

Genetics of common complex psychiatric disorders I

Mark Adams

Part 1: Biometrics

Mark Adams

Division of Psychiatry

mark.adams@ed.ac.uk

Genetics and Environmental Influences on Behaviour and Mental Health

What is a “*common*”, “*complex*” psychiatric disorder?

Common: Affects 1% or more of the population

Complex: Inheritance cannot be explained by a single gene

Psychiatric diagnoses

- Depressive disorder: Marked and persistent low mood and inability to feel pleasure.
- Anxiety disorder: Extensive, pervasive, unrealistic, and disabling worry.
- Bipolar disorder: Intense mood swings between mania and depression.
- Schizophrenia: Persistent hallucinations and delusions that severely impair functioning, highly disorganized thought and speech.
- Eating disorders: self-starvation or excessive over-eating, debilitating preoccupation with body image.
- Attention-deficit/hyperactivity disorder: Early onset, developmentally inappropriate and persistent hyperactivity and inattention.
- Autism spectrum disorders: Markedly impaired interpersonal interactions, non-goal-directed behaviours, and language development.

Prevalences

- Depression: 3% in a week
- Schizophrenia: 1% in lifetime
- Bipolar disorder: 2% in lifetime
- Anxiety disorder: 6% in a week

Why genetics?

Why use genetics to study mental health and psychiatric disorders?

- Biological understanding of genes, pathways
- Shared aetiology with other disorders
- Risk prediction
- Drug repurposing
- Causal analysis of environmental risk factors

Genetics of categorical traits

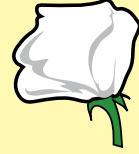
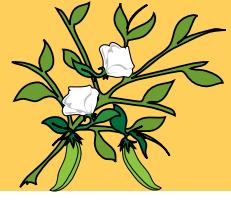
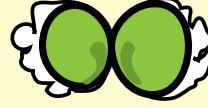
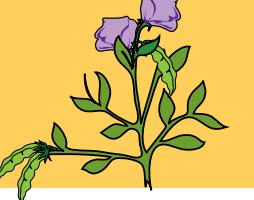
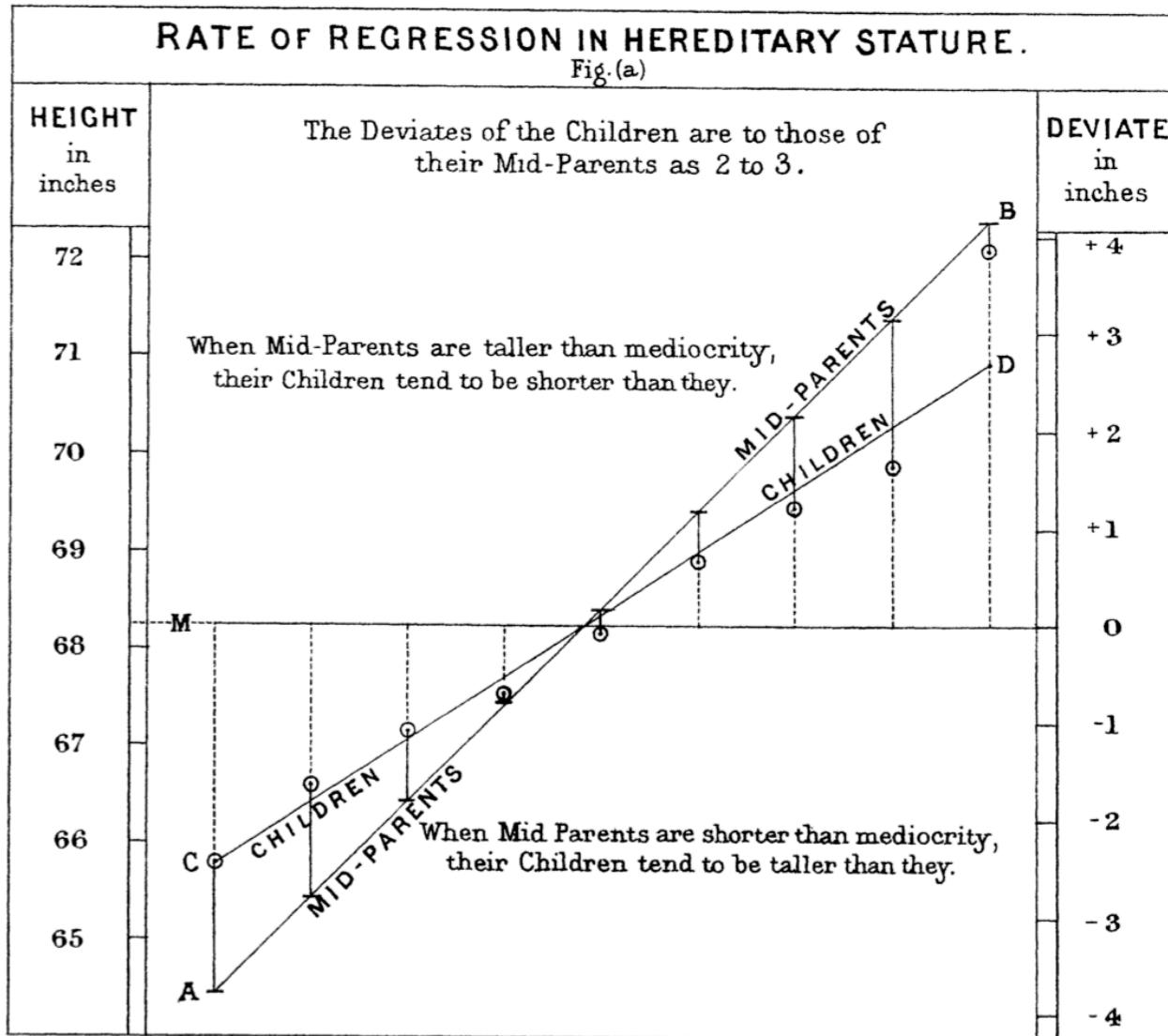
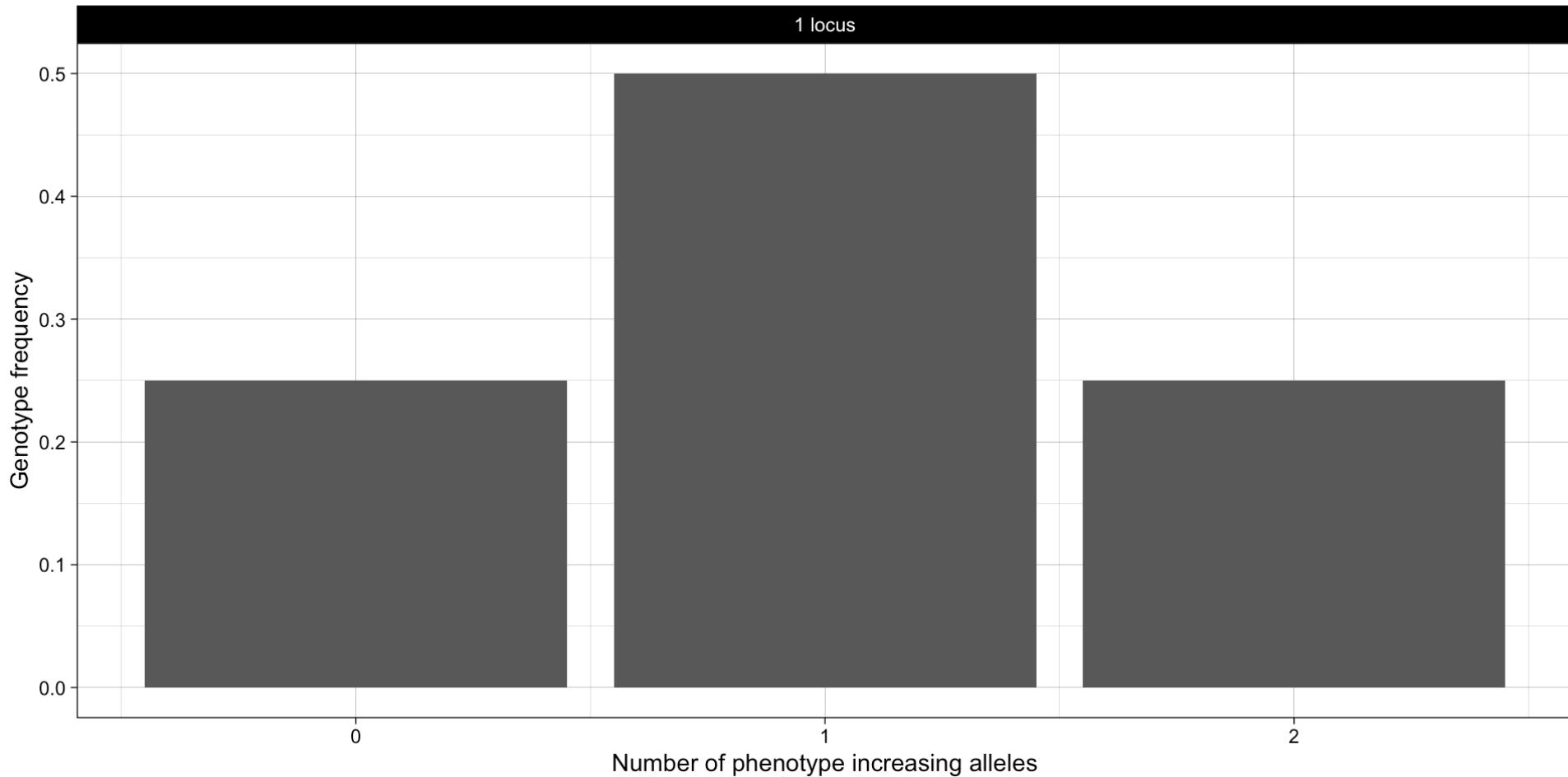
Seed		Flower		Pod		Stem	
Form	Cotyledons	Color		Form	Color	Place	Size
							
Grey & Round	Yellow	White		Full	Yellow	Axial pods, Flowers along	Long (6-7ft)
							Short ($\frac{3}{4}$ 1ft)
White & Wrinkled	Green	Violet		Constricted	Green	Terminal pods, Flowers top	Short ($\frac{3}{4}$ 1ft)
1	2	3		4	5	6	7

