Heritability of diseases

Loic Yengo
Group Leader – Statistical Genomics Group
Institute for Molecular Bioscience
The University of Queensland

Aim of this lecture

In this lecture, we will show why these statements are consistent :

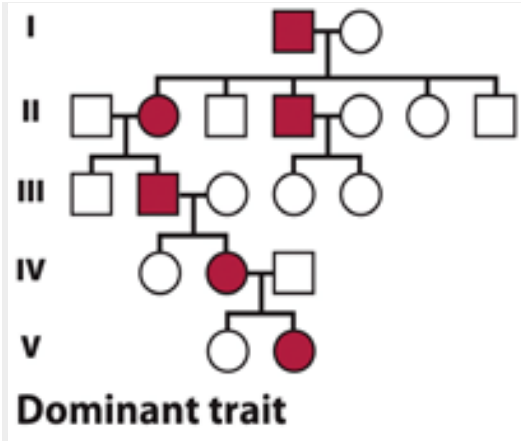
If a disease affects 1% of the population and has heritability 80% then

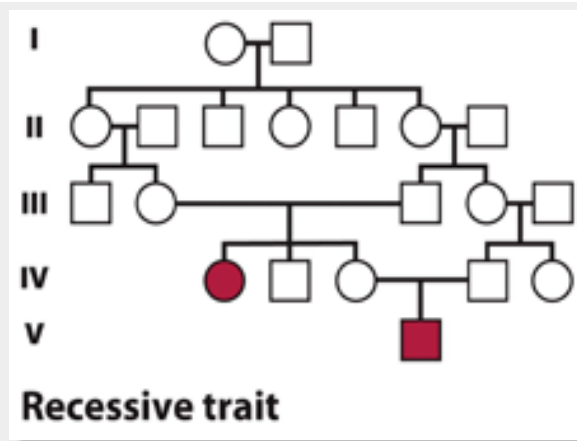
If an individual is affected ~8% of his/her siblings affected

If an MZ twin is affected ~50% of their co-twins are affected

How do we know that there is a genetic contribution to disease?

Disease traits: Some disease traits have a clear pattern of inheritance





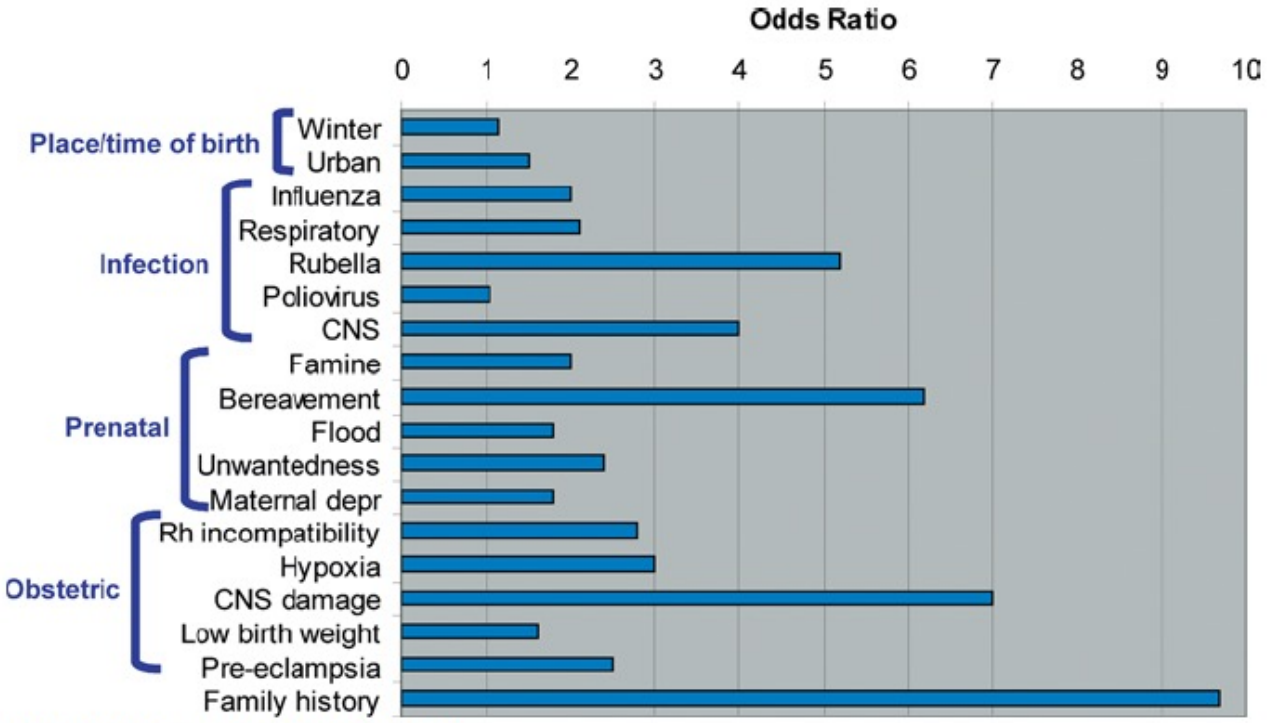
There are thousands of single gene disorders (OMIM database Online Mendelian Inheritance in Man) but each is very rare

Most common diseases do not have a clear pattern of inheritance yet there is increased risk associated with family history

Complex genetic diseases

- Unlike Mendelian disorders, there is no clear pattern of inheritance
- Tend to "run" in families
- Few large pedigrees of multiply affected individuals
- Most people have no known family history
- Common diseases > 0.5% are complex genetic diseases

