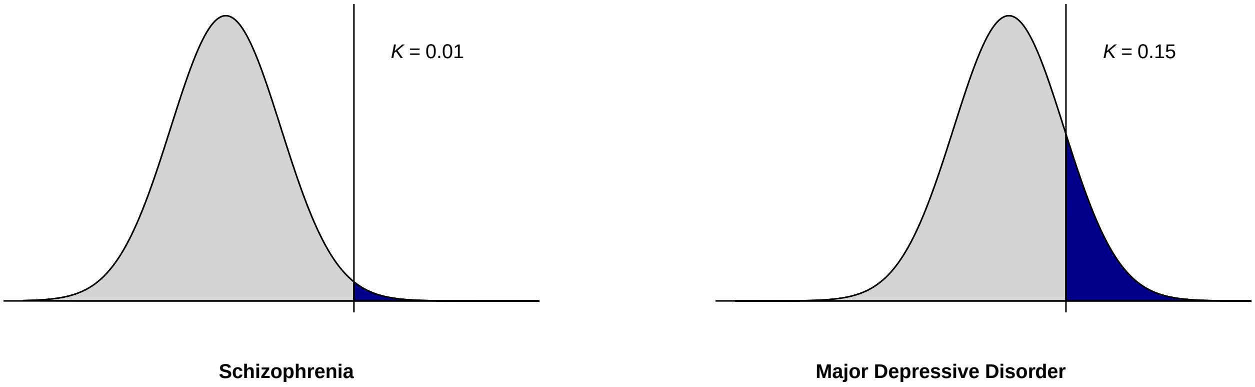
Phenotype	h^2_{twin} (%)	h^2_{family} (%)	h^2_{round} (%)	$h^2_{SNP(sBayesS)}$ (%)	$h^2_{SNP(ldsc)}$ (%)	K
Autism Spectrum Syndrome	80	85	80	11	12	0.01

h^2_{round} is an approximate average of h^2_{family} and h^2_{twin}.
See Supplementary Table 1 for the references from which these estimates were derived.

Liability Threshold model

When many factors contribute to the risk of disease it can be helpful to consider a model of disease where there is a latent distribution of liability to disease representing the dichotomous Case/Control status.(e.g. Falconer, 1965 or Reich 1972). Since this latent liability comprises many genetic and other risks it is reasonable to assume that the liability distribution is approximately normally distributed (since many things added together will make a bell-shaped distribution in a population sample) and that those affected by disease have the combination of genetic and other factors that it places them in the top end of the liability distribution.

```
h2x = 0.7
h2y = 0.35
Kx = 0.01
Ky = 0.15
disorder1 = "Schizophrenia"
disorder2 = "Major Depressive Disorder"

plot_liability_model <- function(h2x,h2y,Kx,Ky, disorder1, disorder2){

Tx = -qnorm(Kx, 0,1)
Ty = -qnorm(Ky, 0,1)

Ks <- c(Kx,Ky)  #Lifetime risk
Heritability <- c(h2x,h2y)
Threshold <-c(Tx,Ty)
disorder <-c(disorder1,disorder2)

layout(matrix(c(1,2),1, 2, byrow = TRUE))
for(j in 1:2){

h2 <- c(as.expression(bquote(italic(h)^2~'='~.(Heritability[j]))),
                  as.expression(bquote(italic(h)^2~'='~.(Heritability[j]))))
K_pop <- c(as.expression(bquote(italic(K)~'='~.(Ks[j]))),
                    as.expression(bquote(italic(K)~'='~.(Ks[j]))))

  T0 = Threshold[j]
  z = dnorm(T0)
  i = z / Ks[j] # mean phenotypic liability of those with disease
  mean = -(T0) ;sd=1
  lb=0; ub=4

  x <- seq(T0-6,T0+3,length=1000)*sd + mean
  hx <- dnorm(x,mean,sd)

  plot(x, hx, type="n", xlab="", ylab="",
       main="", axes=FALSE)

  i <- x >= lb & x <= ub
  l <- x <= lb & x <= ub
  lines(x, hx)
  polygon(c(lb,x[i],ub), c(0,hx[i],0), col="darkblue")
  polygon(c(-5,x[l],lb), c(0,hx[l],0), col="lightgrey")


  #axis(1, at=seq(-5, 2, 1), pos=0, cex.axis =1)
  abline(h=0)
  abline(v=0,h=-2)

  mtext(K_pop[j], side = 3, line = -2.1,at = 2,font = 2, cex = 0.8, adj=1)
  mtext(disorder[j], side = 1, line = 1.5,at = 0,font = 2, cex = 0.8, adj=1)
  }
}

plot_liability_model(h2x,h2y,Kx,Ky, disorder1, disorder2)
```



Hence, those whose liability to disease is above the threshold are affected, and so this model is sometimes called the liability threshold model. The threshold can be calculated using the lifetime risk (K) parameter as K represents the proportion of individuals in the general population have the disorder of interest. The threshold (or boundary) that determines the area of K (indicated in blue) in a normal distribution is given by:

$$T = \Phi^{-1}(1 - K)$$

Conversely, the life-time risk (K) is derived from the threshold parameter

$$K = 1 - \Phi(T)$$

where $\Phi^{-1}(x)$ is the inverse of the cummulative standard normal distribution function.

An example with R code using lifetime risk $K = 0.15$:

```
K = 0.15
Threshold_prevalence <- function(K){
Tx = -qnorm(K,0,1)
Kx = 1-pnorm(Tx,0,1)

cat("Output Parameters:\n")
cat("----------------------------\n")
cat("Tx: ", round(Tx,2),"\n")
cat("Kx: ", round(Kx,2),"\n")
cat("----------------------------\n")
}
Threshold_prevalence(K)
```

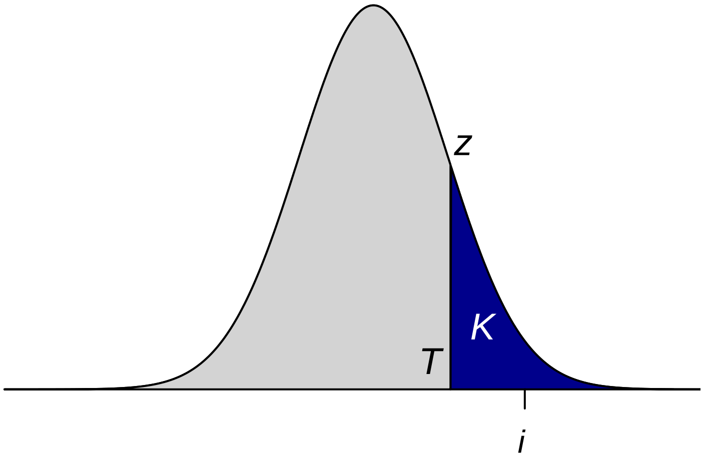
```
## Output Parameters:
## ----------------------------
## Tx:  1.04
## Kx:  0.15
## ----------------------------
```

Risk in relatives and heritability of liability

The lifetime risk of a disease in relatives (K_R) of those affected by the disease is expected to be greater, or equal to, the lifetime risk of the disease in a population sample. It is logical to assume that the threshold in liability associated with disease has the same value in the relatives of those affected. As a result, the liability distribution in first degree members must be shifted (in the direction of increased liability) compared to the general population to an extent consistent with the observed higher risk of K_R.

We can then ask: what proportion of the variation in liability must be attributable to genetic factors (i.e., what is the heritability?) in order to generate this observed increased risk in relatives, given the known coefficient of relationship between the relatives, a_R, i.e., 0.5 for 1st degree relatives. First, we introduce some properties of the normal distribution.

```
h2x = 0.7
Kx = 0.15

plot_liability_model <- function(h2x,Kx){

  Tx = -qnorm(Kx, 0,1)


  Ks <- c(Kx)  #Lifetime risk
  Heritability <- c(h2x)
  Threshold <-c(Tx)
  disorder <-c(disorder1)

  #layout(matrix(c(1,2),1, 2, byrow = TRUE))

    h2 <- c(as.expression(bquote(italic(h)^2~'='~.(Heritability))))

    K_pop <- c(as.expression(bquote(italic(K)~'='~.(Ks))))

    T0 = Threshold
    z = dnorm(T0)
    i = z / Ks # mean phenotypic liability of those with disease
    mean = -(T0) ;sd=1
    lb=0; ub=4

    x <- seq(T0-6,T0+3,length=1000)*sd + mean
    hx <- dnorm(x,mean,sd)

    plot(x, hx, type="n", xlab="", ylab="",
         main="", axes=FALSE)

    i <- x >= lb & x <= ub
    l <- x <= lb & x <= ub
    lines(x, hx)
    polygon(c(lb,x[i],ub), c(0,hx[i],0), col="darkblue")
    polygon(c(-5,x[l],lb), c(0,hx[l],0), col="lightgrey")



    mtext(bquote(italic("K")), side = 1, line = -2.5,at = 0.6,font = 2, cex = 1.2, adj=1, col = "white"
    mtext(bquote(italic("T")), side = 1, line = -1.6,at = -0.12,font = 2, cex = 1.2, adj=1, col = "blac
    mtext(bquote(italic("z")), side = 1, line = -7.3,at = 0.3,font = 2, cex = 1.2, adj=1)
    mtext(bquote(italic("i")), side = 1, line = 0.45,at = 1,font = 2, cex = 1, adj=1)
    xtick <- seq(1,1, by =1)
    axis(side=1, at = xtick, labels =F,pos = 0)
}

plot_liability_model(h2x,Kx)
```



We define z as the height of the standard normal curve at threshold T, that corresponds to the proportion K, and $i\sigma_p$ is the mean phenotypic liability of those with disease. Since the phenotypic variance of liability (σ_p^2) is by definition 1, the mean phenotypic liability is simply i.From mathematical propertis of the normal distribution $i = z/K$.

The mean liability of relatives of those ascertained to have disease, accounts for the coefficient of relationship between the relatives and the proportion of variance that is genetic: $a_R i h^2$. Hence, for a trait with no genetic contribution (h^2=0), the mean liability in relatives is the same as in the general population. Overall, the shift in mean liability between population and 1st degree relatives is higher for diseases with higher heritability and for diseases that are less common.

The difference in mean liabilities between population ($M_{pop} = 0$) and relatives (M_R), can be shown to be equivalent to the difference in thresholds when the liability distribution of relatives is scaled back to a N(0,1) distribution. i.e.,M_R - $M_{pop} = a_R i h^2 = T_R - T$, where $T_R = \Phi^{-1}(1 - K_R)$ Falconer (1965). Hence, $h^2 = \frac{T - T_R}{a_R i}$, i.e. everything on the right hand side of the equation is derived from measurable statistics of K, K_R and a_R.