Diagram showing the seven “characters” observed by Mendel

Genetics of continuous traits



Reconciling categorical + continuous genetics = quantitative genetics



Single locus with additive effect

Single locus with two alleles, A_1 and A_2 , and four genotypes with values:

- $A_1A_1: 2a$
- A_1A_2 and $A_2A_1: a$
- $A_2A_2: 0$

If allele A_1 has a frequency of p and A_2 has frequency $1 - p$, then the genotype frequencies are:

- $A_1A_1: p^2$
- A_1A_2 and $A_2A_1: 2p(1 - p)$
- $A_2A_2: (1 - p)^2$

Genotype	Frequency	Value	Freq. × Val.
A ₁ A ₁	p^2	2a	$2p^2a$
A ₁ A ₂	$p(1 - p)$	a	$p(1 - p)a$
A ₂ A ₁	$p(1 - p)$	a	$p(1 - p)a$
A ₂ A ₂	$(1 - p)^2$	0	0
Mean =			$2pa$
Variance =			$2p(1 - p)a^2$

Genetic values

Each individual i has a genetic value at locus j :

$$g_{ij} \in [0, a_j, 2a_j]$$

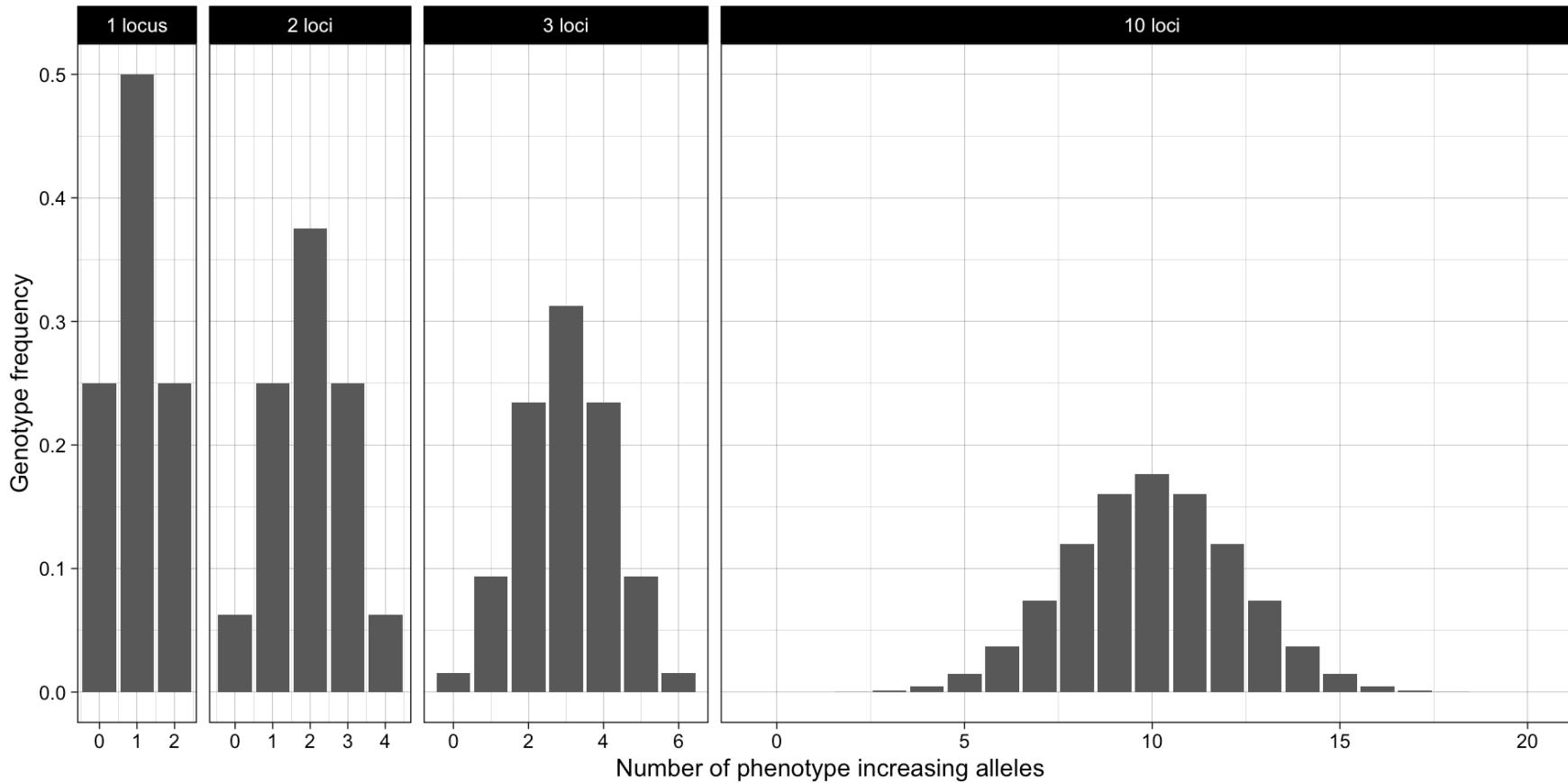
Total genetic value from M loci:

$$G_i = \sum_{j=1}^M g_{ij}$$

$$\text{var}(G) = \sum_{j=1}^M 2p_j(1 - p_j)a_j^2$$

Polygenic traits are quantitative traits

Adding up effects from a large number of genetic loci makes a continuous phenotype.



Biometrics

What are the sources of family resemblance? How do we quantify them numerically?

Heritability

Proportion of similarity in phenotypes that can be attributed to similarity in genotypes.

Model: Phenotype (P) = Genotype (G) + Environment (E)