Risk Factors for Schizophrenia

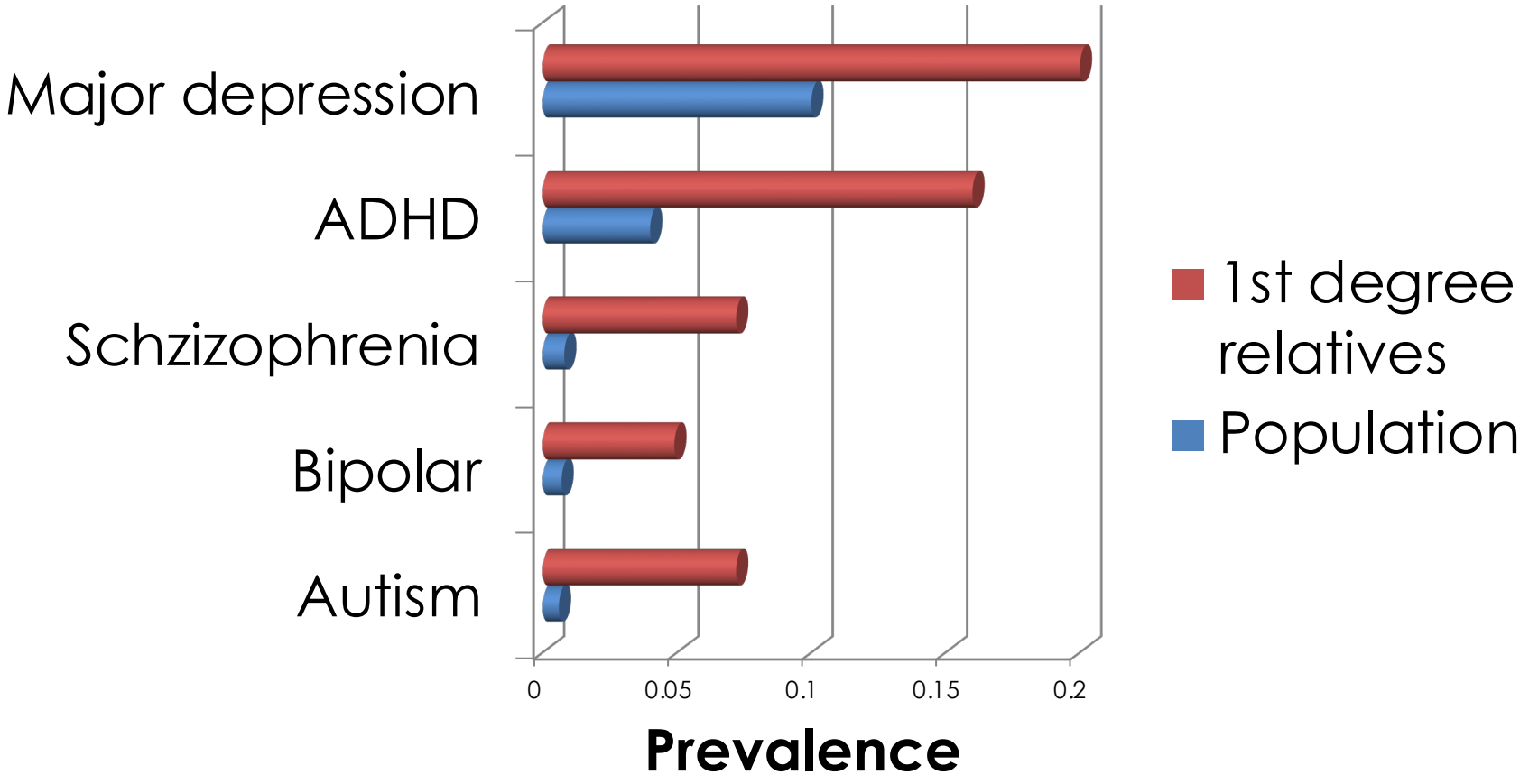


DOI: 10.1371/journal.pmed.0020212.g001

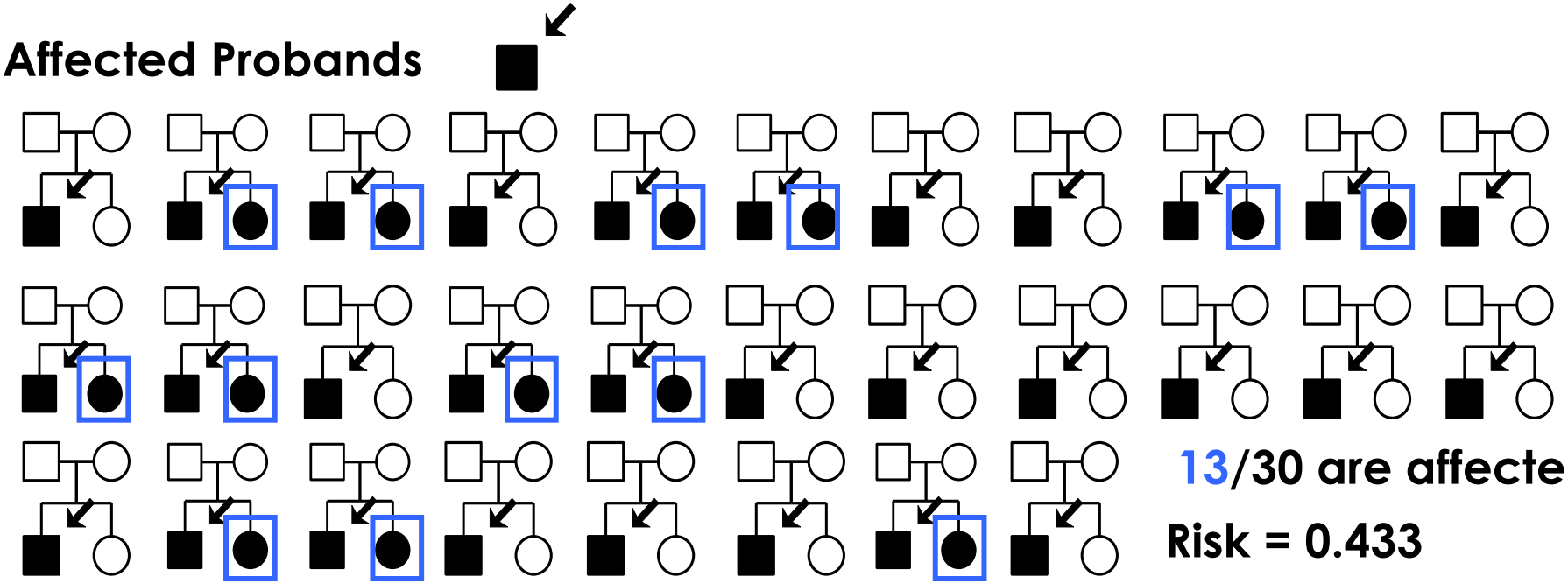
Figure 1. Comparison of a Selected Set of Relatively Well-Established Risk Factors for Schizophrenia, Focusing Mainly on Pre- and Antenatal Factors [6] (abbreviations: CNS, central nervous system; depr, depression; Rh, Rhesus)