However, Falconer's derivation assumed that the genetic (and hence liability) variance amongst the relatives was not changed by ascertainment on disease in the probands. Reich et al.,(1972) showed that while it is OK to assume that the distribution of disease liability in relatives of those affected is approximately normal, it should be recognised that the variance in liability is slightly reduced in the relatives as a result of ascertaining the proband as been affected. They showed that the variance in liability is reduced by a factor of $a_R^2h^4i(i-T)$ so that the scaling of the T_R to a standard N(0,1) distribution has a denominator of the standard deviation of liability in the relatives given that the probands have disease:

$$T_R = \frac{T - a_R i h^2}{\sqrt{1 - a_R^2 h^4 i(i-T)}}$$

Since heritability is in the numerator and denominator, making h^2 the subject of the equation requires solving of a quadratic equation to give:

$$h^2 = \frac{T - T_R\sqrt{1-(1-\frac{T}{i})(T^2-T_R^2)}}{a_R(i+(i-T)T_R^2)}$$

Although, this equation looks complicated, in fact everything on the right hand side again can be calculated from two observations K and K_R, and a_R (and holds as long as $a_R < 1$), and if ascertainment is based on the disease status of one proband. A special case is when ascertainment is based on both parents being affected, hence the K and K_{2PAR} is estimated, with risk ratio $RR_{2PAR} = \frac{K_{2PAR}}{K}$ In this case it can be shown in Wray & Gottesman (2012) that

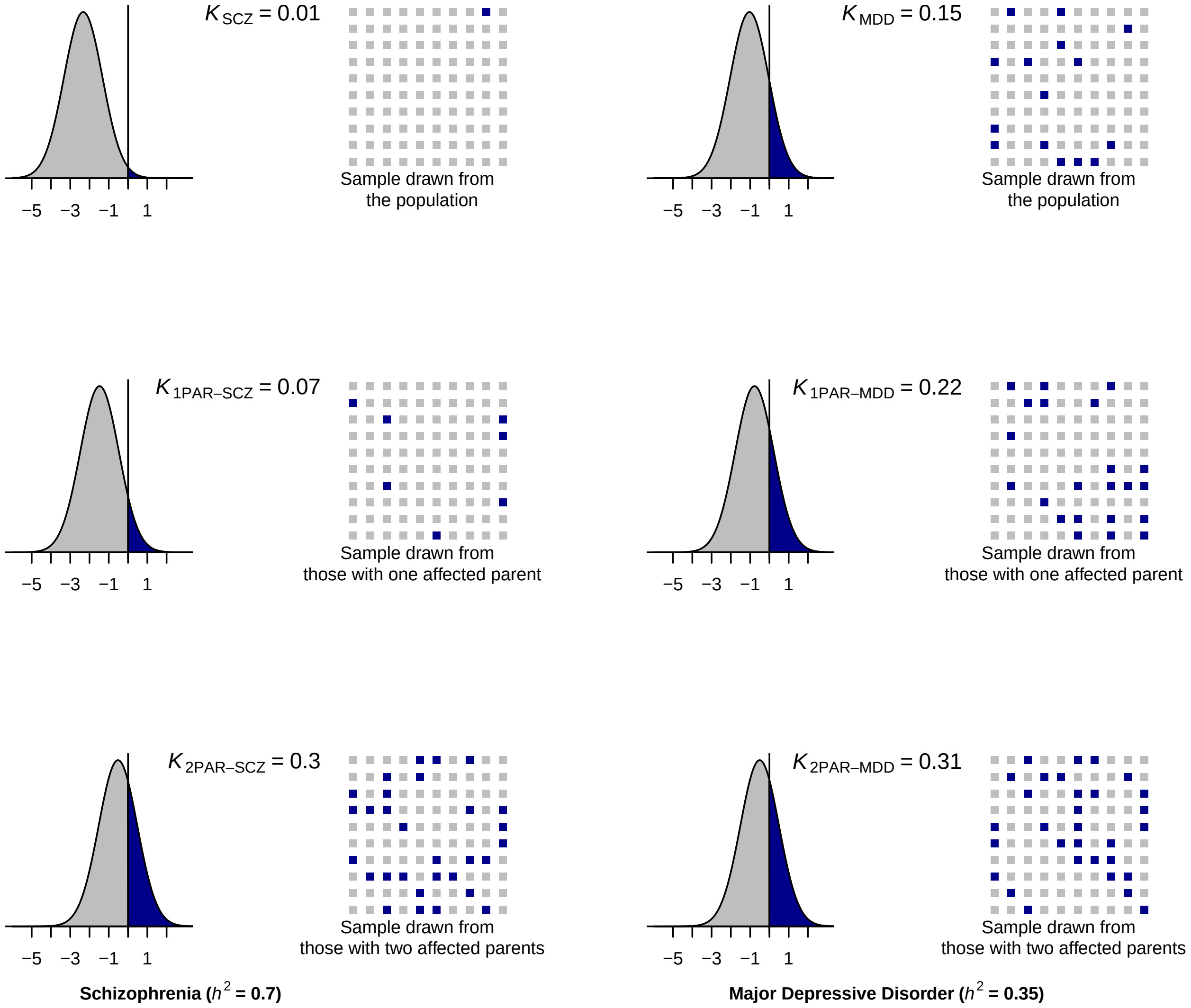
$$T_{2PAR} = \frac{T - ih^2}{\sqrt{1 - 0.5h^4 i(i-T)}}$$

and solving this quadratic equation gives:

$$h^2 = 2T - \sqrt{2}T_{2PAR}\frac{\sqrt{2-(T^2-T_{2PAR})(1-\frac{T}{i})}}{2i+(i-T)T_{2PAR}}$$

To visualize the increase in liability for relatives having one or two affected parents consider examples of schizophrenia with $K = 0.01$, $h^2 = 0.7$ or major depressive disorder with $K = 0.15$, $h^2 = 0.35$ a sample of 100 individuals drawn from the general population, and a sample of 100 children who have one affected parent, or a sample of 100 children who have two affected parent(s).

```
h2x = c(0.7,0.35)
Kx = c(0.01,0.15)
a = 0.5
disorder = c("Schizophrenia","Major Depressive Disorder")


###################################################################################################
liability <- c()
liability_function <- function(h2x, Kx, a, disorder){

  Tx    = -qnorm(Kx, 0,1)
  z    = dnorm(Tx)
  i    = z/Kx
#One affected parent
  Tx1   = (Tx - a * i * h2x) / (sqrt(1 - a * a * h2x * h2x * i * (i-Tx)))
  Kx1   = 1 - pnorm(Tx1)
  RRx1   = Kx1 / Kx
#Two affected parents
  Tx2   = (Tx -i*h2x) / sqrt(1-(0.5*h2x*h2x*i)*(i-Tx))
  Kx2   = 1 - pnorm(Tx2)
  RRx2   = Kx2 / Kx

  x <- as.data.frame(c(disorder,Tx,Tx1, Tx2, Kx, Kx1, Kx2, NA,RRx1, RRx2 ))
  return(x)
}

#Run function for number of included disorders
for(i in 1:length(disorder)){
liability[i] <- liability_function(h2x[i], Kx[i], a, disorder[i])
}

#convert list to dataframe
df <- data.frame(matrix(unlist(liability), nrow=length(liability), byrow=T), stringsAsFactors = F)

#Threshold extraction
Threshold <- df[2:4]
Threshold_row <-cbind(Threshold[1,], Threshold[2,])
Threshold_row  <- as.numeric(Threshold_row)

# Prevalence extraction
K_prevalence <- df[5:7]
K_prevalence_row <- as.numeric(cbind(K_prevalence[1,], K_prevalence[2,]))
K100_prevalence_row <- round((K_prevalence_row*100),0)

#Make plot
layout(matrix(c(1,2,3,7,8,9,4,5,6,10,11,12), 3, 4, byrow = F))
par(mar=c(5.1, 4.1, 4.1, 2.2))
for(j in 1:6){

#Normal distribution Threshold
  Ks <- as.matrix(sapply((K100_prevalence_row/100), as.numeric))
  K_print <- c(as.expression(bquote(italic(K)[SCZ]~'='~.(Ks[j]))),
               as.expression(bquote(italic(K)[1][PAR][-SCZ]~'='~.(Ks[j]))),
               as.expression(bquote(italic(K)[2][PAR][-SCZ]~'='~.(Ks[j]))),
               as.expression(bquote(italic(K)[MDD]~'='~.(Ks[j]))),
               as.expression(bquote(italic(K)[1][PAR][-MDD]~'='~.(Ks[j]))),
               as.expression(bquote(italic(K)[2][PAR][-MDD]~'='~.(Ks[j]))))

  heritability1 <- c(paste0("= ",h2x[1]))
  Phenotype1<- c("","",as.expression((bquote(bold(bold(.(disorder[1])~'('*bold(italic(h)^2)~.(heritabil
  heritability2 <- c(paste0("= ",h2x[2]))
    Phenotype2 <- c("","","","","",as.expression((bquote(bold(bold(.(disorder[2])~'('*bold(italic(h)^2)

  Txp = Threshold_row[j]
  z = dnorm(Txp)
  i = z / Ks[j]
  mean = -(Txp) ;sd=1
  lb=0; ub=4

  x <- seq(Txp-6,Txp+3,length=1000)*sd + mean
  hx <- dnorm(x,mean,sd)

  plot(x, hx, type="n", xlab="", ylab="",
       main="", axes=FALSE)
  i <- x >= lb & x <= ub
  l <- x <= lb & x <= ub
  lines(x, hx)
  polygon(c(lb,x[i],ub), c(0,hx[i],0), col="darkblue")
  polygon(c(-5,x[l],lb), c(0,hx[l],0), col="grey")

  axis(1, at=seq(-5, 2, 1), pos=0, cex.axis =1)
  abline(h=0)
  abline(v=0,h=-2)

  mtext(K_print[j], side = 3, line = -1.1,at = 10,font = 2, cex = 0.8, adj=1)
  mtext(Phenotype1[j], side = 3, line = -12,at = -2.5,font = 2, cex = 0.7, adj = 0)
  mtext(Phenotype2[j], side = 3, line = -12,at = -2.1,font = 2, cex = 0.7, adj = 0)

}

############################
##########################

#Square indicating number of affected individuals
result <- matrix(nrow = 6, ncol = 100)
#layout(matrix(c(1,2,3,4,5,6), 3, 2, byrow = F))
for(j in 1:6){
  result[j,] <- sample(as.matrix(c(rep("b",K100_prevalence_row[j]),rep("a",100-K100_prevalence_row[j]))


  #sum(result[1,]=="a")
  # input parameters - nr * nc should equal length(x)
  cols <- c("grey", "darkblue")
  nr <- 10
  nc <- 10

  # create data.frame of positions and colors
  m <- matrix(cols[factor(result[j,])], nr, nc)
  DF <- data.frame(row = c(row(m)), col = c(col(m)[, nc:1]), value = c(m), gender =  sample(c(rep("male
                   stringsAsFactors = FALSE)

  title_plot <- c("Sample drawn from \n the population", "Sample drawn from \n those with one affected

  plot(col ~ row, DF, col = DF$value, pch = 15, cex = 1, asp = 1,
       xlim = c(0, nr), ylim = c(0, nc),
       axes = FALSE, xlab = "", ylab = "")
  title(xlab=title_plot[j], line=0.2, cex.lab=1)
}
```



For the major psychiatric disorders the increased risk ratio are:

```
h2x = c(0.70,0.65,0.35,0.75,0.5,0.8)
Kx = c(0.01,0.01,0.15,0.05,0.01,0.01)
a = 0.5
disorder = c("Schizophrenia","Bipolar Disorder", "Major Depressive Disorder", "ADHD", "Anorexia Nervosa
liability <- c()
liability_function <- function(h2x, Kx, a, disorder){

  Tx    = -qnorm(Kx, 0,1)
  z    = dnorm(Tx)
  i    = z/Kx
  #One affected parent
  Tx1   = (Tx - a * i * h2x) / (sqrt(1 - a * a * h2x * h2x * i * (i-Tx)))
  Kx1   = 1 - pnorm(Tx1)
  RRx1   = round(Kx1 / Kx,1)
  #Two affected parents
  Tx2   = (Tx -i*h2x) / sqrt(1-(0.5*h2x*h2x*i)*(i-Tx))
  Kx2   = 1 - pnorm(Tx2)
  RRx2   = round(Kx2 / Kx,0)
  x <- as.data.frame(c(disorder,(h2x*100), round(Kx,2), round(Kx1,2), round(Kx2,2),round(RRx1,2), round
  return(x)
}
#Run function for number of included disorders

for(i in 1:length(disorder)){
  liability[i] <- liability_function(h2x[i], Kx[i], a, disorder[i])
}
#convert list to dataframe
Threshold_model_df <- data.frame(matrix(unlist(liability), nrow=length(liability), byrow=T), stringsAsFa
colnames(Threshold_model_df) <- c("Disorder","heritability","K", "K1", "K(2par)", "RR", "RR(2par)")

library(knitr)
kable(Threshold_model_df, caption ="Risk ratio in relatives")
```