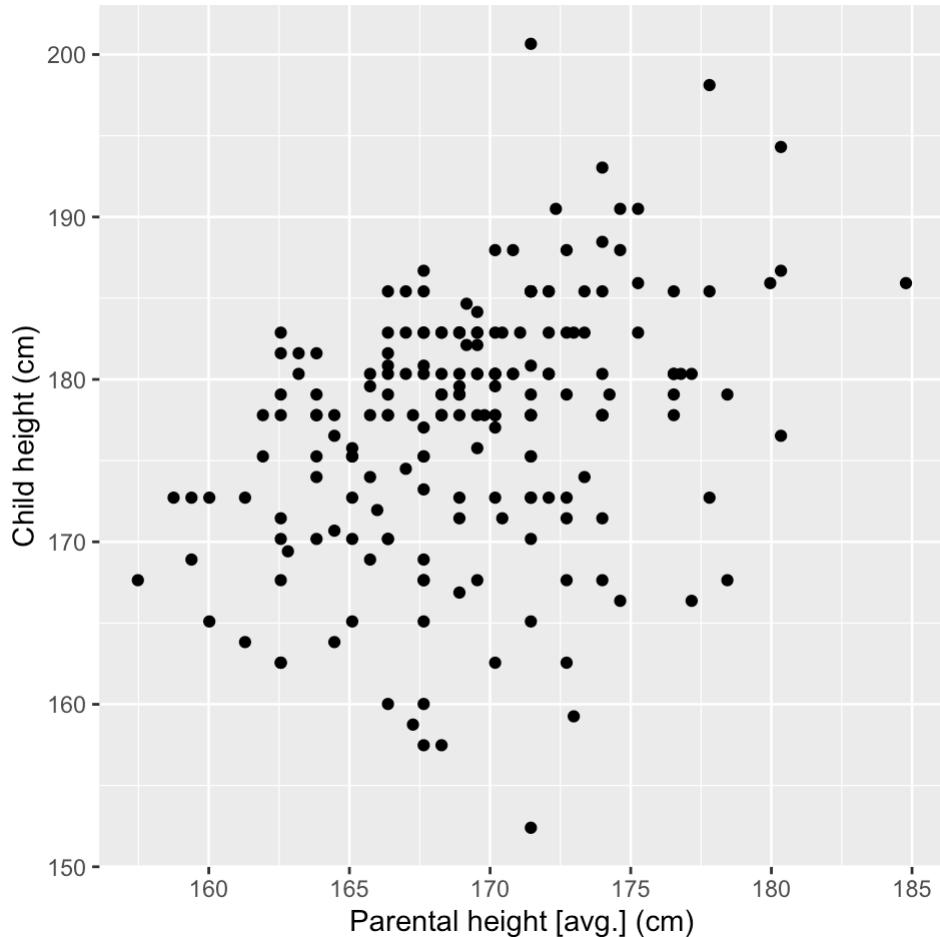
Variance decomposition

$$\text{var}(P) = \text{var}(G) + \text{var}(E)$$

Proportion of variance

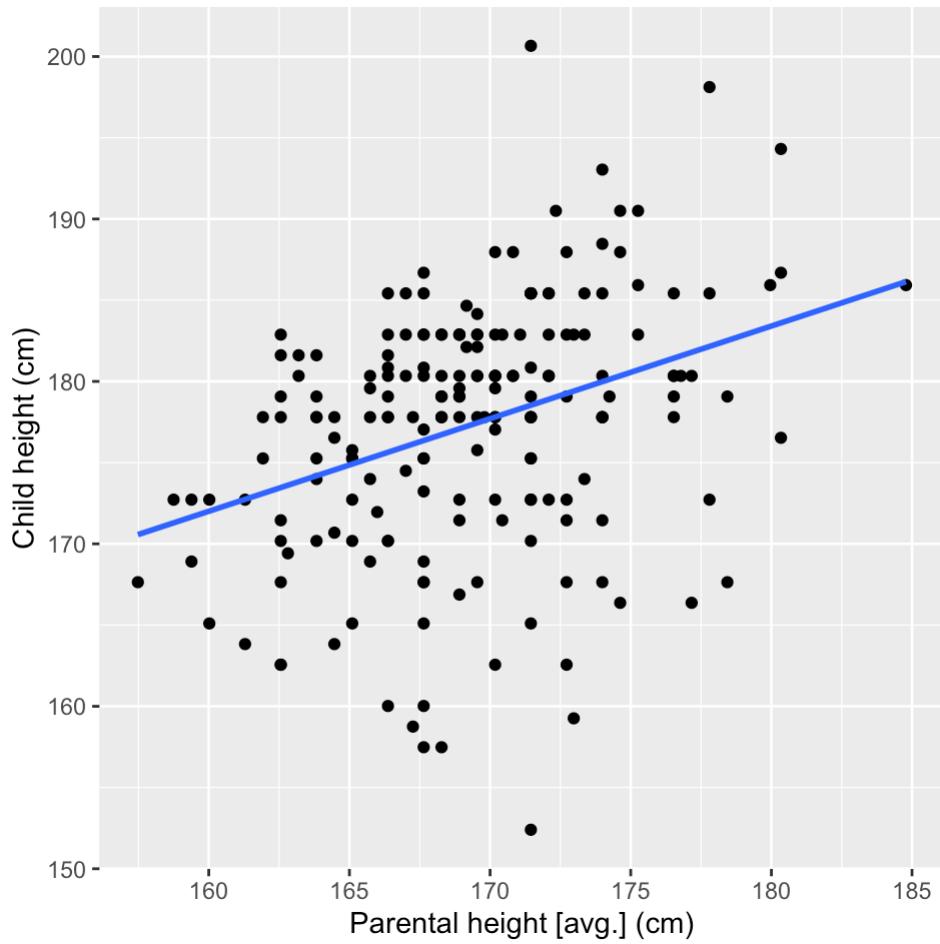
$$H^2 = \frac{\text{var}(G)}{\text{var}(P)}, e^2 = \frac{\text{var}(E)}{\text{var}(P)}, H^2 + e^2 = 1$$

How to estimate heritability from data



Plot of child (offspring) height versus the average of their parents' heights. What is a statistic that can be used to summarise the relationship between these two variables?

How to estimate heritability from data



$$\beta = \frac{\text{cov}(X,Y)}{\text{var}(X)}$$

Estimate the beta coefficient (slope) for a simple regression from the covariance between predictor (X) and outcome (Y) variable divided by the variance of the predictor (X).

Simple model of genetic and environmental effects

$$P = A + E$$

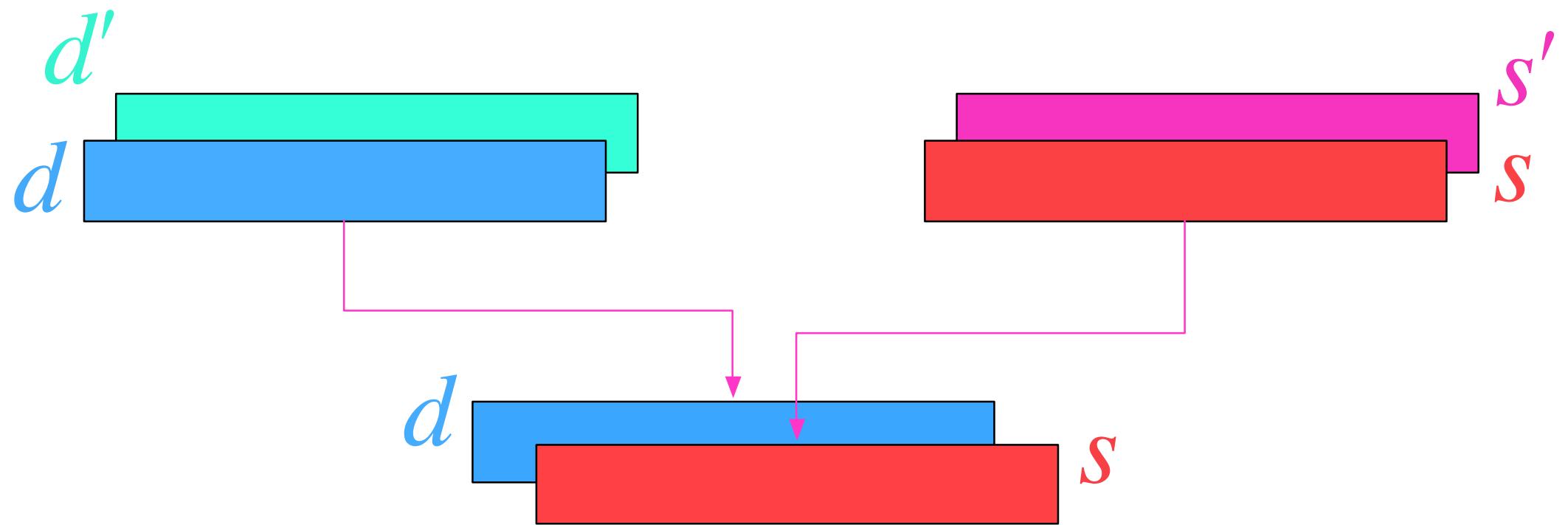
The phenotype value P is influenced by an additive genetic effect A and an environmental effect E .

Simple model of genomics

$$A = d + s$$

Each individual has two copies of the genome, one inherited from each parent.

Simple model of inheritance

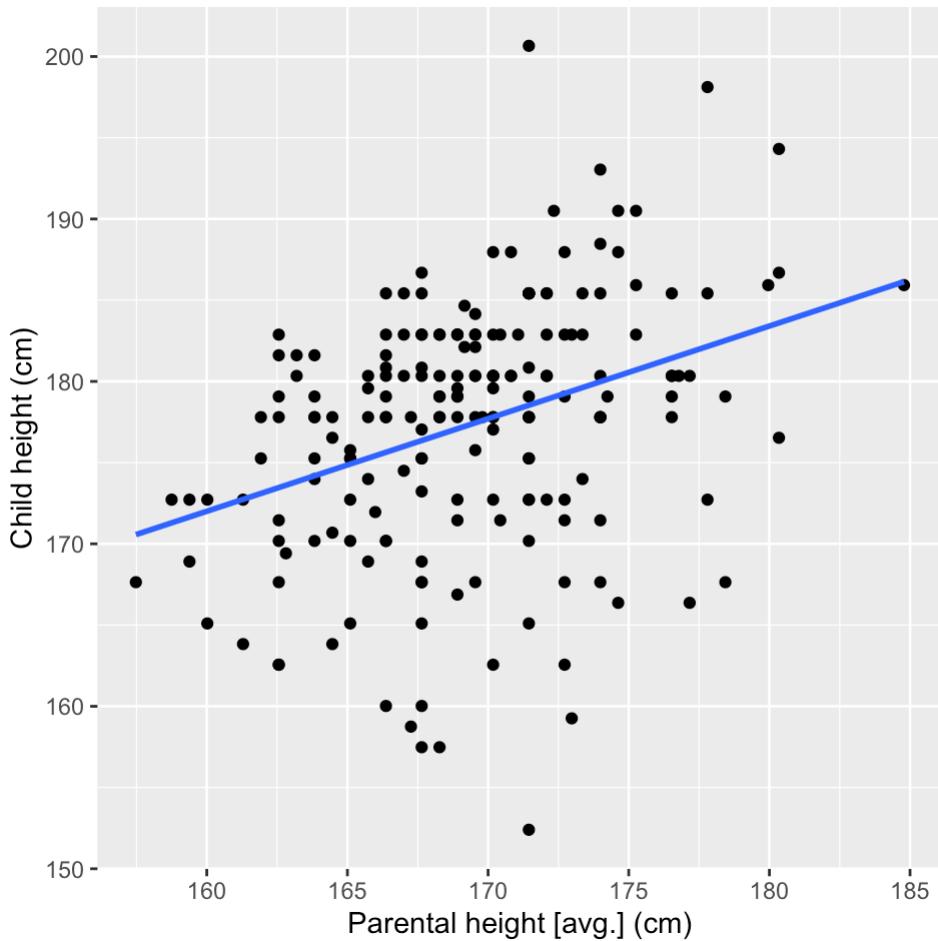


Simple model of genetics, environment, and inheritance

Phenotype (P) value is the sum of the two genetic values plus an environmental value (e).

- Mother's phenotype: $P_d = d + d' + e_d$
- Father's phenotype: $P_s = s + s' + e_s$
- Child's phenotype: $P_o = d + s + e_o$

Regression equation



$$\beta = \frac{\text{cov}(X, Y)}{\text{var}(X)}$$

- X = average of parents' phenotypes
- Y = offspring phenotype

Therefore,

$$\beta = \frac{\text{cov}\left(\frac{P_d + P_s}{2}, P_o\right)}{\text{var}\left(\frac{P_d + P_s}{2}\right)}$$

Parent–offspring covariance

$$\text{cov}\left(\frac{P_d + P_s}{2}, P_o\right)$$

$$= \text{cov}\left(\frac{d + d' + e_d + s + s' + e_s}{2}, d + s + e_o\right)$$

Parent-offspring covariance

Expand the terms. Recall that:

$$\begin{aligned}\text{cov}(A + X, B + Y) &= \\ \text{cov}(A, B) + \text{cov}(A, Y) + \text{cov}(X, B) + \text{cov}(X, Y)\end{aligned}$$

Thus we can do a pairwise expansion to:

$$= \text{cov}\left(\frac{d}{2} + \frac{d'}{2} + \frac{e_d}{2} + \frac{s}{2} + \frac{s'}{2} + \frac{e_s}{2}, d + s + e_o\right)$$

$$= \text{cov}\left(\frac{d}{2}, d\right) + \text{cov}\left(\frac{d'}{2}, d\right) + \cdots + \text{cov}\left(\frac{e_s}{2}, e_o\right)$$

Simplifications

Some terms can be simplified.

Covariance between a genetic effect and itself

$$\text{cov}\left(\frac{d}{2}, d\right), \text{cov}\left(\frac{s}{2}, s\right)$$

Simplifies to:

$$\text{cov}\left(\frac{d}{2}, d\right) = \frac{1}{2} \text{cov}(d, d) = \frac{1}{2} \text{var}(d)$$

Assumptions

For some terms we might make an assumption that they are equal to 0.