Sullivan, PLoS Med 05

Evidence for a genetic contribution comes from risks to relatives



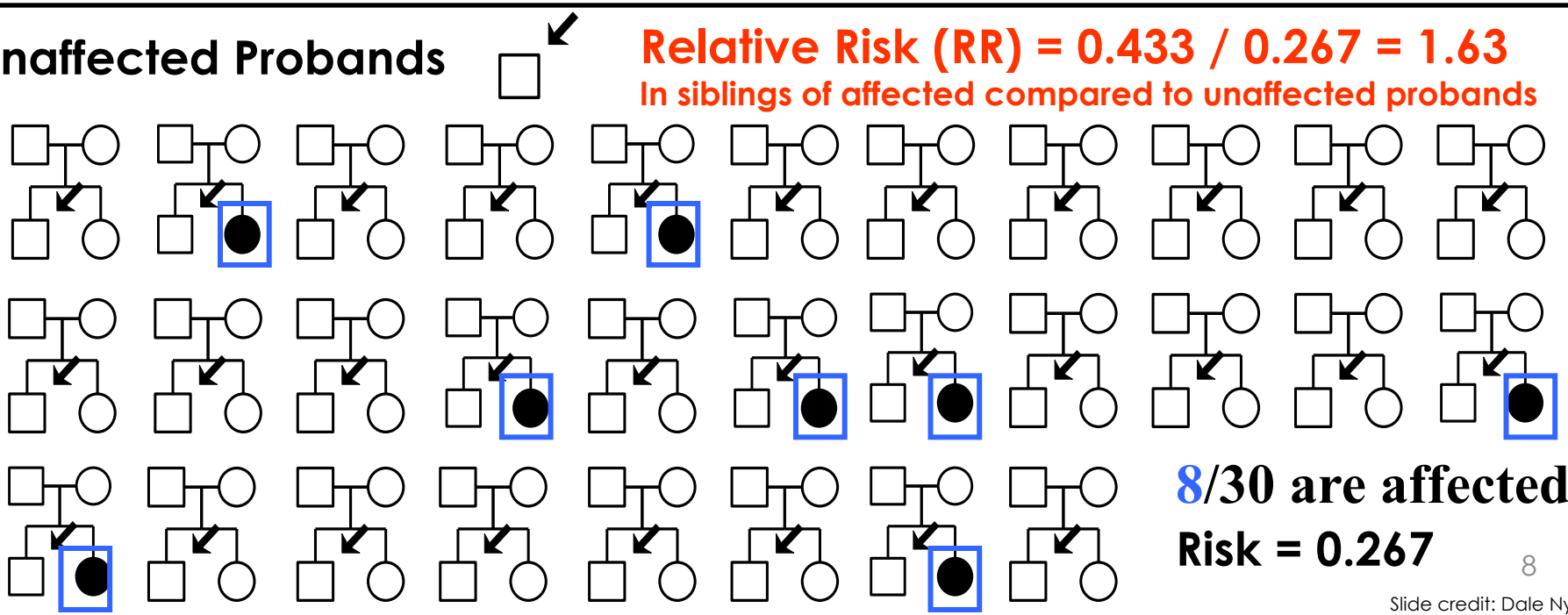
Affected Probands



13/30 are affected;

Risk = 0.433

Unaffected Probands



Relative Risk (RR) = 0.433 / 0.267 = 1.63

In siblings of affected compared to unaffected probands

8/30 are affected;

Risk = 0.267

Slide credit: Dale Nyholt

Relative risk to relatives
Recurrence risk to relatives

How much more likely are you to be diseased if your relative is affected compared to a person selected randomly from the population?

$$\text{Relative risk to relatives } (\lambda_R) = \frac{p(\text{affected} \mid \text{relative affected})}{p(\text{affected in population})} = \frac{K_R}{K}$$

How to estimate p(affected | relative affected) ?

- Collect population samples – cases infrequent
- Collect samples of case families and assess family members

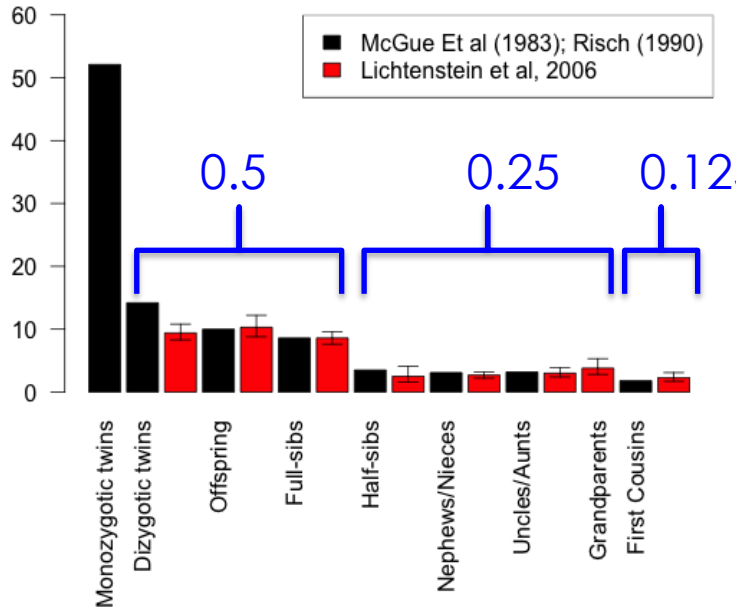
How to estimate p(affected in population) ?