Table 2: Risk ratio in relatives

Disorder	heritability	K	K1	K(2par)	RR	RR(2par)
Schizophrenia	70	0.01	0.07	0.3	7	30
Bipolar Disorder	65	0.01	0.06	0.25	6.2	25
Major Depressive Disorder	35	0.15	0.22	0.31	1.5	2
ADHD	75	0.05	0.18	0.46	3.5	9
Anorexia Nervosa	50	0.01	0.04	0.15	4.4	15
Autism Spectrum Disorder	80	0.01	0.09	0.41	8.7	41

h^2 estimates taken from family studies except ADHD (h^2_{twin})

The R code to calculate T, K, and RR is provided below

```
h2x=0.7
Kx=0.01
a=0.5
disorder="Schizophrenia"

Liability_threshold <- function(h2x, Kx, a, disorder){
  cat("Input Parameters:\n")
  cat("----------------------------\n")
  cat("Disorder:", disorder, "\n")
  cat("h2x   : \t",h2x,"\t heritability \n")
  cat("Kx    : \t",Kx,"\t lifetime risk of disease \n")
  cat("aR    : \t",a, "\t coefficient of relationship \n")
  cat("----------------------------\n")
  Tx    = -qnorm(Kx, 0,1)
  z    = dnorm(Tx)
  i    = z/Kx
  Tx1   = (Tx - a * i * h2x) / (sqrt(1 - a * a * h2x * h2x * i * (i-Tx)))
  Kx1   = 1 - pnorm(Tx1)
  RR   = Kx1 / Kx

  Tx2   = (Tx -i*h2x) / sqrt(1-(0.5*h2x*h2x*i)*(i-Tx))
  Kx2   = 1 - pnorm(Tx2)
  RR2   = Kx2 / Kx
  # return(matrix(data = c(Tx,Kx,NA, Tx1, Kx1,RR, Tx2, Kx2,RR2), nrow = 3, ncol = 3 ))
  cat("Output Parameters:\n")
  cat("----------------------------\n")
  cat("K(1)    : ",round(Kx1,2),"\t lifetime risk in 1st degree relatives of affected person \n")
  cat("K(2par) : ",round(Kx2,2), "\t lifetime risk in children with 2 affected parents \n")
  cat("RR      : ", round(RR,1), "\t relative  risk in 1st degree relatives of affected person \n")
  cat("RR(2par): ", round(RR2,0), "\t relative risk in children with 2 affected parents\n")
  cat("T       : ",round(Tx,2),"\t normal distribution threshold corresponding to proportion K \n")
  cat("T(1)    : ",round(Tx1,2),"\t normal distribution threshold corresponding to proportion K(1) \n")
  cat("T(2par) : ",round(Tx2,2), "\t normal distribution threshold corresponding to proportion K(2par) \

  cat("----------------------------\n")
}

Liability_threshold(h2x,Kx,a, disorder)
```

```
## Input Parameters:
## ----------------------------
## Disorder: Schizophrenia
## h2x   :   0.7     heritability
## Kx    :   0.01    lifetime risk of disease
## aR    :   0.5     coefficient of relationship
## ----------------------------
## Output Parameters:
## ----------------------------
## K(1)    :  0.07   lifetime risk in 1st degree relatives of affected person
## K(2par) :  0.3    lifetime risk in children with 2 affected parents
## RR      :  7      relative  risk in 1st degree relatives of affected person
## RR(2par):  30     relative risk in children with 2 affected parents
## T       :  2.33   normal distribution threshold corresponding to proportion K
## T(1)    :  1.48   normal distribution threshold corresponding to proportion K(1)
## T(2par) :  0.52   normal distribution threshold corresponding to proportion K(2par)
## ----------------------------
```

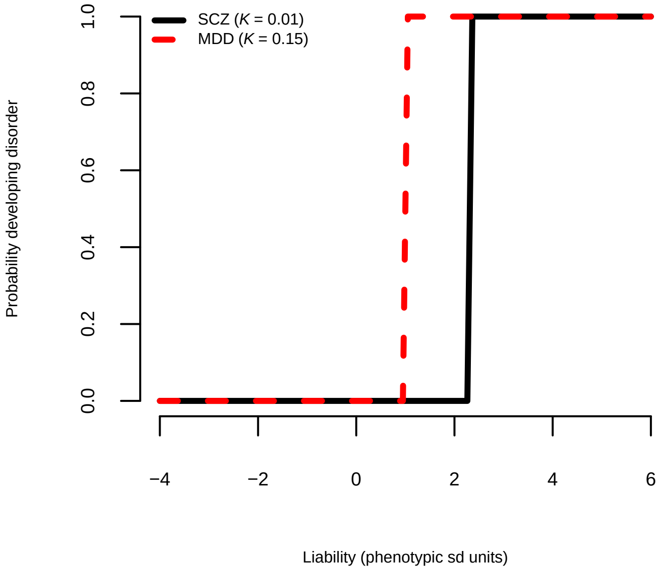
Different views of the liability threshold distribution

The liability model assumes that genetic and non-genetic factors contribute additively to liability to disease. This implies a non-linear relationship between phenotypic and genetic liability with probability of disease.

If we assume that individuals either have disease (probability of disease =1) or do not have disease (probability of disease =0), then when considering the relationship between phenotypic liability and disease, a vertical line will correspond to the threshold T corresponding to the lifetime risk of the disease, K, that bisects risk of disease (as visualized above). For schizophrenia (red), and major depressive disorder (black) with $K = 0.01$ and $K = 0.15$ the relationship between phenotypic liability and probability of developing the disorder looks like:

```
r2x <- 1
Kxp <- 0.01
r2y <- 1
Kyp <- 0.15
a <- 0.5
disorder1 <- "Schizophrenia"
disorder2 <- "Major Depressive Disorder"

xrange <- seq(-4,+6,len=100)
```

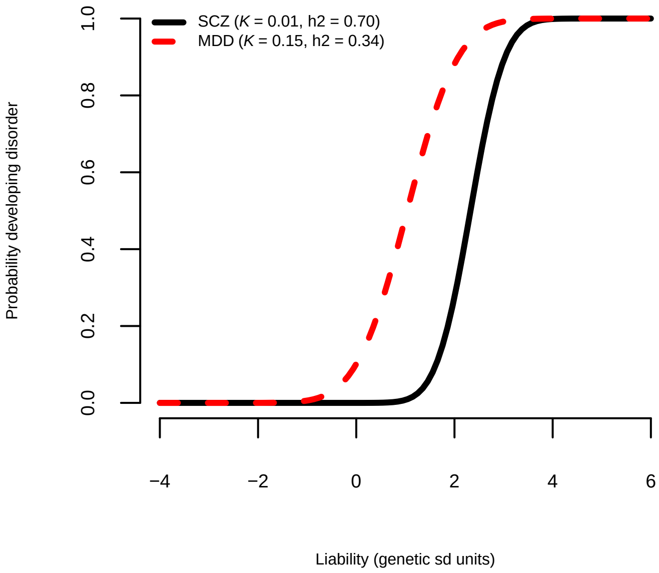
```
ProbDisease<- function(K,r2,xrange){
  sapply(xrange,function(x) pnorm( (x-qnorm(1-K))/sqrt(1-r2) ))
}

par(mar=c(5.1, 4.1, 4.1, 2.1))


l3<- c(as.expression((bquote(SCZ~'('*italic(K)~'= 0.01'*')'))))
l4<- c(as.expression((bquote(MDD~'('*italic(K)~'= 0.15'*')'))))
xrange <- seq(-4,+6,len=100)

SCZ<- ProbDisease(K=Kx,r2=r2x,xrange)
MDD <- ProbDisease(K=Ky,r2=r2y,xrange)

matplot(xrange,cbind(SCZ,MDD), mgp=c(3,1,0),
          frame.plot = FALSE, type="l",lty=1:2,col=1:2,lwd=3,
          xlab = "Liability (phenotypic sd units)",cex.axis=0.6, cex.lab =0.5,
          ylab = "Probability developing disorder")
          legend("topleft", legend = c(l3,l4) ,
          col =1:2, lty = 1:2, lwd=3, cex = 0.5,bty = "n")
```



Now, if we visualize the relationship between genetic liability and risk/probability of disease you will observe a very non-linear relationship with the gradient of relationship becomes steeper when heritability becomes higher. The position along the x-axis when the rise in probability starts is higher for diseases that are less common. Slatkin (2008) in his paper *Exchangeable models of complex inherited disease* nicely demonstrates that any models of polygenic disease have to lead to this non-linear relationship between genetic liability and probability of disease, and that the liability threshold model is the most convenient and mathematically tractable because it use the nice proprties of the normal distribution theory and is based on only two parameters, h^2 and K.

```
r2x <- 0.7
Kxp <- 0.01
r2y <- 0.34
Kyp <- 0.15
a <- 0.5
disorder1 <- "Schizophrenia"
```

```
disorder2 <- "Major Depressive Disorder"

xrange <- seq(-4,+6,len=100)
ProbDisease<- function(K,r2,xrange){
  sapply(xrange,function(x) pnorm( (x-qnorm(1-K))/sqrt(1-r2) ))
}

par(mar=c(5.1, 4.1, 4.1, 2.1))



l3<- c(as.expression((bquote(SCZ~'('*italic(K)~'= 0.01, h2 = 0.70'*')'))))
l4<- c(as.expression((bquote(MDD~'('*italic(K)~'= 0.15, h2 = 0.34'*')'))))
xrange <- seq(-4,+6,len=100)

SCZ<- ProbDisease(K=Kx,r2=r2x,xrange)
MDD <- ProbDisease(K=Ky,r2=r2y,xrange)

matplot(xrange,cbind(SCZ,MDD), mgp=c(3,1,0),
          frame.plot = FALSE, type="l",lty=1:2,col=1:2,lwd=3,
          xlab = "Liability (genetic sd units)",cex.axis=0.6, cex.lab =0.5,
          ylab = "Probability developing disorder")
          legend("topleft", legend = c(l3,l4) ,
          col =1:2, lty = 1:2, lwd=3, cex = 0.5,bty = "n")
```



SNP-based heritability

Heritability is the ratio of additive genetic variance V_A divided by the sum of V_A and the residual variance V_E (and where phenotypic variance, by definition is $V_P = V_A + V_E$. NB. Before we used σ^2 for variance, because we were thinking about parameters, we use V for variance because we are now considering estimates from data):

$$h^2 = \frac{V_A}{V_A + V_E}$$

Estimates of heritability are derived from knowledge of sharing between relatives, therefore, the genetic factors unique to an individual but which impact of disease status for that individual (i.e., *de novo* mutations) would be partitioned into the residual variance. Estimates of h^2 made from family record represent contributions from both rare and common genetic variants. In contrast h^2_{SNP} represents only the proportion of variance associated with measured common DNA variants (typically of allele frequency >1%):

$$h^2{}_{SNP} = \frac{V_{A_{SNP}}}{V_A + V_E}$$

Hence h^2_{SNP} is, by definition, lower than heritability gained from family and twin designs. SNP-based heritability is estimated from genome-wide association study (GWAS) data, either directly from individual level genotype data or from GWAS summary statistics

GREML

To estimate the proportion of phenotypic variance captured by genotyped SNPs from individual-level data, Genome-based Restricted Maximum-likelihood analysis (GREML) is frequently used applied to a linear mixed model:

y = covariates + $g + e$, with V(y)= $\mathbf{A}g\sigma^2_g + \mathbf{I}\sigma^2_e$

Where y is the phenotype, covariates could include age, sex and ancestry principal components (PCs) *etc.*, g is the aggregate effect of all genome-wide SNPs for an individual, and e is the residual effects for an individual, which includes genetic factors not captured by common SNPs. $V\ (y)$ is the variance of y, σ^2_g is the variance of the g values, with the matrix $\mathbf{A}g$ being the genetic relationship matrix (GRM) (which can be estimated from SNP data) since g values are correlated between people, and $\mathbf{I}$ is the identity matrix, since residual effects are assumed to be uncorrelated between individuals. Assuming that indivdiduals are unrelated in the classical sense (< ~3rd degree relatives) then h^2_{SNP} is then estimated as $h^2_{SNP} = \frac{\sigma^2_g}{\sigma^2_g + \sigma^2_e}$.