Covariance between genetic effects from the same parent

$$\text{cov}\left(\frac{d'}{2}, d\right), \text{cov}\left(\frac{s'}{2}, s\right)$$

Covariance between genetic effects from different parents

$$\text{cov}\left(\frac{d'}{2}, s\right), \text{cov}\left(\frac{s'}{2}, d\right)$$

Covariance between parent and offspring environment effects

$$\text{cov}\left(\frac{e_d}{2}, e_o\right), \text{cov}\left(\frac{e_s}{2}, e_o\right)$$

Covariance between parental genetic and offspring environmental effects

$$\text{cov}\left(\frac{d}{2}, e_o\right), \text{cov}\left(\frac{s}{2}, e_o\right)$$

Using those assumptions the parent–offspring covariance simplifies to

$$\text{cov}\left(\frac{P_d + P_s}{2}, P_o\right) = \frac{\text{var}(d) + \text{var}(s)}{2}$$

Parent variance

The denominator in the regression equation was

$$\text{var}\left(\frac{P_d + P_s}{2}\right)$$

Using the identity

$$\text{var}(aX + bY) = a^2 \text{var}(X) + b^2 \text{var}(Y) + 2ab\text{cov}(X, Y)$$

the variance of the average parental phenotypes is:

$$\text{var}\left(\frac{P_d + P_s}{2}\right) = \text{var}\left(\frac{1}{2}P_d + \frac{1}{2}P_s\right)$$

$$= \left(\frac{1}{2}\right)^2 \text{var}(P_d) + \left(\frac{1}{2}\right)^2 \text{var}(P_s) + 2 \cdot \frac{1}{2} \cdot \frac{1}{2} \text{cov}(P_d, P_s)$$

$$= \frac{1}{4} \text{var}(P_d) + \frac{1}{4} \text{var}(P_s) + \frac{1}{2} \text{cov}(P_d, P_s)$$

If we assume as above that there is no covariation between parental effects ($\text{cov}(P_d, P_s) = 0$), this simplifies to

$$= \frac{\text{var}(P_d) + \text{var}(P_s)}{4}$$

Thus the regression equation is:

$$\begin{aligned}\beta &= \frac{\text{cov}\left(\frac{P_d+P_s}{2}, P_o\right)}{\text{var}\left(\frac{P_d+P_s}{2}\right)} \\ &= \frac{\frac{\text{var}(d)+\text{var}(s)}{2}}{\frac{\text{var}(P_d)+\text{var}(P_s)}{4}} \\ &= 2 \frac{\text{var}(d) + \text{var}(s)}{\text{var}(P_d) + \text{var}(P_s)}\end{aligned}$$

Previously we defined

$$A = d + s$$

thus

$$\text{var}(A) = \text{var}(d) + \text{var}(s)$$

and assume variances in parental phenotypes are equal

$$\text{var}(P_d) = \text{var}(P_s) = \text{var}(P)$$

Then substitute into the regression equation

$$\begin{aligned}\beta &= 2 \frac{\text{var}(d) + \text{var}(s)}{\text{var}(P_d) + \text{var}(P_s)} \\ &= 2 \frac{\text{var}(A)}{\text{var}(P) + \text{var}(P)} \\ &= 2 \frac{\text{var}(A)}{2\text{var}(P)} \\ &= \frac{\text{var}(A)}{\text{var}(P)} \\ &= h^2\end{aligned}$$

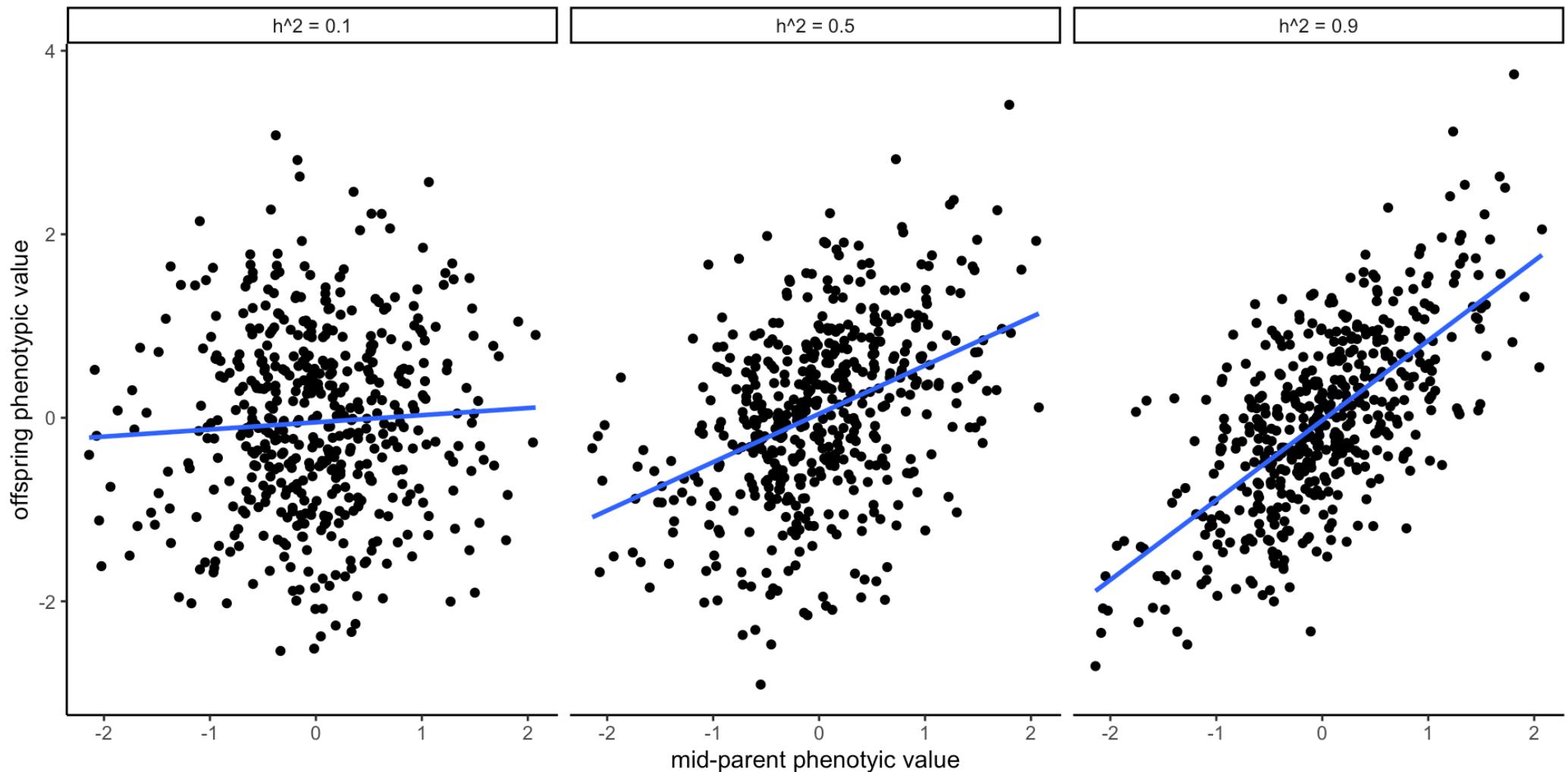
Height data

$$\text{cov}\left(\frac{P_d+P_s}{2}, P_o\right) = 12.57$$

$$\text{var}\left(\frac{P_d+P_s}{2}\right) = 22.04$$

$$h^2 = 12.57 / 22.04 = 0.57$$

Parent and offspring phenotypes become more highly correlated as heritability increases.



Mini review: What assumptions have we made when estimating h^2 ?

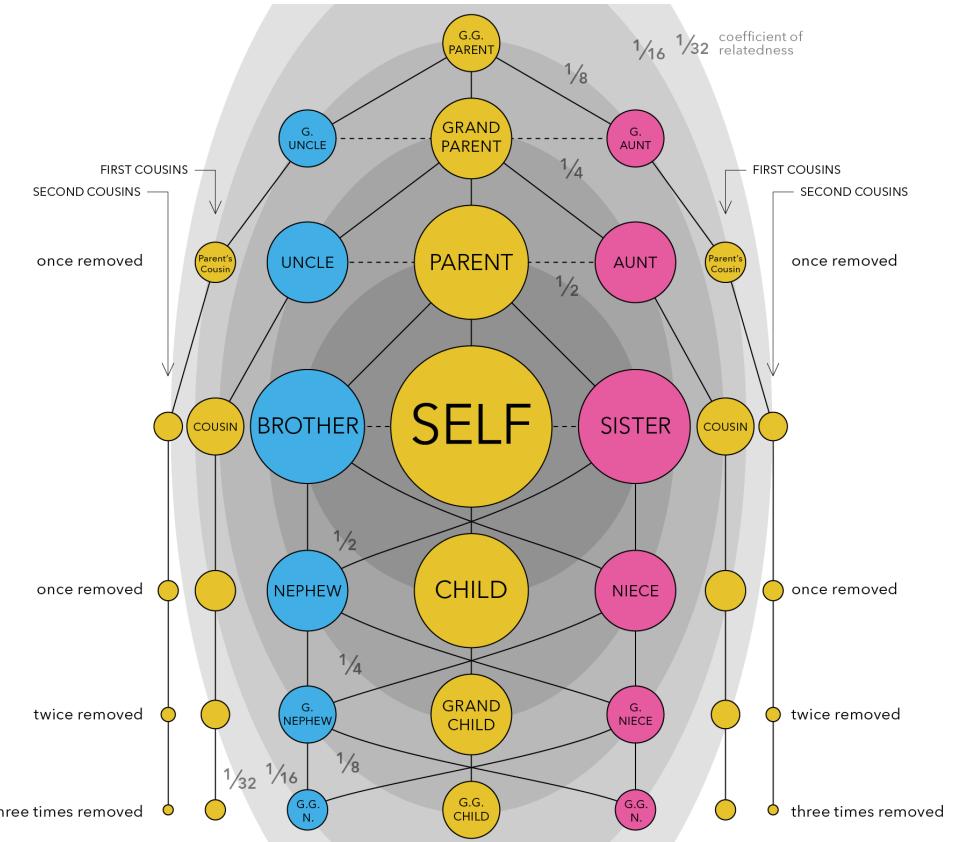


Generalising to other relatives

Heritability can also be estimated from resemblance between different types of related pairs. The general equation is:

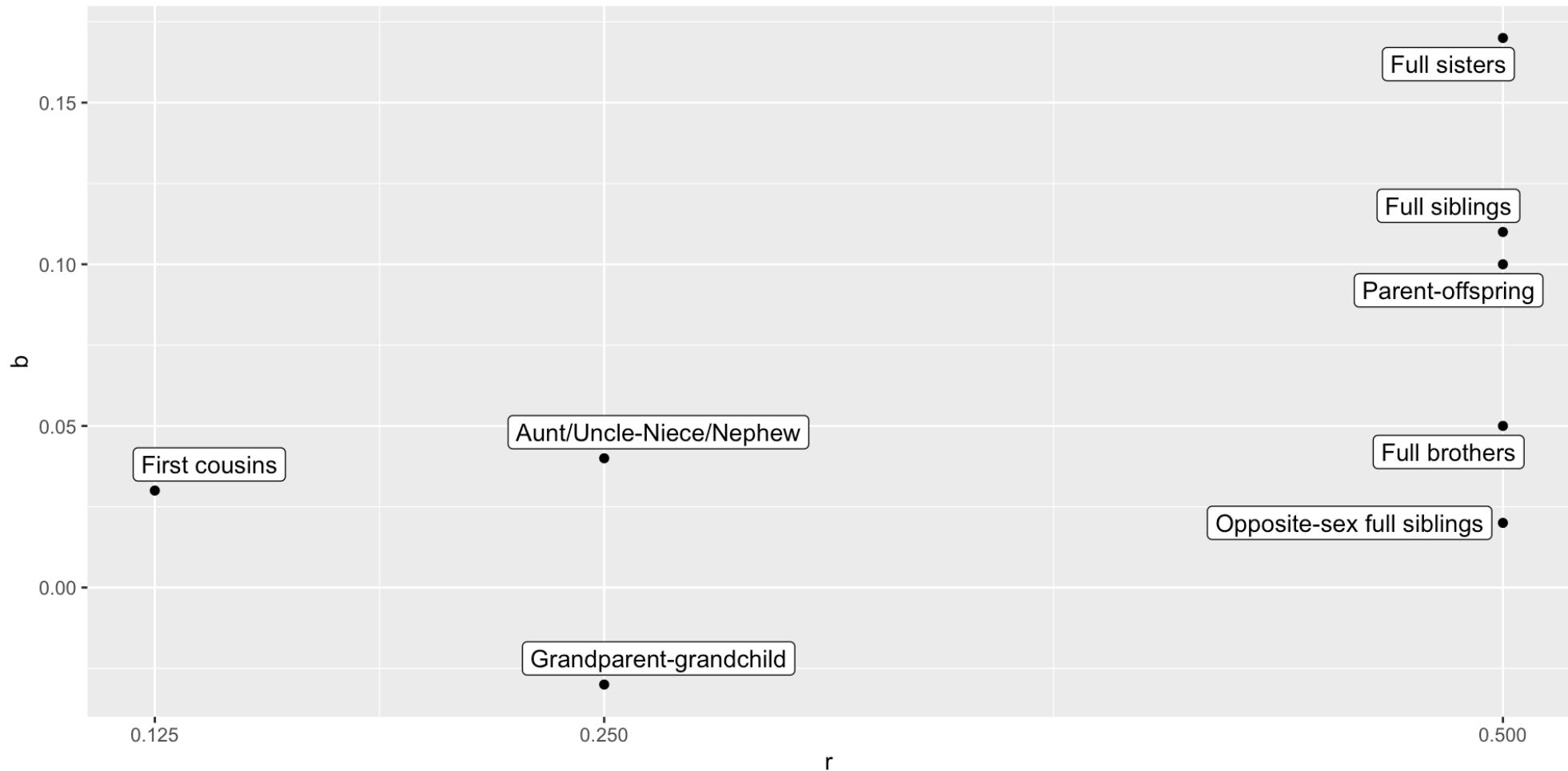
$$h^2 = \frac{b}{r}$$

b = regression coefficient
 r = relatedness coefficient
("coefficient of additive variance")



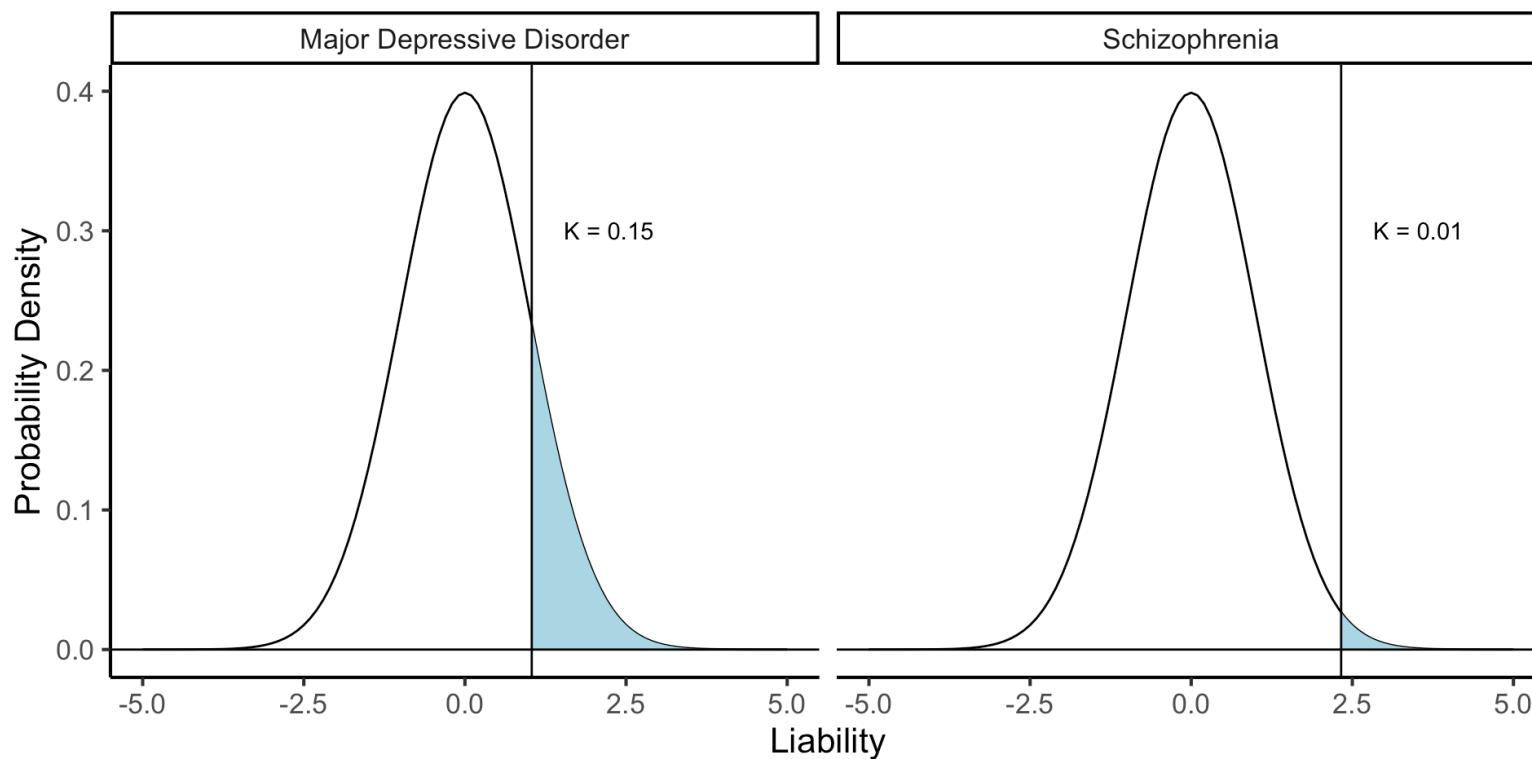
Example data: depression scores

Correlation of depression scores for different pairs of relatives



Recurrence risk to relatives

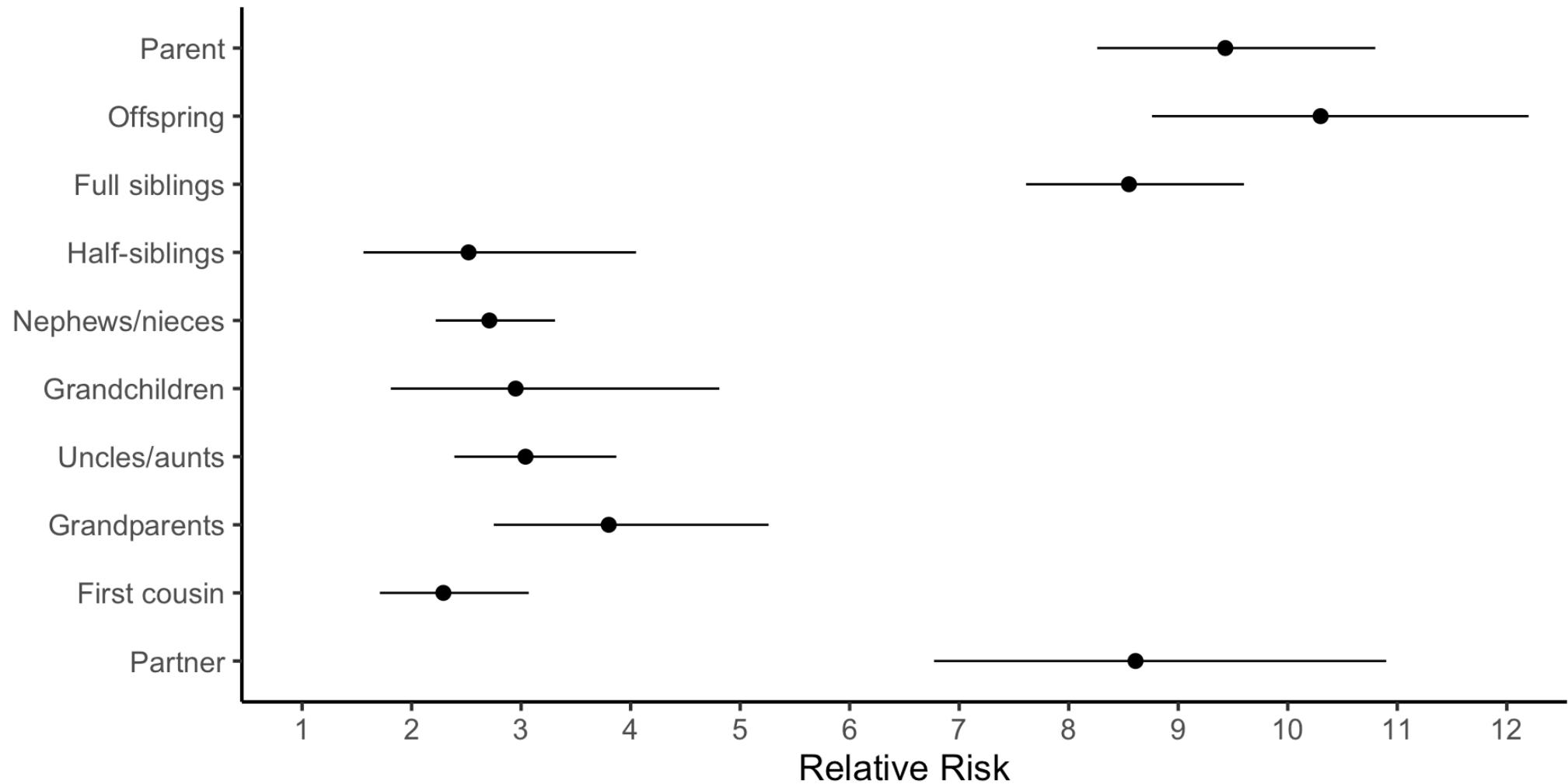
$$\lambda_R = \frac{P(\text{affected} | \text{relative affected})}{P(\text{affected in population})} = \frac{K_R}{K}$$