- Census or national health statistics
  - Is definition of affected same in population sample as family sample
- Collect control families and assess family members

$$\text{If disease is not common} \quad \lambda_R = \frac{p(\text{sibling affected} \mid \text{case family})}{p(\text{sibling affected} \mid \text{control family})}$$

Schizophrenia risks to relatives

Relatives	Coefficient of relationship	Risch McGue et al	Lichtenstein et al Estimate	95% CI
Monozygotic twins	1	**52.1**		
Dizygotic twins	½	**14.2**		
Parent	½		**9.4**	**8.3 - 10.8**
Offspring	½	**10.0**	**10.3**	**8.8 - 12.2**
Full-sibs	½	**8.6**	**8.6**	**7.6 - 9.6**
Half-sibs	¼	**3.5**	**2.5**	**1.6 - 4.1**
Nephews/Nieces	¼	**3.1**	**2.7**	**2.2 - 3.2**
Uncles/Aunts	¼	**3.2**	**3.0**	**2.4 - 3.9**
Grandparents	¼		**3.8**	**2.8 - 5.3**
First Cousins	1/8	**1.8**	**2.3**	**1.7 – 3.1**
Offspring of 2 affected parents	½ but ascertained		**89**	**19 - 672**



Baseline risk, K = 0.85% McGue et al
= 0.407% Lichtenstein et al

Risch(1990) Linkage Strategies for Genetically Complex Traits AJHG
McGue et al (1983) Genetic Epidemiology 2: 99
Lichtenstein et al (2006) Recurrence risks for schizophrenia in a Swedish National Cohort.Psychological Medicine

For which disease is the genetic contribution more important?

Lifetime Risk 1%

Relative risk to 1st degree relatives: 10

Absolute Risk to 1st degree relatives 10%

Vs

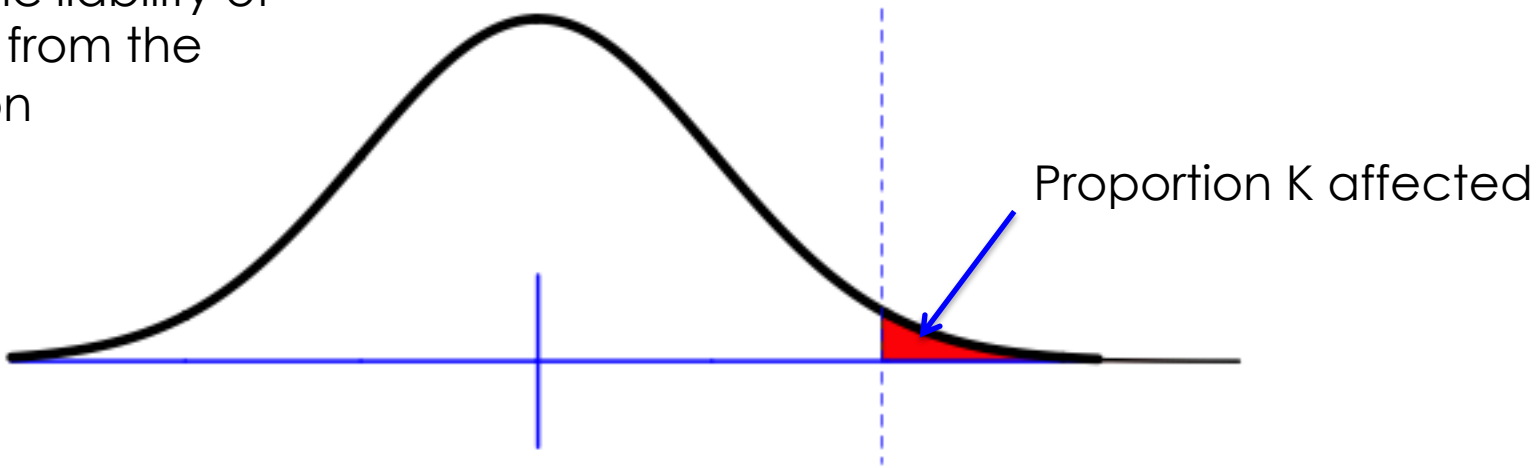
Lifetime Risk 15%

Relative risk to 1st degree relatives: 2

Absolute Risk to 1st degree relatives 30%

Liability threshold model

Phenotypic liability of a sample from the population



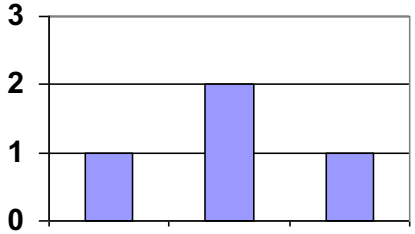
Assumption of normality

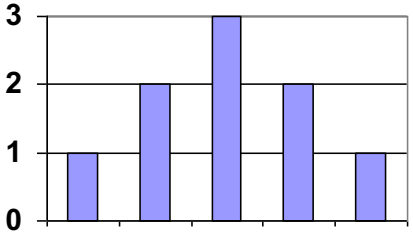
- Only appropriate for multifactorial disease
- i.e. more than a few genes but doesn't have to be highly polygenic
- Key – unimodal

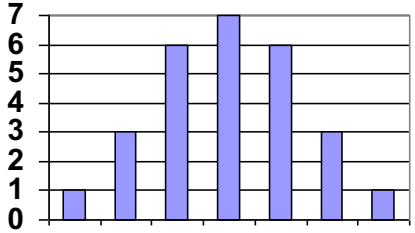
Does an underlying normality assumption make sense?