GRM matrix

A key component of the GREML approach is the GRM (**A**g) matrix:

Humans have diploid genomes, so at each (biallelic) SNP individuals can be homozygous (AA) for the reference allele, heterozygous (AC) or homozygous (CC) for the non-reference (or alternative) allele, and genotypes for each SNP for an individual can be coded as 0,1,2 with respect to the count of alternative alleles. Under a random mating assumption, the number of alternative alleles for a SNP and individual follows a binomial distribution with 2 draws (since we are diploid) and probability equal to the minor allele frequency of that SNP.

If we assume that the minor allele frequency (MAF) of a given number of SNPs (M) come from a uniform distribution between 0 and 0.5. We then can generate genotypes for one person using:

```
set.seed(666)
m <- 1000                                 # number of SNPs
maf <- runif(m, 0, .5)                    # random MAF for each SNP
x012 <- rbinom(m, 2, maf)
n <- 500                                    # number of individuals
X012 <- t(replicate(n, rbinom(m, 2, maf)))  # n x m genotype matrix (500 individuals X 1000 SNPs)
```

Monomorphic SNPs (SNPs of which frequencies do not vary between individuals in the sample) can cause problems in GRM analyses and should be removed. Therefore, to end up with only polymorphics SNPS, more SNPs are generated below and subsequently monomorphic SNPs will be removed.

```
X012 = t(replicate(n, rbinom(2*m, 2, c(maf, maf))))
polymorphic = apply(X012, 2, var) > 0
X012 = X012[,polymorphic][,1:m]
maf = c(maf, maf)[polymorphic][1:m]
```

```
X012[1:5, 1:10]
```

```
##      [,1] [,2] [,3] [,4] [,5] [,6] [,7] [,8] [,9] [,10]
## [1,]    0    0    1    0    0    1    1    0    0     0
## [2,]    0    0    2    0    0    1    0    0    0     1
## [3,]    0    0    0    0    1    0    2    0    0     1
## [4,]    1    0    2    0    0    1    1    0    0     0
## [5,]    1    1    1    0    1    0    1    1    0     1
```

A GRM is constructed from scaled genotype matrix in which the genotypes of each SNP in our sample have mean 0 and variance 1, therefore the genotype matrix created above has to be scaled

```
X = scale(X012, scale=TRUE)
```

With the scaled genotype matrix, a GRM can be calculated as follows, where M is the number of SNPs:

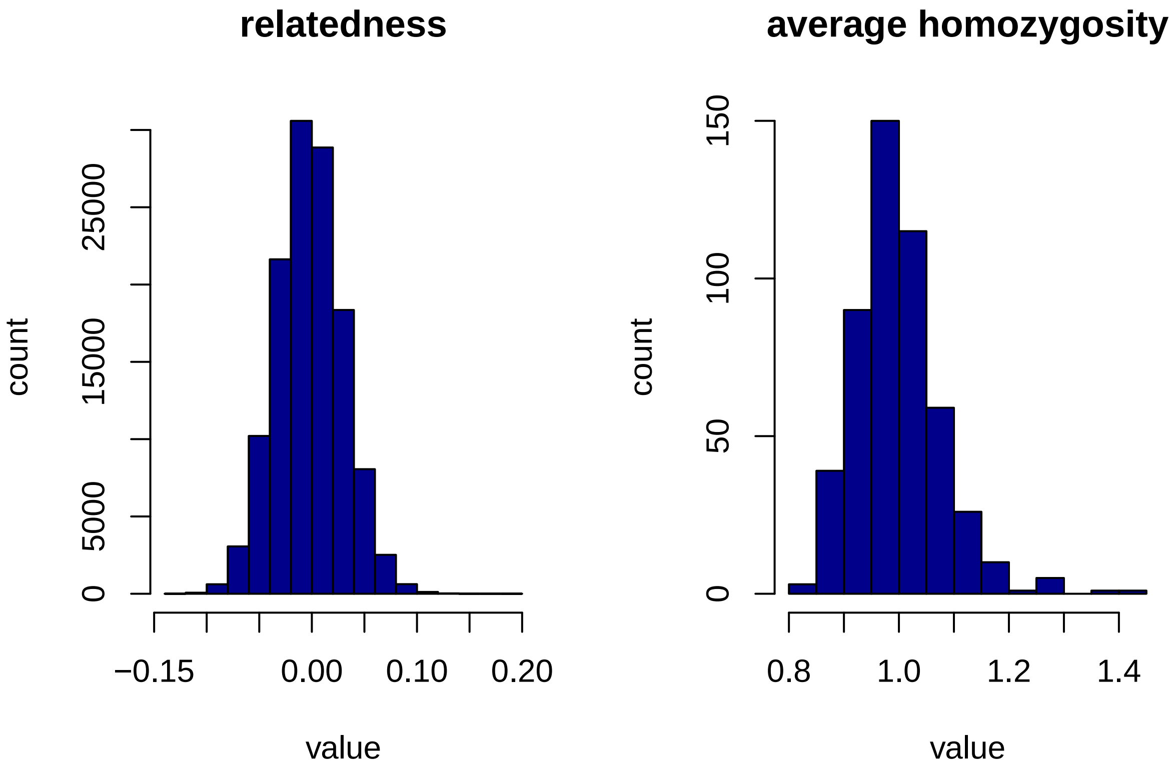
$$A = \frac{XX^t}{M}$$

```
grm = (X %*% t(X))/m
round(grm[1:5,1:10],3)
```

```
##        [,1]   [,2]   [,3]   [,4]   [,5]   [,6]   [,7]   [,8]   [,9]  [,10]
## [1,]  0.983  0.038  0.036  0.030 -0.008  0.017 -0.043 -0.004  0.019  0.004
## [2,]  0.038  0.995 -0.017 -0.016 -0.051  0.007  0.056  0.000  0.021 -0.011
## [3,]  0.036 -0.017  0.915  0.019  0.005 -0.001 -0.011 -0.021  0.038  0.038
## [4,]  0.030 -0.016  0.019  0.983 -0.011  0.000 -0.056  0.000  0.008  0.004
## [5,] -0.008 -0.051  0.005 -0.011  1.114 -0.037 -0.029 -0.033 -0.025 -0.019
```

Here, the offdiagonal elements represent the relative relatedness between two indivuals. Positive values indicate closer genetic relationship than on average, and negative values indicate a less close genetic relationship than average. The diagonal elements represents the relatedness of an individual with itself, which is 1+ the average homozygosity or the level of inbreeding. If we visualize this:

```
layout(matrix(c(1,2),1,2))
hist(grm[upper.tri(grm)],  ylab = "count", xlab = "value", col ="darkblue", main = "relatedness")
hist(diag(grm), ylab = "count", xlab = "value", col = "darkblue", main = "average homozygosity")
```



We see that the relatedness (GRM off-diagonals) are centered around 0 (left panel) and the relatedness with itself (GRM diagonals) is centered around 1.

LD Score regression

Linkage disequilibrium (LD) score regression (LDSC) was the first of now many methods to estimate SNP-based heritability from GWAS summary statistics. LDSC assumes a polygenic model of genetic effects on a trait. Under this assumption, and recognising that there is local correlation between DNA variants, along chromosomes (a reflection of past population growth, selection, drift), a SNP that is correlated to many other DNA variants (has a high linkage LD score, and) is expected to have a higher association test statistic, on average, compared to SNPs with low LD scores.

Theory shows that the expected ($E[\,]$) relationship between the chi-square test statistics (χ^2)) of SNPs and the LD score of SNPs is a function of the SNP-based heritability (h^2_{SNP}), sample size (N) and the number of SNPs (M)

$$E[\chi^2] = \frac{Nh^2_{SNP}}{M} l_j + intercept$$

Using GWAS summary statistics, h^2_{SNP} can be estimated from the regression coefficient (e.g. $\frac{Nh^2_{SNP}}{M}$) when the observed GWAS χ^2 test statistics are regressed on the LD scores (l_j).

Linkage disequilibrium (LD) Matrix

SNPs are often correlated with one another. Especially when they are physically close, since it is unlikely that recombination breaks up any correlation between them. This correlation between a pair of SNPs (two columns in our genotype matrix created earlier) is called linkage disequilibrium (LD) and can be estimated by calculating the correlation coefficient between the SNP genotypes.

The LD matrix contains the correlations of all SNP pairs in the genotype matrix and has therefore dimensions $M \times M$.

$$cor(x,y) = \frac{cov(x,y)}{\sqrt{var(x)var(y)}}$$

Since correlation implicitly scales by the variance it actually does not matter whether the scaled (X) or unscaled (X_012) genotype matrix is used. Further, when X is already scaled, the covariance matrix is the same as the correlation matrix, the LD matrix can be calculated as:

$$LD = \frac{X^t X}{N}$$

Therefore, if we calculate the LD matrices from our scaled or unscaled genotype matrices, we end up with the same results:

```
ld1 = cor(X012)
ld2 = cor(X)
mean(ld1)
```

```
## [1] 0.001016983
```

```
mean(ld2)
```

```
## [1] 0.001016983
```

These values are close to zero, because we simulated the SNPs to be independent

LD Scores

LD scores. l_j are defined as the sum of squared correlations of a SNP j with al other SNPs within a predefined region (usually, 1-Cm):

$$l_j = \sum_{k=1}^{M} cor^2(X_j, X_k)$$

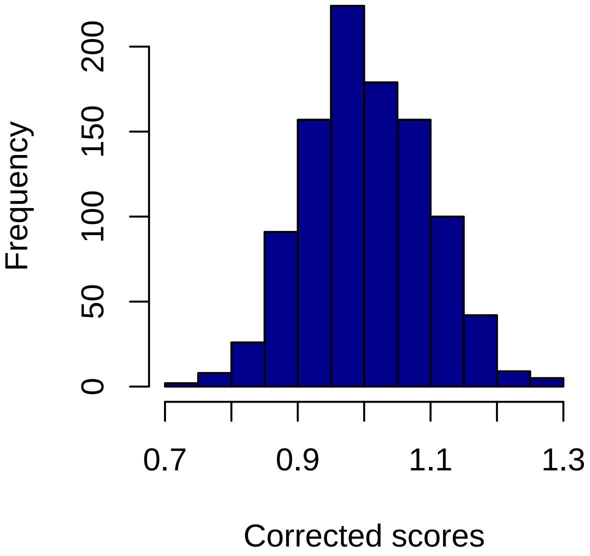
As X is standardized, we can calculate the sum of the squared sample correlations:

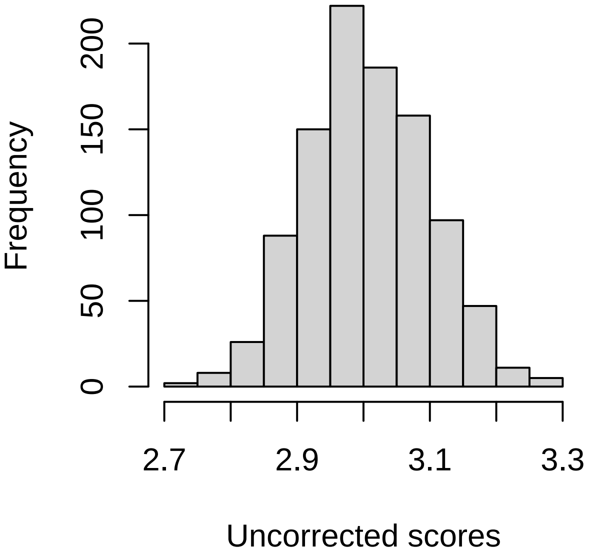
$$\tilde{l_j} = \frac{1}{N^2} X_j^t X X^t X_j$$

To correct for bias (just as in the LDSC formula) we can:

$$l_j = \frac{\tilde{l_j} N - M}{N + 1}$$

```
ldscores_sample = colSums(ld1^2)
ldscores = (ldscores_sample*n - m) / (n + 1)

layout(matrix(c(1,2),1,2))
hist(ldscores, col = "darkblue", xlab = "Corrected scores", main = "")
hist(ldscores_sample, col = "lightgrey", xlab = "Uncorrected scores", main = "")
```





Scale Transformation

Transformation of heritability estimates from observed to the liability scale

SNP-based heritability estimates can be calculated from GWAS summary statistics. The estimates are made on the case/control scale (h^2_{occ}). To be interpretable they are converted to the liability scale (h^2_l), which depends on the lifetime risk of disease (K) and the proportion of the sample that are controls (P). The standard transformation equation is

$$h^2_l = \frac{h^2_{occ} K^2 (1-K)^2}{P(1-P)z^2}$$

where z is the height of the normal curve at the threshold derived from K (see section 2.2). See Lee et al., 2010, Zhou et al., 2013, Golan et al., 2015 for derivations.

```
h2occtoh2l<-function(h2occ, K, ncase, ncont) {
P = ncase/(ncase+ncont)
T0<--qnorm(K,0,1) # threshold
z<-dnorm(T0) #height of normal curve at threshold
h2l=h2occ*K*(1-K)*K*(1-K)/(P*(1-P)*z*z) #heritability on liability scale
cat("SNP-based heritability transformation:\n")
  cat("----------------------------\n")
  cat("h2occ    :  ",round(h2occ[1],2),"\t SNP-based heritability as estimated from a linear model\n")
  cat("h2occ_se :  ",round(h2occ[2],2),"\t SNP-based heritability standard error\n")
  cat("K        :  ",round(K,2), "\t assumed lifetime risk of disease \n")
  cat("ncase    :  ", round(ncase,1), "\t number of cases in GWAS used to estimate SNP-based heritabili
  cat("ncontrol :  ", round(ncont,0), "\t number of controls in GWAS used to estimate SNP-based heritab
  cat("P        :  ",round(P,2),"\t proportion of the GWAS sample that are cases \n")
  cat("h2liab   :  ",round(h2l[1],2),"\t SNP-based heritability on the liability scale \n")
  cat("h2liab_se:  ",round(h2l[2],2),"\t SNP-based heritability standard error\n")
  cat("----------------------------\n")



}
h2occ=c(0.45,0.06)
```

```
ncase=34241
ncont=45064
K=0.01
h2occtoh2l(h2occ, K, ncase, ncont)
```

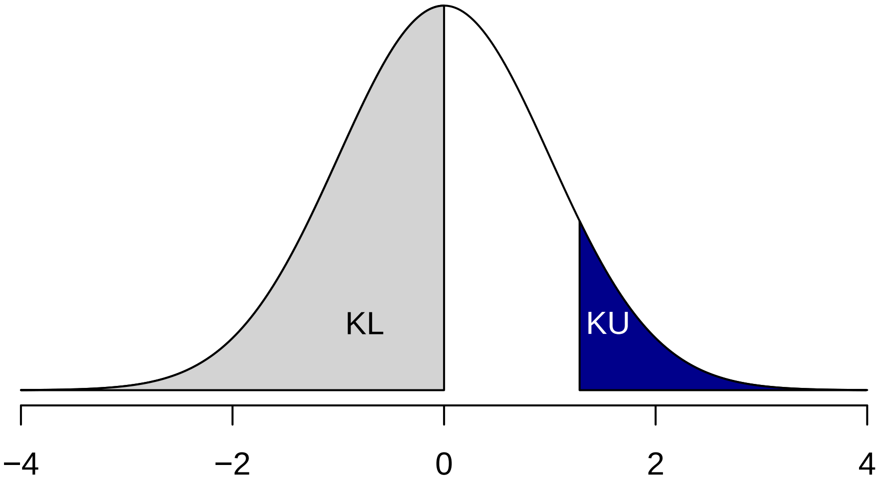
```
## SNP-based heritability transformation:
## ----------------------------
## h2occ    :   0.45     SNP-based heritability as estimated from a linear model
## h2occ_se :   0.06     SNP-based heritability standard error
## K        :   0.01     assumed lifetime risk of disease
## ncase    :   34241    number of cases in GWAS used to estimate SNP-based heritability
## ncontrol :   45064    number of controls in GWAS used to estimate SNP-based heritability
## P        :   0.43     proportion of the GWAS sample that are cases
## h2liab   :   0.25     SNP-based heritability on the liability scale
## h2liab_se:   0.03     SNP-based heritability standard error
## ----------------------------
```

Screened and unscreened controls

It is noteworthy that the standard scale transformation for SNP-based heritability Lee et al., (2011)) assumes that controls are screened. When K is small (for example, for schizophrenia, bipolar disorder or autism) there is little impact on estimates if controls are unscreened. However, for more common disorders the impact can be substantial and updated transformations are needed to account for unscreened or for super-screened controls.

Consider the following Liability Threshold model for a disorder with life time risk ($K = 0.1$).

```
mx=4
z <- seq( from=-mx, to=+mx, by=.01)
dens <- dnorm(z)
KU=0.10
KL=0.5
plot( z, dens, type="l", lwd=1,ylab="",yaxt="n",xlab="",
      frame.plot=F,main="")
t=qnorm(1-KU)

x.shade <- seq(t,mx,0.01)
n=length(x.shade)
polygon(c(rev(x.shade),x.shade),c(rep(0,n),dnorm(x.shade,0,1)),col="darkblue")
text(1.55,0.07,"KU", col = "white")

t=qnorm(1-KL)
x.shade <- seq(-mx,-t,0.01)
n=length(x.shade)
polygon(c(rev(x.shade),x.shade),c(rep(0,n),dnorm(x.shade,0,1)),col="lightgrey")
text(-0.75,0.07,"KL")
```



Here:

K_U = proportion of the general population that are cases (Corresponding to K)
K_L = proportion of the general population selected at the lower-end as controls.

Using K_U and K_L in general, for example:
If K_L=1, then controls are **Unscreened**
If K_L=1-K_U then controls are **Screeened**
If $K_L = 0.5$, then controls are **Heavily screened**
If K_L=K_U, then controls are **Super-screened**

$$h_l^2 = h_{occ}^2 P(1-P)(\frac{Z_U}{K_U} + \frac{Z_L}{K_L})^2$$

See Gianola (1979), Golan et al.,2015 and Yap et al., 2018 for derivations. For a given value estimate of SNP-based heritability calculated from a linear model on the observed scale, we can calculate the SNP-based heritability according to the different designs from which it could be estimated.

```
h2occtoh2l_screen<-function(h2occ, K_U, ncase, ncont) {
P = ncase/(ncase+ncont)
z_U=dnorm( qnorm(K_U))
K_L=1
z_L=dnorm( qnorm(K_L))
trans= P*(1-P)*(z_U/K_U + z_L/K_L)^2
h2l_uns= h2occ / trans
K_L=1-K_U
z_L=dnorm( qnorm(K_L))
trans= P*(1-P)*(z_U/K_U + z_L/K_L)^2
h2l_scr= h2occ / trans
K_L=0.5
z_L=dnorm( qnorm(K_L))
trans= P*(1-P)*(z_U/K_U + z_L/K_L)^2
h2l_heavy= h2occ / trans
K_L=K_U
z_L=dnorm( qnorm(K_L))
trans= P*(1-P)*(z_U/K_U + z_L/K_L)^2
h2l_super= h2occ / trans
cat("SNP-based heritability transformation:\n")
  cat("----------------------------\n")
  cat("h2occ       :  ",round(h2occ,2),"\t SNP-based heritability as estimated from a linear model\n")
  cat("K           :  ",round(K,2), "\t assumed lifetime risk of disease, K=KU \n")
  cat("ncase       :  ", round(ncase,1), "\t number of cases in GWAS used to estimate SNP-based heritab
  cat("ncontrol    :  ", round(ncont,0), "\t number of controls in GWAS used to estimate SNP-based herit
  cat("P           :  ",round(P,2),"\t proportion of the GWAS sample that are cases \n")
  cat("h2 on liability scale when controls are: \n")
  cat("unscreend       :",round(h2l_uns,2),"\t \n")
  cat("screened        :",round(h2l_scr,2),"\t  standard transformation \n")
  cat("heavily screened:",round(h2l_heavy,2),"\t  KL=0.5 \n")
  cat("super-screened  :",round(h2l_super,2),"\t  KL=KU \n")
  cat("NB: In a real analysis only one transformation would apply depending on the ascertainment criter
  cat("----------------------------\n")
}
h2occ=0.45
ncase=34241
ncont=45064
K=0.01
h2occtoh2l_screen(h2occ, K, ncase, ncont)
```

```
## SNP-based heritability transformation:
## ----------------------------
## h2occ       :   0.45      SNP-based heritability as estimated from a linear model
## K           :   0.01      assumed lifetime risk of disease, K=KU
## ncase       :   34241     number of cases in GWAS used to estimate SNP-based heritability
## ncontrol    :   45064     number of controls in GWAS used to estimate SNP-based heritability
## P           :   0.43      proportion of the GWAS sample that are cases
## h2 on liability scale when controls are:
## unscreend       : 0.26
## screened        : 0.25     standard transformation
## heavily screened: 0.15     KL=0.5
## super-screened  : 0.06     KL=KU
## NB: In a real analysis only one transformation would apply depending on the ascertainment criteria fo
```

Nagelkerke's R^2

Careful consideration of the scale of estimates is required in SNP-based genetic studies of disease traits. For example, estimates of SNP-based heritability are derived from a linear model, and a transformation must be applied, in order for them to be interpretable on the liability scale, which is a scale interpretable across studies. This transformation can also be applied in the context of polygenic risk score prediction, where polygenic risk scores are the weighted count of the DNA variants, for which both the DNA variants to include in the score and their weight are derived from GWAS summary statistics.

The predictive ability of polygenic risk scores are evaluated in case-control cohorts that are independent of the samples contributing to the GWAS summary statistics. For quantitative traits, the predictive ability is evaluated by the R^2 statistic (the proportion of phenotypic variance explained by the risk score). For disease traits the Nagelkerke's R^2 derived from a logistic regression is commonly reported. However, for a given variance in liability explained by a predictor, the measured Nagelkerke's R^2 depends on both P and K.