Example:

- $K_{\text{sib}} = P(\text{affected} | \text{sibling affected}) = 0.09$
- $K = P(\text{affected in population}) = 0.02$
- $\frac{K_{\text{sib}}}{K} = \frac{0.09}{0.02} = 4.5$

Recurrence risk for schizophrenia



Recurrence risk and heritability

- Score unaffected = 0, affected = 1
- If population prevalence is K , then phenotypic variance is $V_P = K(1 - K)$ (Bernoulli distribution)

- Y = score of individual (proband)
- Y_R = score of relative of proband
- Expectation: $E[Y] = E[Y_R] = K$
- $K_R = E[Y_R | Y = 1]$
- Probability that both Y and $Y_R = 1$: $E[YY_R] = K \times K_R$

$$\begin{aligned}\text{cov}(Y, Y_R) &= E[YY_R] - E[Y]E[Y_R] \\ &= K \times K_R - K^2\end{aligned}$$

$$\text{cov}(Y, Y_R) = E[YY_R] - E[Y]E[Y_R]$$

$$= K \times K_R - K^2$$

$$= K(K_R - K)$$

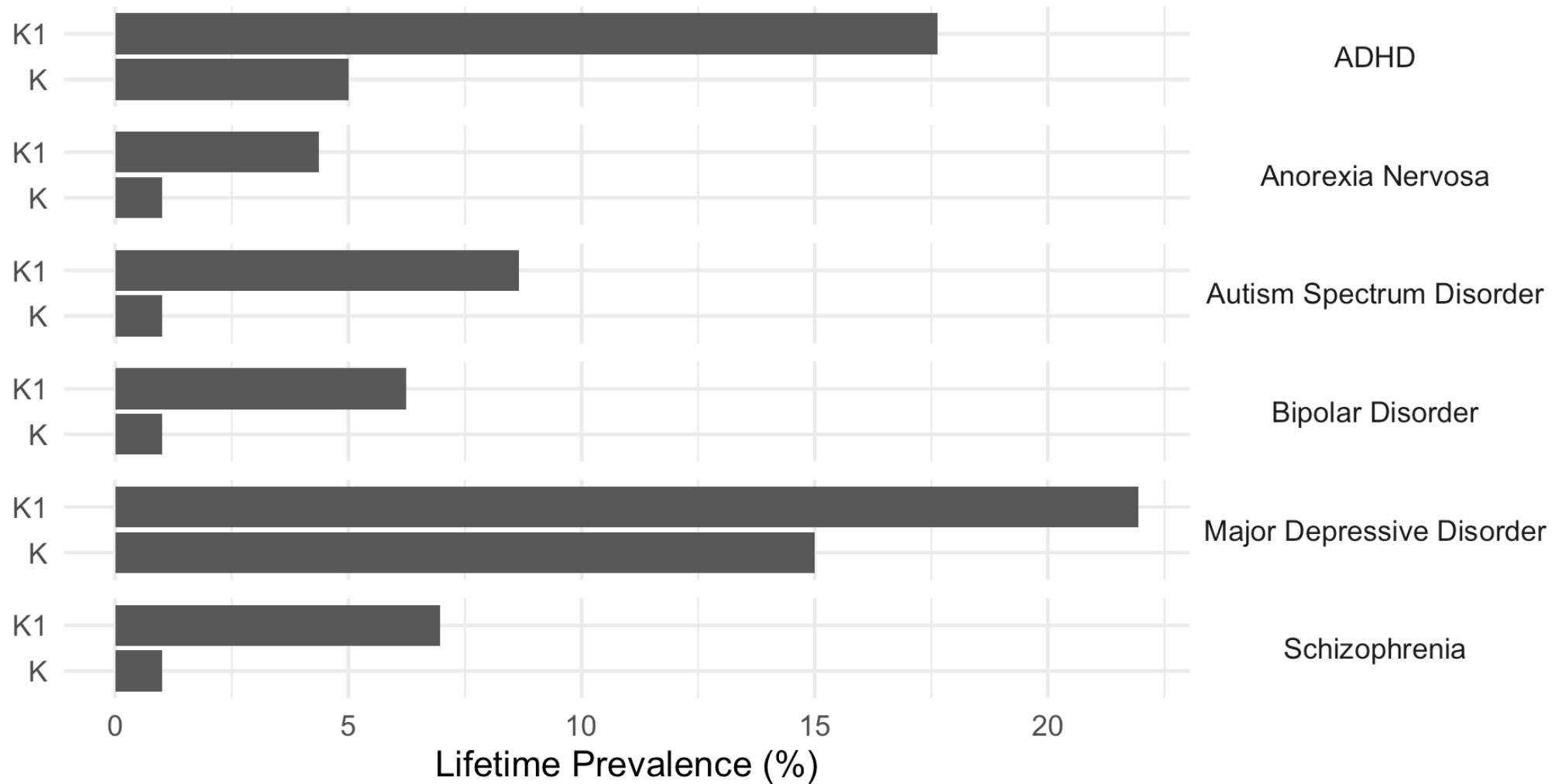
$$= K^2(\frac{K_R}{K} - 1)$$

$$= K^2(\lambda_R - 1)$$

Heritability estimate

$$\begin{aligned} h^2 &= \frac{\text{COV}_R}{rV_P} \\ &= \frac{K^2(\lambda_R - 1)}{rK(1 - K)} \\ &= \frac{K(\lambda_R - 1)}{r(1 - K)} \end{aligned}$$

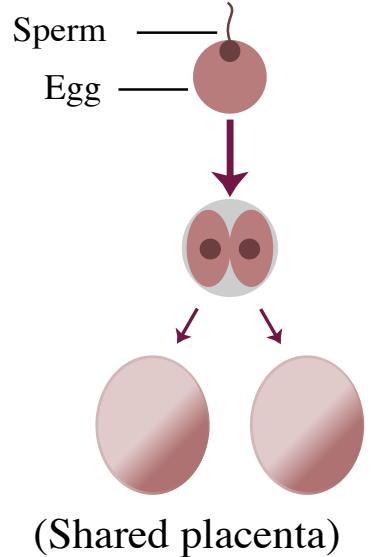
Recurrence risk of psychiatric disorders



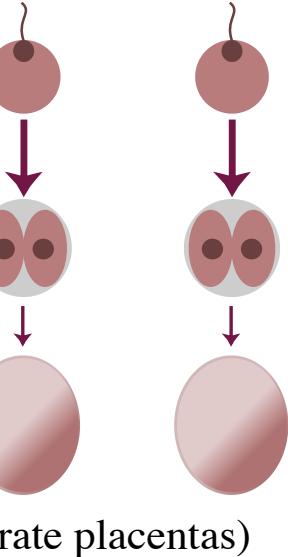
Estimating environmental effects

Contrast pairs of relatives that have comparable environmental similarity but different genetic similarity.