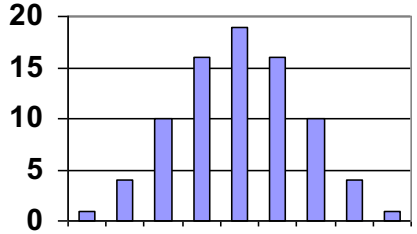
Assumes approximately normal distribution of liability
Makes sense for many genetic variants and environmental/noise factors

1 Locus	2 Locus	3 Locus	4 Locus
→ 3 Genotypes → 3 Classes	→ 9 Genotypes → 5 Classes	→ 27 Genotypes → 7 Classes	→ 81 Genotypes → 9 Classes



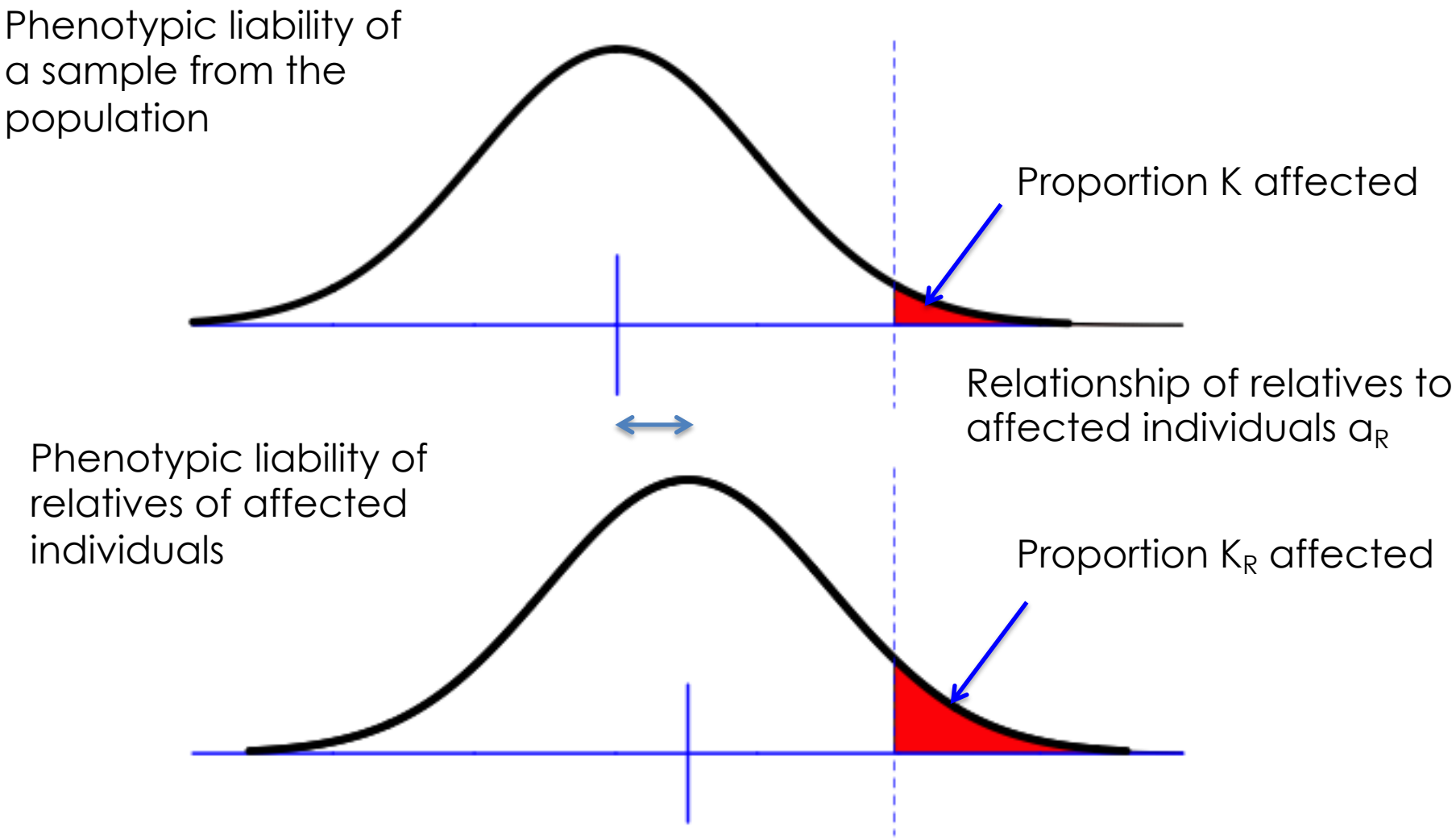






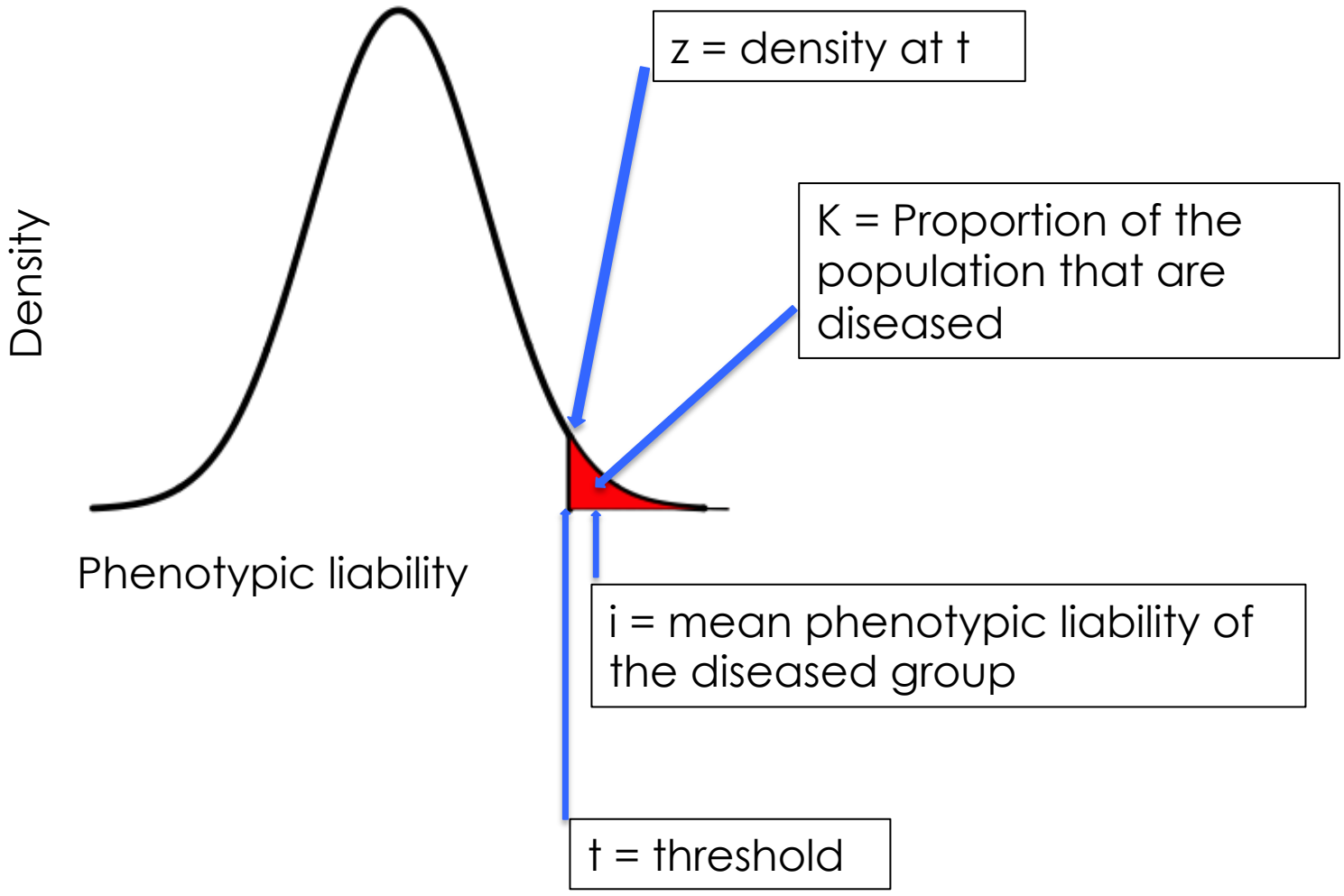
Each Locus has alleles R and r, R = risk alleles.
Each class has a different count of number of risk alleles