The relationship between Nagelkerke's Rˆ2 calculation and R2 on the liability scale was described by: Lee et al., (2012)

Here, Nagelkerke's Rˆ2 is calculated as:

$$R^2_N = \frac{R^2_{C\&S}}{R^2_{max}}$$

where $R^2_{C\&S}$ is the Cox and Snell's R^2 on the observed scale for which the theoretical expectation can be estimated as:

$$R^2_{C\&S} = h^2_l \frac{z^2}{K(1-K)}$$

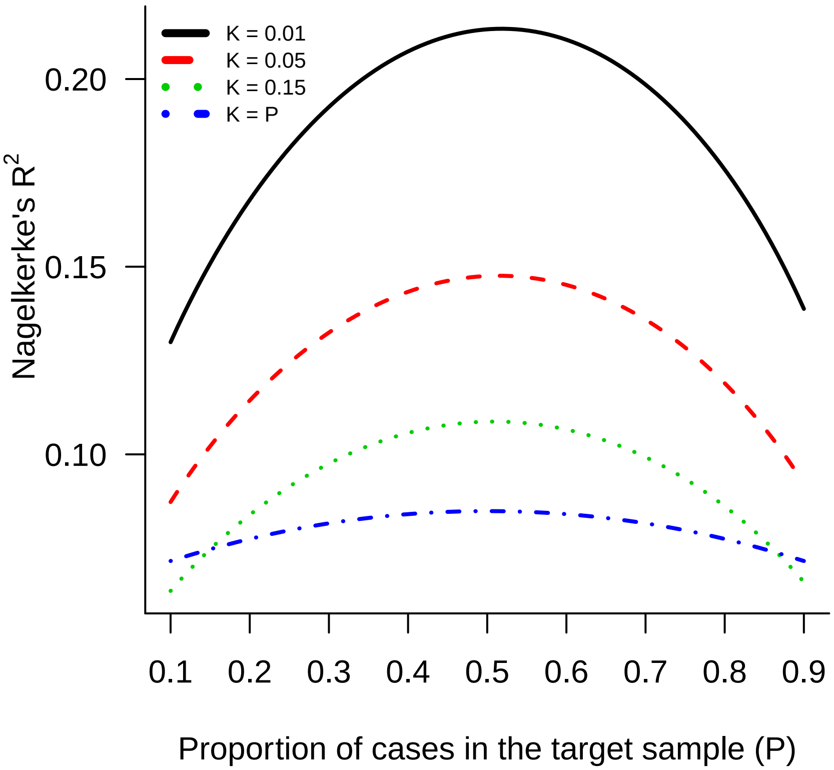
where h^2_l is the heritability on the liability scale, z is the height of the normal curve of threshold T, and $K(1-K)$ is the phenotypic variance (in the absence of covariates), from binominal theory

Subsequently, R^2_{max} is the maximum value $R^2_{C\&S}$ can ever attain and is estimated as:

$$R^2_{max} = 1 - K^{2K}(1-K)^{2(1-K)}$$

The code below shows the effect of the proportion of cases in the target sample (P) on the estimated Nagelkerke's R^2. For simplicity, we assume that the variance in liability explained by the predictor (sometimes called h^2_{PRS}) is constant at $R^2 = 0.1$ and we vary the proportion of cases between 0.1 and 0.9. Furthermore, we choose two scenarios, in which $K < P$, with $K = 0.01$, $K = 0.05$, and $= 0.15$), and one scenario where $K = P$.

```
xrange <- seq(0.1,0.9,len=100)
h2l = 0.1
P = 0

#Function to calculate Nagelkerke's R2
nagelkerke <- function(h2l,K){
  P=P+xrange
  #  K=P
  x= qnorm(1-K)
  z= dnorm(x)
  i=z/K
  C= K*(1-K)*K*(1-K)/(z^2*P*(1-P))
  theta= i*((P-K)/(1-K))*(i*((P-K)/(1-K))-x)
  CS=h2l/(C-h2l*theta*C)

  rmax=1-P^(2*P)*(1-P)^(2*(1-P))

  NR2=CS/rmax
}

#K=0.01
K = 0.01
K.0.01  <- sapply(0,function(x) nagelkerke(h2l,K))#[,1]

#K=0.1
K = 0.05
K.0.05  <- sapply(0,function(x) nagelkerke(h2l,K))
#K=0.1
K = 0.15
K.0.15  <- sapply(0,function(x) nagelkerke(h2l,K))

#K=P
nagelkerkeKP <- function(h2l,K){
  P=P+xrange
  K=P
  x= qnorm(1-K)
  z= dnorm(x)
  i=z/K
  C= K*(1-K)*K*(1-K)/(z^2*P*(1-P))
  theta= i*((P-K)/(1-K))*(i*((P-K)/(1-K))-x)
  CS=h2l/(C-h2l*theta*C)

  rmax=1-P^(2*P)*(1-P)^(2*(1-P))

  NR2=CS/rmax
}

KP  <- sapply(0,function(x) nagelkerkeKP(h2l,K))

#Visualize Nagelkerke's R2


l3<- c(as.expression((bquote("K = 0.01"))),
       as.expression((bquote("K = 0.05"))),
       as.expression((bquote("K = 0.15"))),
       as.expression((bquote("K = P"))))

par(mar=c(5.1, 5.1, 4.1, 2.1))
matplot(xrange,cbind(K.0.01,K.0.05,K.0.15,KP),
        mgp=c(3,1,0),
        frame.plot = FALSE, type="l",lty=1:10,col=1:4,lwd=2,
        xlab = "Proportion of cases in the target sample (P)", cex.lab =1,
        ylab = "", xaxt="n", yaxt="n") #, ylim = c(0,0.3)
axis(1, at = seq(0,1,0.1), cex.axis=1)
axis(2, at = seq(0,0.3,by = 0.05) , cex.axis=1, las=2)
mtext(as.expression((bquote("Nagelkerke's"~R^2))),side = 2, line = 2.5,at = 0.15, cex =1)
legend("topleft", legend = l3 ,
       col =1:6, lty = 1:6, lwd=4, cex = 0.7,bty = "n")
```



As shown in the above figure, estimates of Nagelkerke's R^2, are highest when the proportion of cases is 0.5, and this effect is stronger in disorders with a low lifetime prevalence (e.g. $K = 0.01$) compared to disorders that are more frequent in the population (e.g. $K = 0.15$)

Instead of Nagelkerke's R, the transformation derived for SNP-based heritability estimates described in Lee et al. (2011) can be applied to the R^2 from a linear regression, allowing the predictive ability of polygenic risk scores to be evaluated on the liability scale and hence be comparable with both heritabilities and SNP-based heritabilities see Lee et al. (2012).

Cross-disorder risk and genetic correlation

Cross-disorder risk in relatives

Just as epidemiological studies can investigate the increased risk of disorder x in relatives of those with disorder x, they can also collect the data to estimate the increased risk of disorder y in relatives of those with disorder x.

Calculating the cross-disorder risk in relatives can be done in a rather similar way as described in section 2 above following Falconer, 1965 and Wray and Gottesman, (2012) . Here, the threshold $T_{Rx,y}$ bisects the normal distribution for the proportion $K_{Rx,y}$ which is the lifetime risk of disease y in relatives of probands with disease x, and which relates to the threshold for disease x ($T_x) defined as:

$$T_{Rx,y} = \frac{T_x - a_R r_g h_x h_y i_y}{\sqrt{1 - a_R^2 r_g^2 h_x^2 h_y^2 i_y (i_y - T_y)}},$$

where r_g is the genetic correlation between the traits, and $r_g h_x h_y$ is the co-heritability between the traits (sometimes denoted $h_{x,y}$), and where and i_x and i_y , are the mean phenotypic liabilities of the two diseases.

If the heritabilities of two diseases, their lifetime risks and the genetic correlation between them are known, then this equation can be used to estimate $K_{Rx,y}$ as:

$$K_{Rx,y} = \Phi^{-1}(T_{Rx,y})$$

where $\Phi^{-1}(x)$ is the inverse of the cummulative standard normal distribution function. And the risk ratio for relatives is defined as:

$$RR_{Rx,y} = \frac{K_{Rx,y}}{K_x}$$

In R, you can calculate the cross disorder life-time risk and risk ratio of disease x in relatives of probands with disease y using the code below

```r
get_CDRR = function(h2x, h2y, Kx, Ky, rg, a, disorder1, disorder2){
  cat("Input Parameters:\n")
  cat("----------------------------\n")
  cat("aR   :\t",a, "\t coefficient of relationship, e.g. 0.5 for parent/offspring \n")
  cat("rg   :\t",rg, "\t genetic correlation between disorders \n")
  cat("disorder-x: \t", disorder1, "\n")
  cat("h2x  :\t",h2x,"\t heritability of disorder-x \n")
  cat("Kx   :\t",Kx,"\t lifetime risk of disorder-x \n")
  cat("disorder-y: \t", disorder2, "\n")
  cat("h2y  : \t",h2y,"\t heritability of disorder-y \n")
  cat("Ky   : \t",Ky,"\t lifetime risk of disorder-y \n")
  cat("----------------------------\n")

  Tx = -qnorm(Kx, 0, 1)
  Ty = -qnorm(Ky, 0, 1)
  zx = dnorm(Tx)
  zy = dnorm(Ty)
  ix = zx / Kx
  iy = zy / Ky

  #risk of disorder x in relatives of those with disorder x
  Txx   = (Tx - a * ix * h2x) / (sqrt(1 - a * a * h2x * h2x * ix * (ix-Tx)))
  Kxx   = 1 - pnorm(Txx)
  RRxx   = Kxx / Kx

  #risk of disorder y in relatives of those with disorder y
  Tyy   = (Ty - a * iy * h2y) / (sqrt(1 - a * a * h2y * h2y * iy * (iy-Ty)))
  Kyy   = 1 - pnorm(Tyy)
  RRyy   = Kyy / Ky

  # calculate the genetic covariance:
  covg = rg * sqrt(h2x * h2y)

  # calculate risk ratio for relatives
  Txy  = (Tx - a * covg * iy) / sqrt(1 - a^2 * covg^2 * iy * (iy-Ty))
  Kxy  = 1 - pnorm(Txy)
  RRxy = Kxy/Kx # relative risk of disease x given that parent has disease y

  # calculate risk ratio for relatives other way around
  Tyx  = (Ty - a * covg * ix) / sqrt(1 - a^2 * covg^2 * ix * (ix-Tx))
  Kyx  = 1 - pnorm(Tyx)
  RRyx = Kyx/Ky # relative risk of disease x given that parent has disease y

cat("Output Parameters:\n")
cat("----------------------------\n")
cat("Disorder x, consistent with heritability",h2x," and lifetime risk",Kx,":\n")
cat("Kxx  :\t",round(Kxx,2), "\t lifetime risk of disorder x in relatives (aR) of those with disorder x
cat("RRxx :\t",round(RRxx,2), "\t lifetime risk of disorder  in relatives (aR) of those with disorder x
cat("Disorder y, consistent with heritability",h2y," and lifetime risk",Ky,":\n")
cat("Kyy  :\t",round(Kyy,2), "\t lifetime risk of disorder x in relatives (aR) of those with disorder y
cat("RRyy :\t",round(RRyy,2), "\t lifetime risk of disorder x in relatives (aR) of those with disorder y
cat("Cross-disorder, consistent with above and rg",rg,": \n")
cat("Kxy  :\t",round(Kxy,2), "\t lifetime risk of disorder y in relatives (aR) of those with disorder x
cat("RRxy :\t",round(RRxy,2), "\t lifetime risk of disorder y in relatives (aR) of those with disorder x
cat("Kyx  :\t",round(Kyx,2), "\t lifetime risk of disorder x in relatives (aR) of those with disorder y
cat("RRyx :\t",round(RRyx,2), "\t lifetime risk of disorder x in relatives (aR) of those with disorder y
cat("----------------------------\n")
}
h2x <- 0.7
h2y <- 0.35
Kx <- 0.01
Ky <- 0.15
rg <- 0.34
a <- 0.5
disorder1 <- "schizophrenia"
disorder2 <- "major depressive disorder"
get_CDRR(h2x, h2y, Kx, Ky, rg, a, disorder1, disorder2)
```

```
## Input Parameters:
## ----------------------------
## aR   :    0.5     coefficient of relationship, e.g. 0.5 for parent/offspring
## rg   :    0.34    genetic correlation between disorders
## disorder-x:   schizophrenia
## h2x  :    0.7     heritability of disorder-x
## Kx   :    0.01    lifetime risk of disorder-x
## disorder-y:   major depressive disorder
## h2y  :    0.35    heritability of disorder-y
## Ky   :    0.15    lifetime risk of disorder-y
## ----------------------------
## Output Parameters:
## ----------------------------
## Disorder x, consistent with heritability 0.7  and lifetime risk 0.01 :
## Kxx  :    0.07    lifetime risk of disorder x in relatives (aR) of those with disorder x
## RRxx :    6.97    lifetime risk of disorder  in relatives (aR) of those with disorder x
## Disorder y, consistent with heritability 0.35  and lifetime risk 0.15 :
## Kyy  :    0.22    lifetime risk of disorder x in relatives (aR) of those with disorder y
## RRyy :    1.46    lifetime risk of disorder x in relatives (aR) of those with disorder y
## Cross-disorder, consistent with above and rg 0.34 :
## Kxy  :    0.01    lifetime risk of disorder y in relatives (aR) of those with disorder x
## RRxy :    1.38    lifetime risk of disorder y in relatives (aR) of those with disorder x
## Kyx  :    0.21    lifetime risk of disorder x in relatives (aR) of those with disorder y
## RRyx :    1.38    lifetime risk of disorder x in relatives (aR) of those with disorder y
## ----------------------------
```