Identical
(Monozygotic)



Fraternal
(Dizygotic)



- Monozygotic (MZ) twins
 $r = 1.0$
- Dizygotic (DZ) twins $r = 0.5$

Additive genetic and shared environment effects

Add a shared (C or “common”) environment to the basic genetic model, to capture similarity between relatives attributable to environmental factors. E represents the unique, non-shared environment.

$$P = A + C + E$$

$$h^2 = \frac{\text{var}(A)}{\text{var}(P)}, c^2 = \frac{\text{var}(C)}{\text{var}(P)}, e^2 = \frac{\text{var}(E)}{\text{var}(P)}$$

$$h^2 + c^2 + e^2 = 1$$

Twin correlations

MZ twins: $r_{MZ} = h^2 + c^2$

DZ twins: $r_{DZ} = \frac{1}{2}h^2 + c^2$

Solve for genetic similarity (h^2)

Calculate difference between MZ and DZ correlations

$$r_{MZ} - r_{DZ} = (h^2 + c^2) - \left(\frac{1}{2}h^2 + c^2\right)$$

$$r_{MZ} - r_{DZ} = h^2 - \frac{1}{2}h^2 + c^2 - c^2$$

$$r_{MZ} - r_{DZ} = \frac{1}{2}h^2$$

$$h^2 = 2(r_{MZ} - r_{DZ})$$

Substitute h^2 into MZ equation and solve for shared environment similarity (c^2)

$$r_{\text{MZ}} = h^2 + c^2$$

$$r_{\text{MZ}} = 2(r_{\text{MZ}} - r_{\text{DZ}}) + c^2$$

$$r_{\text{MZ}} - 2(r_{\text{MZ}} - r_{\text{DZ}}) = c^2$$

$$c^2 = r_{\text{MZ}} - 2r_{\text{MZ}} + 2r_{\text{DZ}}$$

$$c^2 = 2r_{\text{DZ}} - r_{\text{MZ}}$$

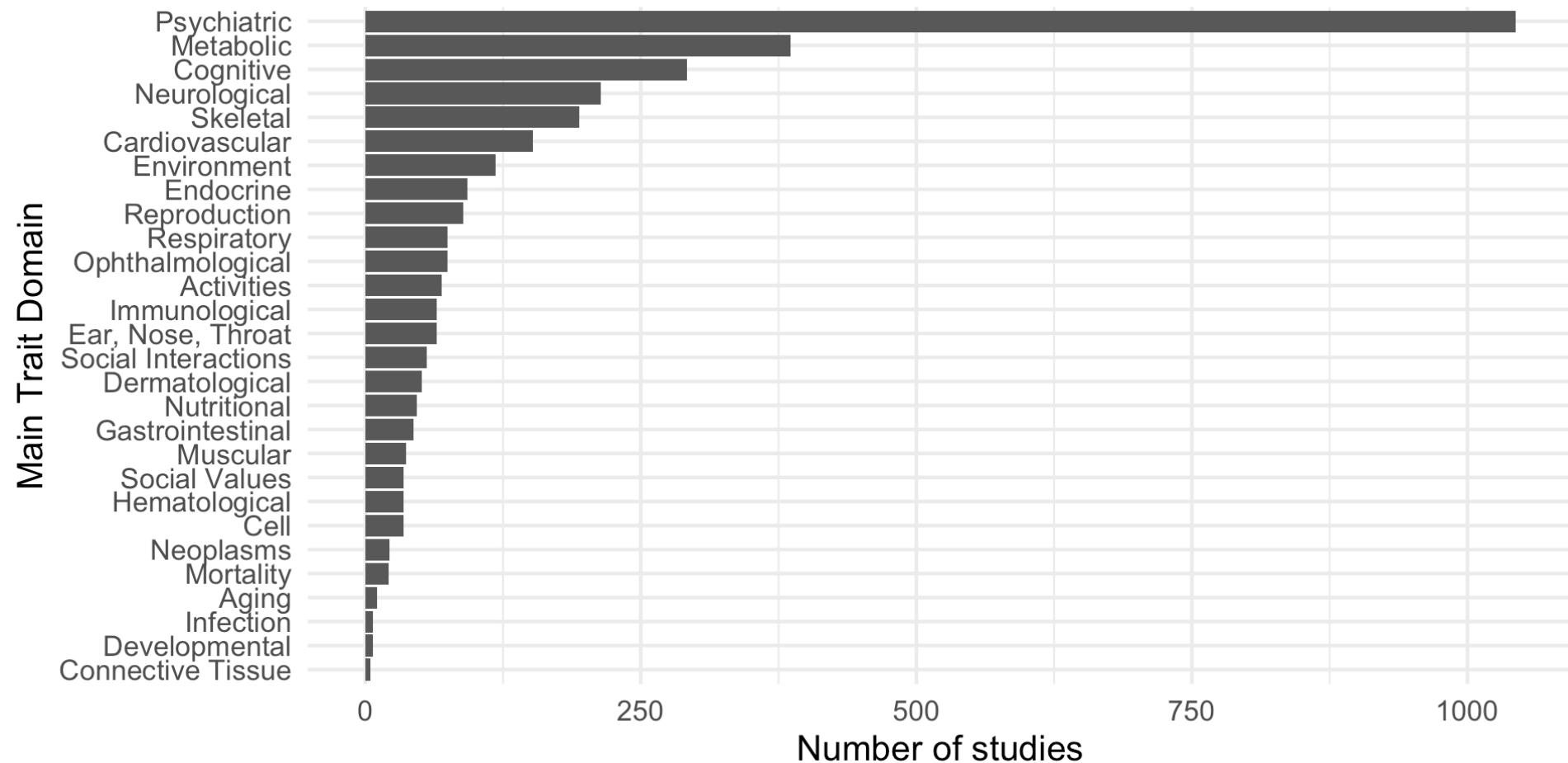
Therefore from MZ and DZ twin correlations we can estimate:

$$h^2 = 2(r_{\text{MZ}} - r_{\text{DZ}})$$

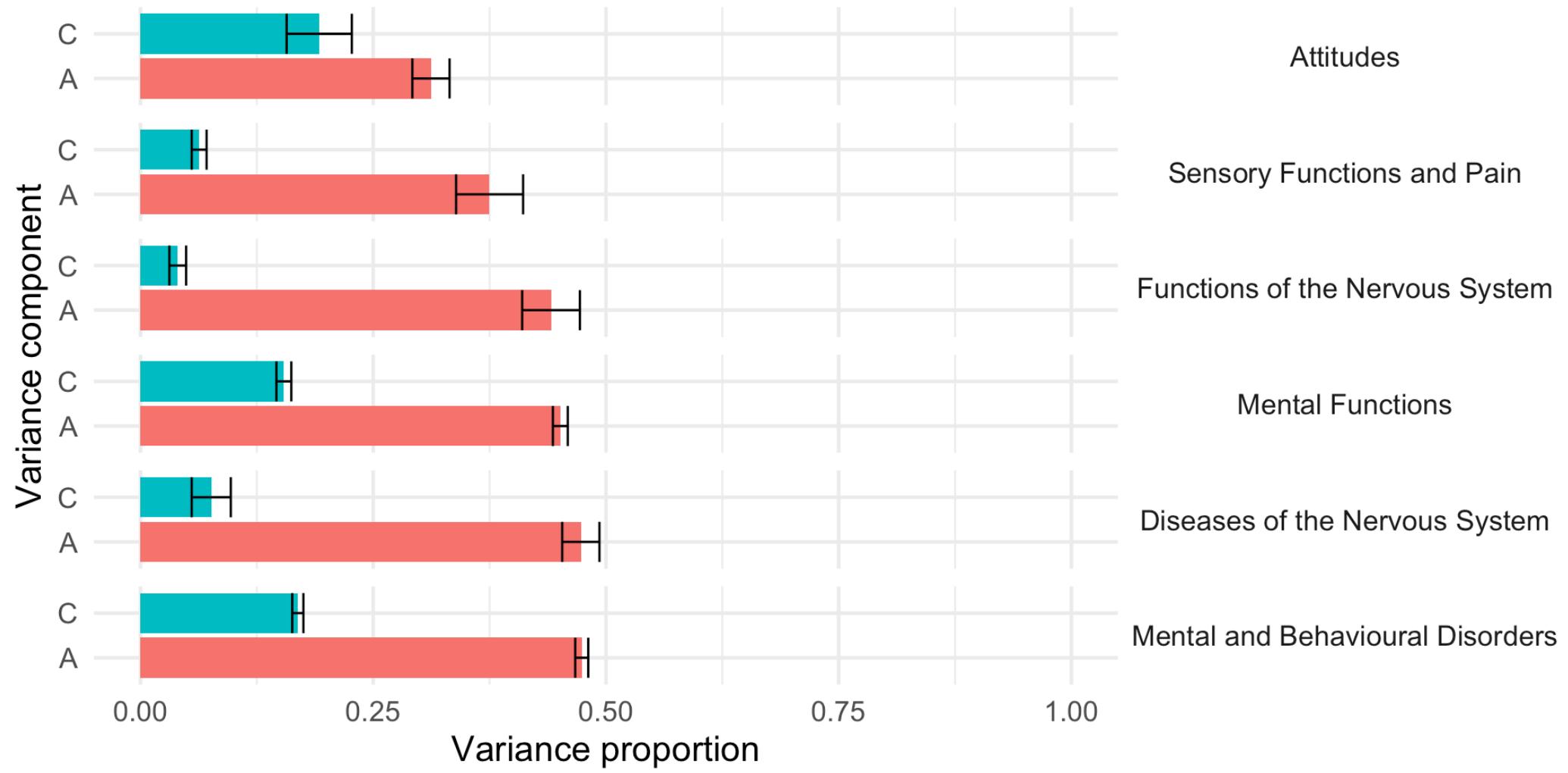
$$c^2 = 2r_{\text{DZ}} - r_{\text{MZ}}$$