Falconer (1965)



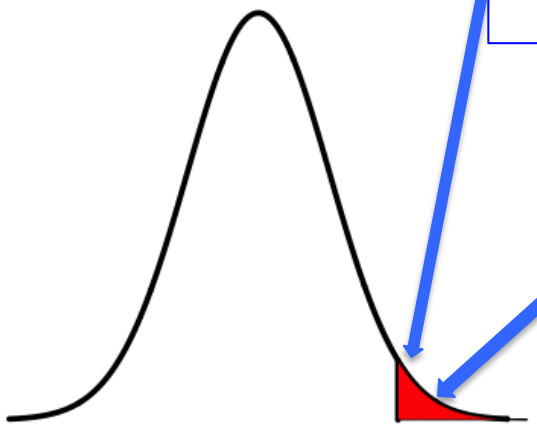
Using normal distribution theory what percentage of the variance in liability is attributable to genetic factors given K, K_R and a_R

Definitions



Liability Threshold Model –truncated normal distribution theory

Φ(x) =cumulative density until liability x
standard normal distribution function
φ (x) = probability density at x
Phi



Standard Deviation =1
$\sigma_p = 1$

z = density at t

z = φ (t) $\frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}t^2}$ = dnorm(t)

K = Proportion of the population that are diseased

K= 1-Φ(t) = 1-pnorm(t)

i = mean phenotypic liability of the diseased group

i= z/K "selection intensity"

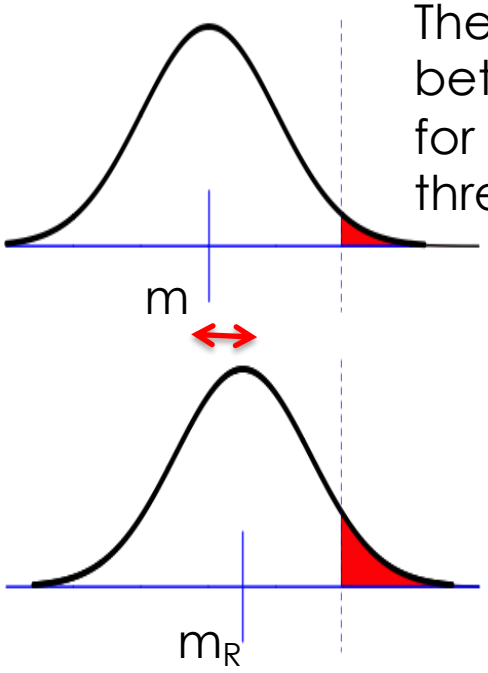
Variance in liability amongst the diseased individuals
= σ_p^2 (1-k), where k = i(i-t)

t = threshold

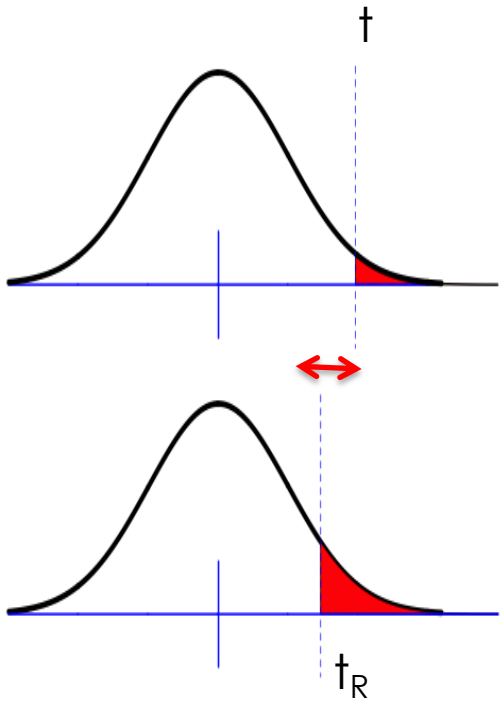
t= $\Phi^{-1}(1-K)$ = qnorm(1-K)

Inverse standard normal distribution (probit) function

Falconer (1965)



The difference between the means for the same threshold



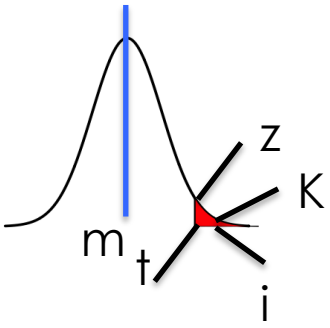
The difference between the thresholds when standardised to have the same mean

$$m_R - m = t - t_R$$

Given the difference in thresholds, and given known additive genetic relationship between relatives, what proportion of the total variance must be due to genetic factors