Genetic correlation

If from real data the lifetime risks of the two disorders are known, i.e., K_x and K_y then these can be used to calculate the thresholds T_x, T_y. If the risks in relatives are measured of disease x in relatives of disease x, $K_{Rx,x}$ and of disease y in relatives of disease y, $K_{Rx,x}$ these can be used to calculate the thresholds of $T_{Rx,x}$ and $T_{Ry,y}$, then these can be used to estimate the heritability of each disorder. Last, if the lifetime risk of disease x is measured in relatives of those with disease y $K_{Rx,y}$ then this can be used to calculate and $T_{Rx,y}$ then these can be used to calculate the genetic correlation between the two disorders, by making r_g the subject of the equation above

$$r_g = \frac{T_y - T_{Ry,x}\sqrt{1-(1-T_x/i_x)(T_y^2 - T_{Ry,x}^2)}}{a_R(i_x + (i_x - T_x)T_{Ry,x}^2)\sqrt{h_x^2 h_y^2}}$$

```
get_rg_fromRR = function(h2x, h2y, Kx, Ky, Kyx, a, disorder1, disorder2){
  cat("Input Parameters:\n")
  cat("----------------------------\n")
  cat("aR   :\t",a, "\t coefficient of relationship, e.g. 0.5 for parent/offspring \n")
  cat("Kyx   :\t",Kyx, "\t increased risk of disorder y in those with disorder x \n")
  cat("disorder-x: \t", disorder1, "\n")
  cat("h2x  :\t",h2x,"\t heritability of disorder-x \n")
  cat("Kx   :\t",Kx,"\t lifetime risk of disorder-x \n")
  cat("disorder-y: \t", disorder2, "\n")
  cat("h2y  : \t",h2y,"\t heritability of disorder-y \n")
  cat("Ky   : \t",Ky,"\t lifetime risk of disorder-y \n")
  cat("----------------------------\n")

  Tx = -qnorm(Kx, 0, 1)
  Ty = -qnorm(Ky, 0, 1)
  zx = dnorm(Tx)
  zy = dnorm(Ty)
  ix = zx / Kx
  iy = zy / Ky

  # calculate risk ratio for relatives
  RRyx = Kyx/Kx # relative risk of disease x given that parent has disease y
  if(RRyx< 1){cat ("Kyx must be greater than Kx: STOP")}
  Tyx = -qnorm(Kyx, 0, 1)

  rg_num=Ty-Tyx*sqrt(1-(1-Tx/ix)*(Ty*Ty-Tyx*Tyx))
  rg_den=a*(ix+(ix-Tx)*Tyx*Tyx)*sqrt(h2x*h2y)
  rg=rg_num/rg_den


cat("Output Parameters:\n")
cat("----------------------------\n")
cat("genetic correlation:\t",round(rg,2),"\n")
cat("----------------------------\n")
}
h2x <- 0.7
h2y <- 0.35
Kx <- 0.01
Ky <- 0.15
Kyx <- 0.207
```

```
a <- 0.5
disorder1 <- "schizophrenia"
disorder2 <- "major depressive disorder"
get_rg_fromRR(h2x, h2y, Kx, Ky, Kyx, a, disorder1, disorder2)
```

```
## Input Parameters:
## ----------------------------
## aR   :    0.5     coefficient of relationship, e.g. 0.5 for parent/offspring
## Kyx    :   0.207   increased risk of disorder y in those with disorder x
## disorder-x:   schizophrenia
## h2x  :    0.7     heritability of disorder-x
## Kx   :    0.01    lifetime risk of disorder-x
## disorder-y:   major depressive disorder
## h2y  :    0.35    heritability of disorder-y
## Ky   :    0.15    lifetime risk of disorder-y
## ----------------------------
## Output Parameters:
## ----------------------------
## genetic correlation:  0.34
## ----------------------------
```

SNP-based genetic correlation

Bivariate extensions of both the GREML and LDSC methods allow estimation of the a SNP-based genetic correlation from GWAS data sets that have been collected independently for the two traits.

bivariate GREML

In essence, bivariate GREML detects if cases of the two diseases are significantly more similar genetically than they are to controls (or significantly less similar in the case of negative correlation) Lee et al 2012 It got too complicated to call the traits x and y so here we call them traits 1 and 2!

Consider
$y_1 = X_1 b_1 + Z_1 g_1 + e_1$ for trait 1 and
$y_2 = X_2 b_2 + Z_2 g_2 + e_2$ for trait 2.

Here, y_1 and y_2 are two vectors of observations for trait 1 and 2, b_1 and b_2 are vectors of fixed effects, g_1 and g_2 are vectors of random polygenic effects for each individual in both trait 1 and trait 2, while e_1 and e_2 the residuals for trait 1 and trait 2 are. Additionally, X and Z are incidence matrices for the effects b and g. The variance covariance matrix (V) is subsequently defined as:

$$V = \begin{pmatrix} Z_1 A Z_1' \sigma^2_{g1} + I\sigma^2_{e1} & Z_1 A Z_2' \sigma^2_{g1g2} \\ Z_2 A Z_1' \sigma^2_{g1g2} & Z_2 A Z_2' \sigma^2_{g2} + I\sigma^2_{e2} \end{pmatrix}$$

Where A is the genomic relatedness matrix see above and I is an identity matrix, σ^2_g, σ^2_e and σ^2_{g1g1}, which are respectively the genetic variance, residual variance and covariance between $g1$ and $g2$.

Genetic correlation using LD score regression

LD Score regression Bulik-Sullivan et al. 2015 can estimate the genetic correlation between tow traits from GWAS sumary statistics. The method is based on the knowlegde that the GWAS effect size estimate for a given SNP incorporates the effect of all SNPs in linkage disequilibrium with that SNP click here for more information on LD scores. For complex traits, SNPs with high LD-scores have on average higher χ^2 statistics compared to SNPs with low LD-scores under a polygenic model. This observation holds if the χ^2 statistic representing a single trait is replaced with the product of the z scores from two traits (bivariate genetic correlation). Here the expected (E) value of $z_{1j}z_{2j}$ for SNP_j is:

$$E[z_{1j}z_{2j}] = \frac{\sqrt{N_1 N_2 h_{1,2}}}{M} l_j + \frac{\varrho N_s}{\sqrt{N_1 N_2}}$$

Here, N_i is the sample size for the included studies, $h_{1,2}$ is the genetic covariance (or coheritability, since the phenotypic varianced of liability are 1 by definition) and l_j is the LD score. N_s represents the sample overlap between both studies and ϱ is the phenotypic correlation between the overlapping samples.

In general, sample overlap generates a correlation between z_{1j} and z_{2j}, which inflates $z_{1j}z_{2j}$. However, the effect of this inflation is expecetd to be the same across all markers in is independent of the LD scores. Therefore, sample overlap only affects the intercept, which is represented by $\varrho N_s/\sqrt{N_1 N_2}$. The slope itself will not be biased by sample overlap.

Scale

Genetic correlations are scale free, and so no scaling trasnformation from the scale of estimation is needed.

Simulations

(Simulations phenotypes partly based on van Rheenen et al., 2019)

Simulation can be a useful tool. Here, we simulate phenotypes (disease liability and case/control status) in two correlated disorders SCZ (h^2_x) and MDD (h^2_y) in N parent-offspring pairs ($a_R = 0.5$) Phenotypic liability is simulated as $P = G + E$, where G is the genetic value and E the non-genetic (residual) value. G is drawn from the multivariate normal distribution $N(\mathbf{0}, V_g)$ where (V_g) is a symmetric variance-covariance matrix:

$$V_g = \begin{pmatrix} h^2_x & r_g\sqrt{h^2_x h^2_y} & \frac{1}{2}h^2_x & \frac{1}{2}r_g\sqrt{h^2_x h^2_y} \\ & h^2_y & \frac{1}{2}r_g\sqrt{h^2_x h^2_y} & \frac{1}{2}h^2_y \\ & & h^2_x & r_g\sqrt{h^2_x h^2_y} \\ & & & h^2_y \end{pmatrix}$$

Where h^2 is there heritability of the trait and r_g is the genetic correlation. Note, that shared non-genetic effects are not modelled, such that E are drawn from multivariate lnormal distribution $N(\mathbf{0}, V_e)$ where (V_e) is a symmetric variance-covariance matrix:

$$V_e = \begin{pmatrix} (1-h^2_x) & 0 & 0 & 0 \\ & (1-h^2_y) & 0 & 0 \\ & & (1-h^2_x) & 0 \\ & & & (1-h^2_y) \end{pmatrix}$$

```r
N=1000000
Kx=0.01
Ky=0.15
h2x=0.7
h2y=0.35
rg=0.34
a =0.5

h2_estimator = function(N, Kxp, Kyp, h2x, h2y, rg,a){
```

```
cat("Input Parameters:\n")
cat("----------------------------\n")
cat("h2x  : ", h2x, "\n")
cat("h2y  : ", h2y, "\n")
cat("rg   : ", rg, "\n")
cat("aR  : ", a, "\t coefficient of relationship \n")
cat("N    : ", N, "\n")
cat("Estimates:\n")
cat("----------------------------\n")

# simulate the phenotypes for x and y for parent (p) and child (c)
# define genetic parameters
gcovxy = rg*sqrt(h2x*h2y)

Vg  = matrix(c(h2x, gcovxy, a*h2x, a*gcovxy,
               gcovxy, h2y, a*gcovxy, a*h2y,
               a*h2x,  a*gcovxy, h2x, gcovxy,
               a*gcovxy, a*h2y, gcovxy, h2y), nrow=4)

# simulate genetic values
G  = MASS::mvrnorm(n=N, mu=c(0,0,0,0), Sigma=Vg)
Vg  = matrix(c(h2x, gcovxy, a*h2x, a*gcovxy,
               gcovxy, h2y, a*gcovxy, a*h2y,
               a*h2x,  a*gcovxy, h2x, gcovxy,
               a*gcovxy, a*h2y, gcovxy, h2y), nrow=4)

# simulate genetic values
G  = MASS::mvrnorm(n=N, mu=c(0,0,0,0), Sigma=Vg)

# simulate trait liability for x and y
Yxp = G[,1] + rnorm(N, mean=0, sd=sqrt(1-h2x))
Yyp = G[,2] + rnorm(N, mean=0, sd=sqrt(1-h2y))
Yxc = G[,3] + rnorm(N, mean=0, sd=sqrt(1-h2x))
Yyc = G[,4] + rnorm(N, mean=0, sd=sqrt(1-h2y))

# define parameter from normal distribution theory
Tx = -qnorm(Kx, 0, 1)
Ty = -qnorm(Ky, 0, 1)
zy = dnorm(Ty)
zx = dnorm(Tx)
iy = zy/Ky
ix = zx/Kx

# dichotomize to define binary trait
Yxp_cc = rep(0, N) ; Yxp_cc[Yxp >= Tx] = 1
Yyp_cc = rep(0, N) ; Yyp_cc[Yyp >= Ty] = 1
Yxc_cc = rep(0, N) ; Yxc_cc[Yxc >= Tx] = 1
Yyc_cc = rep(0, N) ; Yyc_cc[Yyc >= Ty] = 1

# estimate heritability using normal distribution theory:
Kx_est = sum(Yxp_cc, Yxc_cc) / (2*N)
Ky_est = sum(Yyp_cc, Yyc_cc) / (2*N)
```

```
  Tx_est = -qnorm(Kx_est, 0, 1)
  zx_est = dnorm(Tx_est)
  ix_est = zx_est / Kx_est

  Ty_est = -qnorm(Ky_est, 0, 1)
  zy_est = dnorm(Ty_est)
  iy_est = zy_est / Ky_est

  Trxx_est  = -qnorm(sum(Yxc_cc[Yxp_cc == 1]) / sum(Yxp_cc), 0, 1)
  Tryy_est  = -qnorm(sum(Yyc_cc[Yyp_cc == 1]) / sum(Yyp_cc), 0, 1)

  h2x_est = (Tx_est - Trxx_est * sqrt(1-(1 - Tx_est/ix_est)*(Tx_est^2 - Trxx_est^2)) ) / (a * (ix_est +

  h2y_est = (Ty_est - Tryy_est * sqrt(1-(1 - Ty_est/iy_est)*(Ty_est^2 - Tryy_est^2)) ) / (a * (iy_est +

  # estimate the genetic correlation using normal distribution theory:
  Tryx_est = -qnorm(sum(Yyc_cc[Yxp_cc == 1]) / sum(Yxp_cc), 0, 1)
  rg_est = (Ty_est - Tryx_est * sqrt(1 - (1 - Tx_est/ix_est)*(Ty_est^2 - Tryx_est^2)))  / (a*(ix_est +

cat("Output Parameters:\n")
cat("----------------------------\n")
cat("Kx_est : ", round(Kx_est,2), "\n")
cat("Ky_est : ", round(Ky_est,2), "\n")
cat("h2x_est: ",round(h2x_est,2),"\n")
cat("h2y_est: ",round(h2y_est,2),"\n")
cat("rg_est : ",round(rg_est,2),"\n")
cat("----------------------------\n")
}

h2_estimator(N, Kxp, Kyp, h2x, h2y, rg,a)
```

```
## Input Parameters:
## ----------------------------
## h2x  :  0.7
## h2y  :  0.35
## rg   :  0.34
## aR  :  0.5    coefficient of relationship
## N    :  1e+06
## Estimates:
## ----------------------------
## Output Parameters:
## ----------------------------
## Kx_est :  0.01
## Ky_est :  0.15
## h2x_est:  0.69
## h2y_est:  0.35
## rg_est :  0.38
## ----------------------------
```

Note, if you increase N (N was kept relatively low for computational reasons), the estimated values (h_x^2, h_y^2, and r_g) will

Code for figures

Figure 1

Code used To calculate K, and Kx, x

```
liability_function2 <- function(h2x, Kx, a, disorder){

  Tx    = -qnorm(Kx, 0,1)
  z    = dnorm(Tx)
  i    = z/Kx
  #One affected parent
  Tx1   = (Tx - a * i * h2x) / (sqrt(1 - a * a * h2x * h2x * i * (i-Tx)))
  Kx1   = 1 - pnorm(Tx1)
  RRx1   = Kx1 / Kx

  x <- as.data.frame(c(disorder,Tx,Tx1, Kx, Kx1, NA,RRx1 ))
  return(x)
}
```

Figure 1b: Heritability estimated derived from family data and GWAS
Data for this figure can be found in supplementary Table 1.

Code for Figure 2

Figure 2 can be reproduced with code coverd in section 2.3: Risk in relatives and heritability of liability

Code fo Figure 3

Figure 3 can be reproduced with code covered in section 2.4: Different views of the liability threshold distribution

Code for Figure 5

Genetic correlations (**Figure 5a**) were estimated using LDSC Bulik-Sullivan, 2015. Summary statistics used for these analyses are presented in **supplementary table 1**.

The following code was used to estimate CDRR for the major psychiatric disorders

```
get_CDRR = function(h2x, h2y, Kx, Ky, rg, a, disorder1, disorder2){
  cat("Input Parameters:\n")
  cat("----------------------------\n")
  cat("aR   :\t",a, "\t coefficient of relationship, e.g. 0.5 for parent/offspring \n")
  cat("rg   :\t",rg, "\t genetic correlation between disorders \n")
  cat("disorder-x: \t", disorder1, "\n")
  cat("h2x  :\t",h2x,"\t heritability of disorder-x \n")
  cat("Kx   :\t",Kx,"\t lifetime risk of disorder-x \n")
  cat("disorder-y: \t", disorder2, "\n")
  cat("h2y  : \t",h2y,"\t heritability of disorder-y \n")
  cat("Ky   : \t",Ky,"\t lifetime risk of disorder-y \n")
  cat("----------------------------\n")

  Tx = -qnorm(Kx, 0, 1)
  Ty = -qnorm(Ky, 0, 1)
  zx = dnorm(Tx)
  zy = dnorm(Ty)
  ix = zx / Kx
  iy = zy / Ky

  #risk of disorder x in relatives of those with disorder x
  Txx   = (Tx - a * ix * h2x) / (sqrt(1 - a * a * h2x * h2x * ix * (ix-Tx)))
  Kxx   = 1 - pnorm(Txx)
  RRxx   = Kxx / Kx

  #risk of disorder y in relatives of those with disorder y
  Tyy   = (Ty - a * iy * h2y) / (sqrt(1 - a * a * h2y * h2y * iy * (iy-Ty)))
  Kyy   = 1 - pnorm(Tyy)
  RRyy   = Kyy / Ky

  # calculate the genetic covariance:
  covg = rg * sqrt(h2x * h2y)

  # calculate risk ratio for relatives
  Txy  = (Tx - a * covg * iy) / sqrt(1 - a^2 * covg^2 * iy * (iy-Ty))
  Kxy  = 1 - pnorm(Txy)

  RRxy = Kxy/Kx # relative risk of disease x given that parent has disease y

  # calculate risk ratio for relatives other way around
  Tyx  = (Ty - a * covg * ix) / sqrt(1 - a^2 * covg^2 * ix * (ix-Tx))
  Kyx  = 1 - pnorm(Tyx)
  RRyx = Kyx/Ky # relative risk of disease x given that parent has disease y

  cat("Output Parameters:\n")
  cat("----------------------------\n")
  cat("Disorder x, consistent with heritability",h2x," and lifetime risk",Kx,":\n")
  cat("Kxx  :\t",round(Kxx,2), "\t lifetime risk of disorder x in relatives (aR) of those with disorder
  cat("RRxx :\t",round(RRxx,2), "\t lifetime risk of disorder  in relatives (aR) of those with disorder
  cat("Disorder y, consistent with heritability",h2y," and lifetime risk",Ky,":\n")
  cat("Kyy  :\t",round(Kyy,2), "\t lifetime risk of disorder y in relatives (aR) of those with disorder
  cat("RRyy :\t",round(RRyy,2), "\t lifetime risk of disorder x in relatives (aR) of those with disorder
  cat("Cross-disorder, consistent with above and rg",rg,": \n")
  cat("Kxy  :\t",round(Kxy,2), "\t lifetime risk of disorder y in relatives (aR) of those with disorder
  cat("RRxy :\t",round(RRxy,2), "\t lifetime risk of disorder y in relatives (aR) of those with disorder
  cat("Kyx  :\t",round(Kyx,2), "\t lifetime risk of disorder x in relatives (aR) of those with disorder
  cat("RRyx :\t",round(RRyx,2), "\t lifetime risk of disorder x in relatives (aR) of those with disorder
  cat("----------------------------\n")
}
```

Code for Supplementary Figure 1

To estimate standard errors of heritability estimates based on family data (1 parent or both parents measured), we followed Falconer, (1965). To estimate standard errors of heritability estimates based on GWAS data we followed Visscher et al., 2014.

The following code was used:

```
xrange <- seq(100,100000, len=1000)
#Page  166-167 Falconer- INtroduction in QT
```

```
#SE approximation and relation with N families
options(scipen=999)
SE_2parent <- function(N){
  sapply(xrange, function(N) sqrt(2/N))
}

SE_1parent <- function(N){
  sapply(xrange, function(N) 2/(sqrt(N)))
}

##### relation SE and N with genome-wide data
# Statistical Power to Detect Genetic (Co)Variance of Complex Traits Using SNP Data in Unrelated Sample
SE_SNP <- function(N){
  sapply(xrange, function(N) 316/N)
}
```

References

1) Falconer DS (1965): The inheritance of liability to certain diseases, estimated from the incidence among relatives. Ann Hum Genet 29: 51–76.

2) Reich T, James JW, Morris CA (1972): The use of multiple thresholds in determining the mode of transmission of semi-continuous traits. Ann Hum Genet 36: 163–184.

3) Wray NR, Gottesman II (2012): Using summary data from the Danish National Registers to estimate heritabilities for schizophrenia, bipolar disorder, and major depressive disorder. Front Genet 3: 1–12.

4) Slatkin (2008): Exchangeable Models of Complex Inherited Diseases. GENETICS, 179,4.

5) Lee SH, Wray NR, Goddard ME, Visscher PM (2011): Estimating missing heritability for disease from genome-wide association studies. Am J Hum Genet 88: 294–305.

6) Zhou, X., Carbonetto, P., & Stephens, M. (2013). Polygenic modeling with Bayesian sparse linear mixed models. PLoS genetics, 9(2).

7) Gianola D (1979): Heritability of polychotomous characters. Genetics 93: 1051–1055.

8) Golan, D., Lander, E. S., & Rosset, S. (2014). Measuring missing heritability: inferring the contribution of common variants. Proceedings of the National Academy of Sciences, 111(49), E5272-E5281.

9) Yap CX, Sirodenko J, Marioni RE, Yengo L, Wray NR, Visscher PM (2018): Misestimation of heritability and prediction accuracy of male-pattern baldness. Nat Commun 9: 9–11.

10) Lee SH, Yang J, Goddard ME, Visscher PM, Wray NR (2012): Estimation of pleiotropy between complex diseases using single-nucleotide polymorphism-derived genomic relationships and restricted maximum likelihood. Bioinformatics 28: 2540–2542.

11) Bulik-Sullivan B, Finucane HK, Anttila V, Gusev A, Day FR, Loh PR, et al. (2015): An atlas of genetic correlations across human diseases and traits. Nat Genet 47: 1236–1241.

12) van Rheenen W, Peyrot WJ, Schork AJ, Lee SH, Wray NR (2019): Genetic correlations of polygenic disease traits: from theory to practice. Nat Rev Genet 20: 567–581.

13) Visscher, P.M., Hemani, G., Vinkhuyzen, A.A., Chen, G.B., Lee, S.H., Wray, N.R., Goddard, M.E. and Yang, J., 2014. Statistical power to detect genetic (co) variance of complex traits using SNP data in unrelated samples. PLoS genetics, 10(4).