Falconer (1965) The inheritance of liability to certain diseases, estimated from incidences in relatives, Ann. Hum Genet. 29 51
Crittenden (1961) an interpretation o familial aggregation based on multiple genetic and environmental factors Ann NY Acad Sci 91 769

Calculate heritability of liability using regression theory



Y = phenotypic liability for individuals

Y_R = phenotypic liability for relatives of Y

$Y_R = a_R h^2 Y + \varepsilon$

For affected individuals Y = i

Expected phenotypic liability of relatives of those affected

$E(Y_R \mid Y>t) = m_R = m_R - m = t - t_R$

Substitute $t - t_R = a_R h^2 i$

Rearrange $h^2 = (t - t_R)/i a_R$

Falconer (1965) The inheritance of liability to certain diseases, estimated from incidences in relatives, Ann. Hum Genet. 29 51

Crittenden (1961) an interpretation o familial aggregation based on multiple genetic and environmental factors Ann NY Acad Sci. 91 769

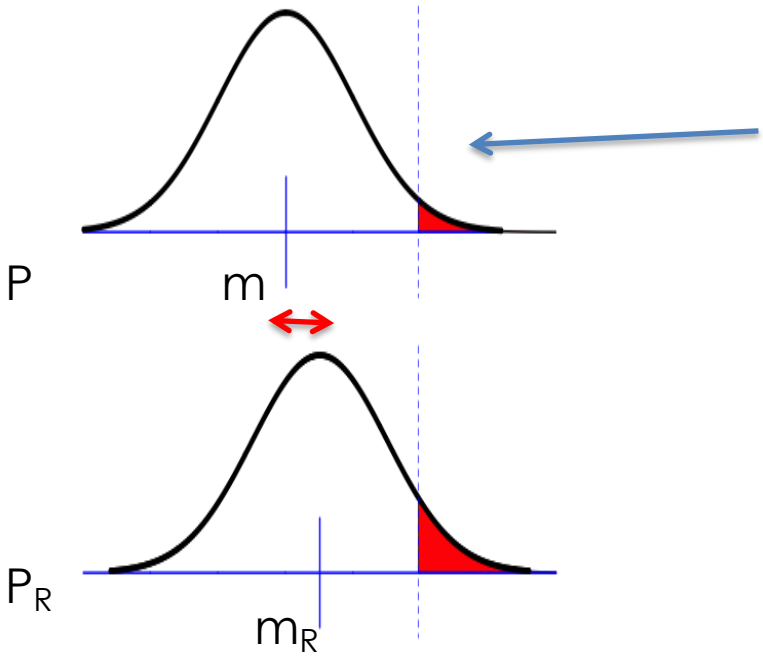
Assumptions made by Falconer (1965)

Assumption: Covariance between relatives reflects only shared additive genetic effects

Check: Use different types of relatives with different a_R and different u_R(dominance coefficient) and different shared environment to see consistency of estimates of h^2

Assumption: Phenotypic variance in relatives is unaffected by ascertainment on affected probands

Accounting for reduction in variance in relatives as a result of ascertainment on affected individuals



Variance in liability amongst the diseased individuals
= σ_p^2 (1-k), where k = i(i-t)

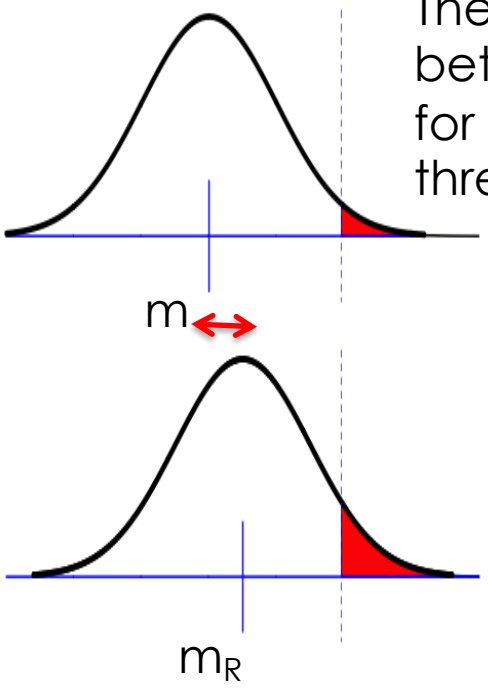
Variance in liability amongst relatives of diseased individuals

$V(Y_R \mid Y>t) = V(Y_R) - kCov(Y_R,Y)^2$

$= 1 - k(a_R h^2)^2 = 1 - ka_R^2 h^4$

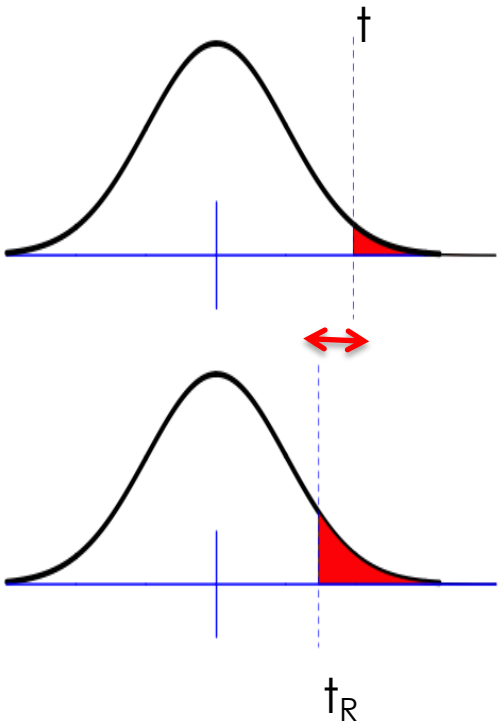
Reich, James, Morris (1972) The use of multiple thresholds in determining the mode of transmission of semi-continuous traits. Ann Hum Gen 36: 163.

Reich et al: heritability of liability



The difference between the means for the same threshold

The difference between the thresholds when standardised to have the mean 0 and variance 1



$$m_R - m = t - t_R \sqrt{1 - ka_R^2 h^4}$$

Reich, James, Morris (1972) The use of multiple thresholds in determining the mode of transmission of semi-continuous traits. Ann Hum Gen 36: 163.

Reich et al: heritability of liability

Y = phenotypic liability for individuals

Y_R = phenotypic liability for relatives of those with Y

$Y_R = a_R h^2 Y + \varepsilon$

For affected individuals Y = i

Expected phenotypic liability of relatives of those affected

$E(Y_R \mid Y>t) = m_R - m = t - t_R\sqrt{1 - ka_R^2h^4}$

Substitute $t - t_R\sqrt{1 - ka_R^2h^4} = a_R h^2 i$

Rearrange

$$h^2 = \frac{t - t_R\sqrt{1 - (1 - t/i)(t^2 - t_R^2)}}{a_R(i + (i - t)t_R^2)}$$

Also useful – calculation of t_R when K and h^2 are known

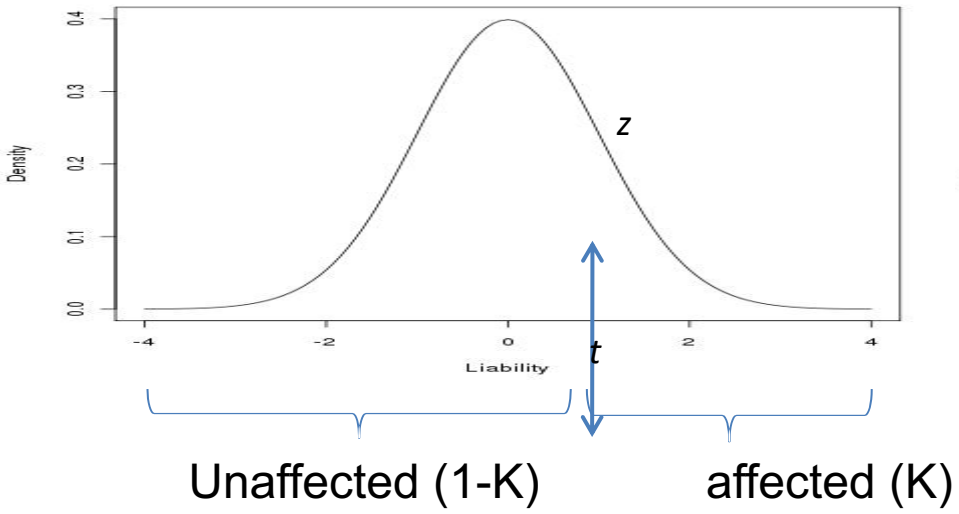
$$t_R = \frac{t - a_R i h^2}{\sqrt{1 - a_R^2 h^4 k}}$$

GREML: h^2-SNP for disease

- Observations are on disease scale but heritability is most interpretable on the liability scale
  - Use linear regression
  - Estimate on observed scale
  - Transform to Liability scale via Robertson Transformation
- Case-control samples are ascertained
  - Up date transformation
- Differences between case and control samples may reflect artefacts
  - Very stringent QC

Lee et al (2011) Estimating missing heritability for Disease from GWAS AJHG
Lee et al (2013) Estimation and partition of polygenic variation captured by common SNPs for AD, MS & Endo, HMG

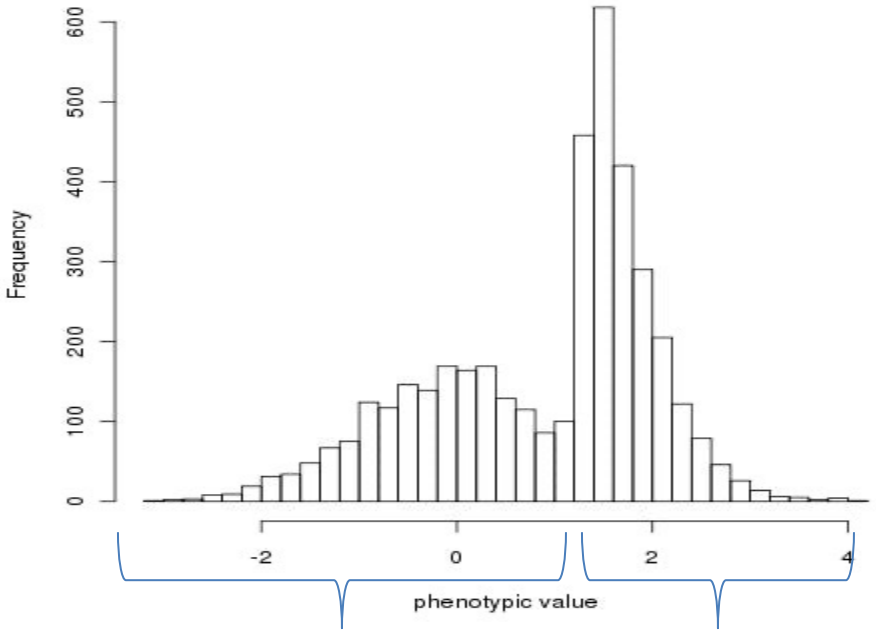
Ascertainment in case-control studies



$$h_l^2 = h_o^2 \frac{K(1-K)}{z^2}$$

Robertson (1950)
Appendix of Dempster and Lerner (1950)
See Lecture 1

$\widehat{h^2_{o_{cc}}}$ · Estimate of proportion of variance explained
· by SNP between cases and controls



$$h_l^2 = \widehat{h^2_{o_{cc}}} \frac{K(1-K)}{z^2} \frac{K(1-K)}{P(1-P)}$$

Lee et al (2011)AJHG
Zhou & Stephens (2013) Polygenic Modeling with Bayesian Sparse Linear Mixed Models PLoSG Text S3
Golan et al (2014) Measuring missing heritability: Inferring the contribution of common variants PNAS

Summary

If a disease affects 1% of the population and has heritability 80%

We have shown why these statements are consistent :

If an individual is affected ~8% of his/her siblings affected

If an MZ twin is affected ~50% of their co-twins are affected