$$e^2 = 1 - h^2 - c^2$$

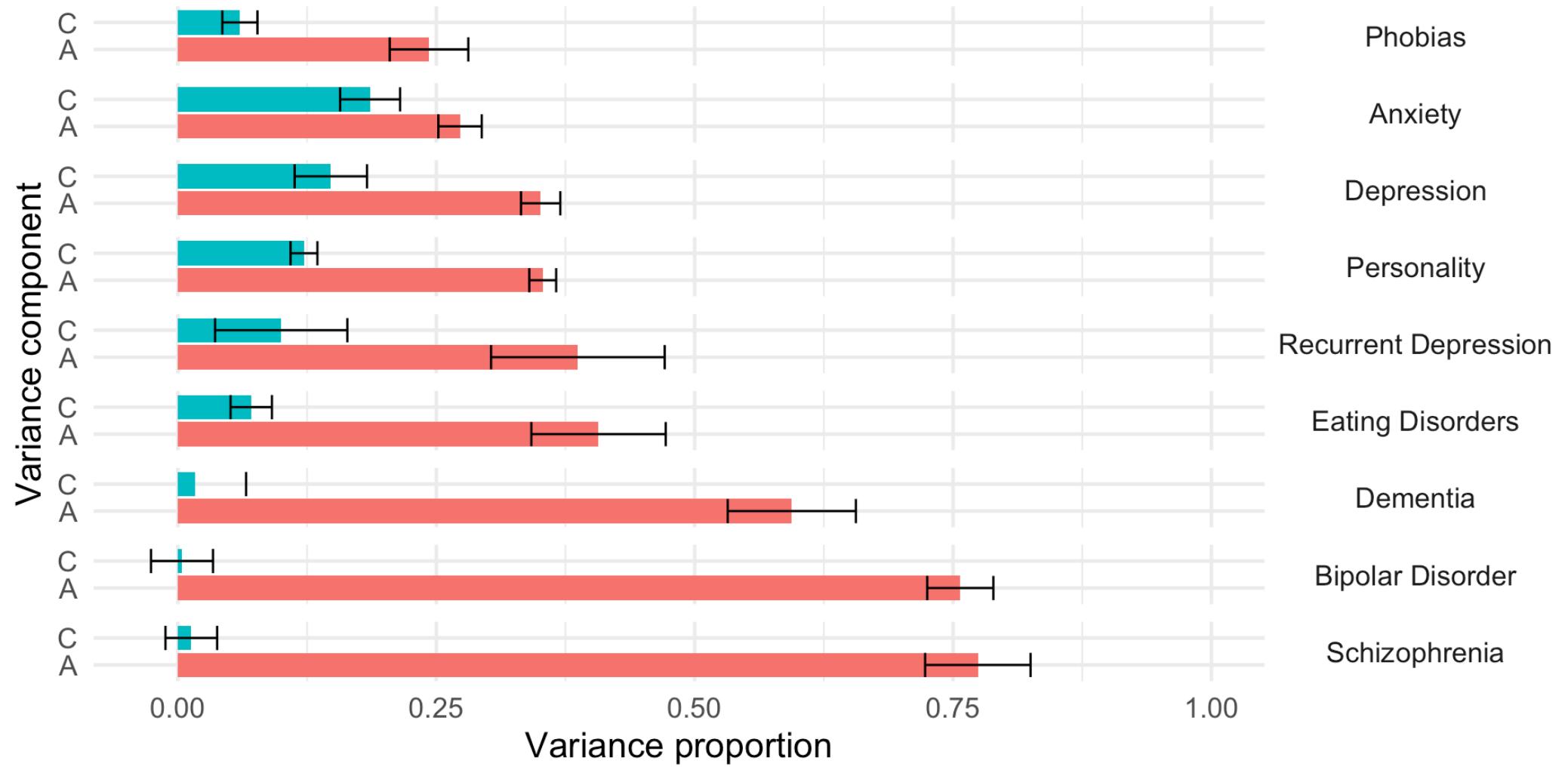
What do we know about psychiatric genetics from twins studies



Meta-analysis of twin heritability



Meta-analysis of phenotype domains



Meta-analysis of psychiatric, neurological, and psychological phenotypes

Genetics of depression and schizophrenia

Major depressive disorder (MDD)

- 1+ in 10 affected
- 7-11 years of life are lost
- Global cause of ill-health

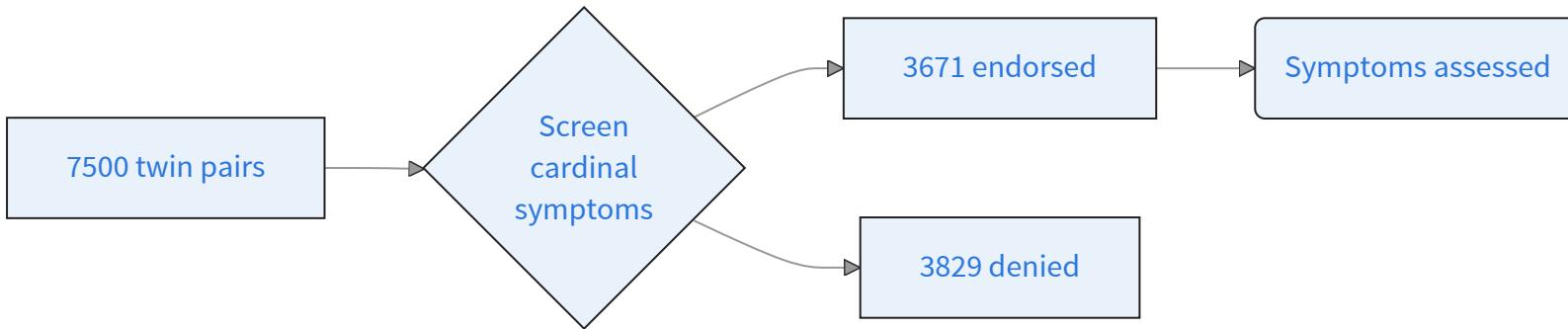
MDD diagnostic criteria

1 of the 2 **cardinal** symptoms with 5 symptoms in total:

1. Low mood
2. Anhedonia
3. Increase or decrease in weight or appetite
4. Insomnia or hypersomnia
5. Psychomotor agitation or slowing
6. Fatigue
7. Feelings of worthlessness/guilt
8. Concentration problems
9. Suicidal ideation

Heterogeneous disorder: 227 possible symptom profiles.

“Evidence for Multiple Genetic Factors Underlying DSM-IV Criteria for Major Depression”



Multivariate twin model

- analysed MZ and DZ twin cross-correlations among depression symptoms
- tested 1, 2, or 3 factor models of genetic, shared environment, and unique environment

Best fit model

Genetic factors

1. Psychomotor/cognitive symptoms (psychomotor, guilt, concentration, suicidality)
2. Mood symptoms (low mood, anhedonia)
3. Neurovegetative symptoms (weight, sleep, fatigue)

Individual-specific (unique) environment factors

1. General depression (low mood, anhedonia, weight, sleep, psychomotor, fatigue, concentration)
2. Mood symptoms (low mood, anhedonia, concentration)
3. Cognitive symptoms (guilt, suicidality)

External validators of genetic factor scores

Data from Table 4 Kendler, Aggen, & Neale. $p < 0.001$

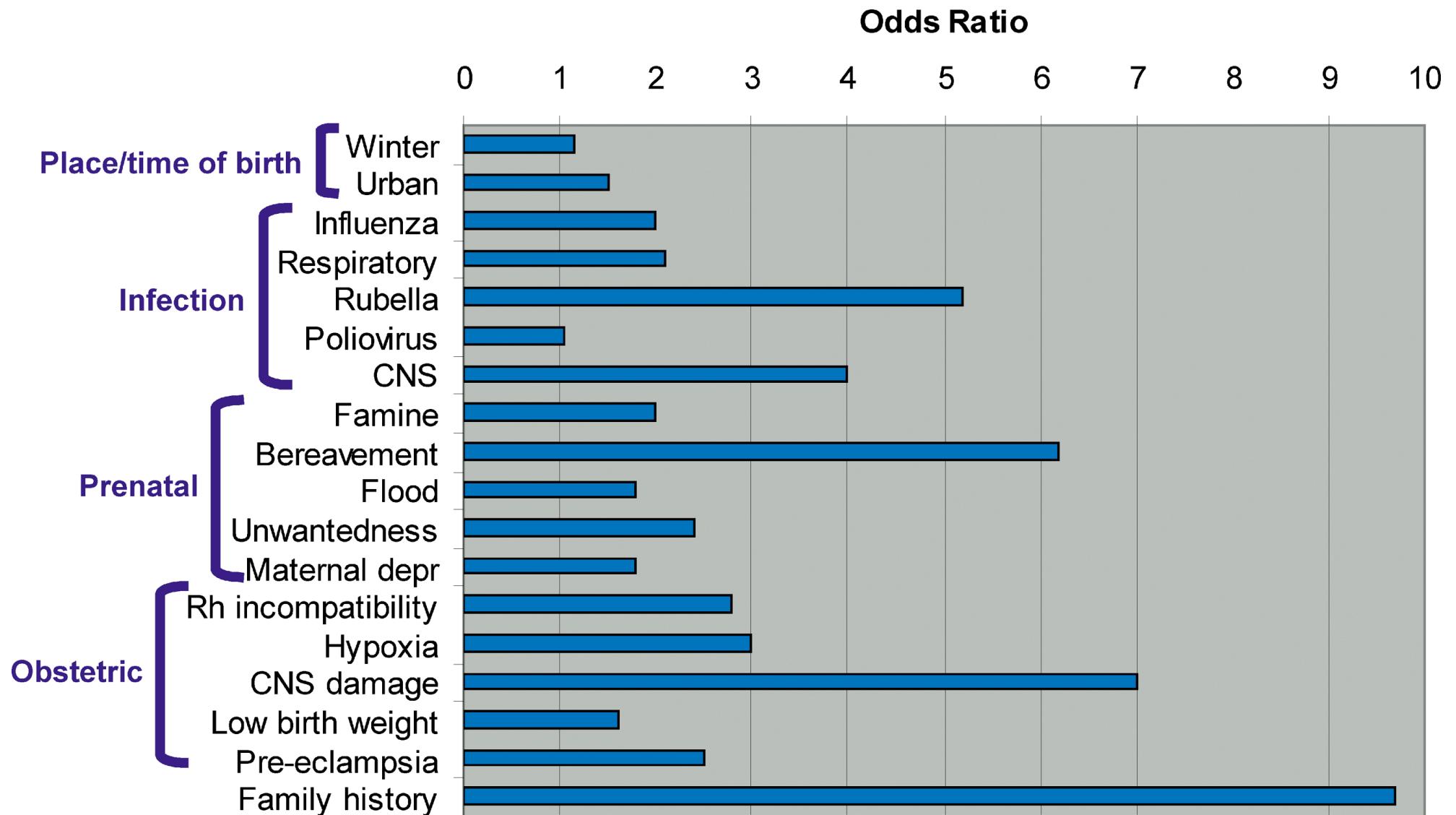
Validator variable	Cognitive/Motor	Core Mood	Neurovegetative
Neuroticism	0.54	0.27	0.30
Anxiety disorder	0.77	0.49	0.47
Age at onset	-3.00	-1.15	0.37
No. of episodes	0.55	0.16	0.23
Melancholia subtype	1.05	-0.43	1.34
Unreactive mood	0.82	0.10	0.79

Schizophrenia diagnostic criteria

At least one **core** symptom and two or more symptoms present for a significant period of time

1. Delusions
2. Hallucinations
3. Disorganised speech
4. Grossly disorganized or catatonic behaviour
5. Negative symptoms

Schizophrenia risk factors



Why is h^2 the symbol for heritability?

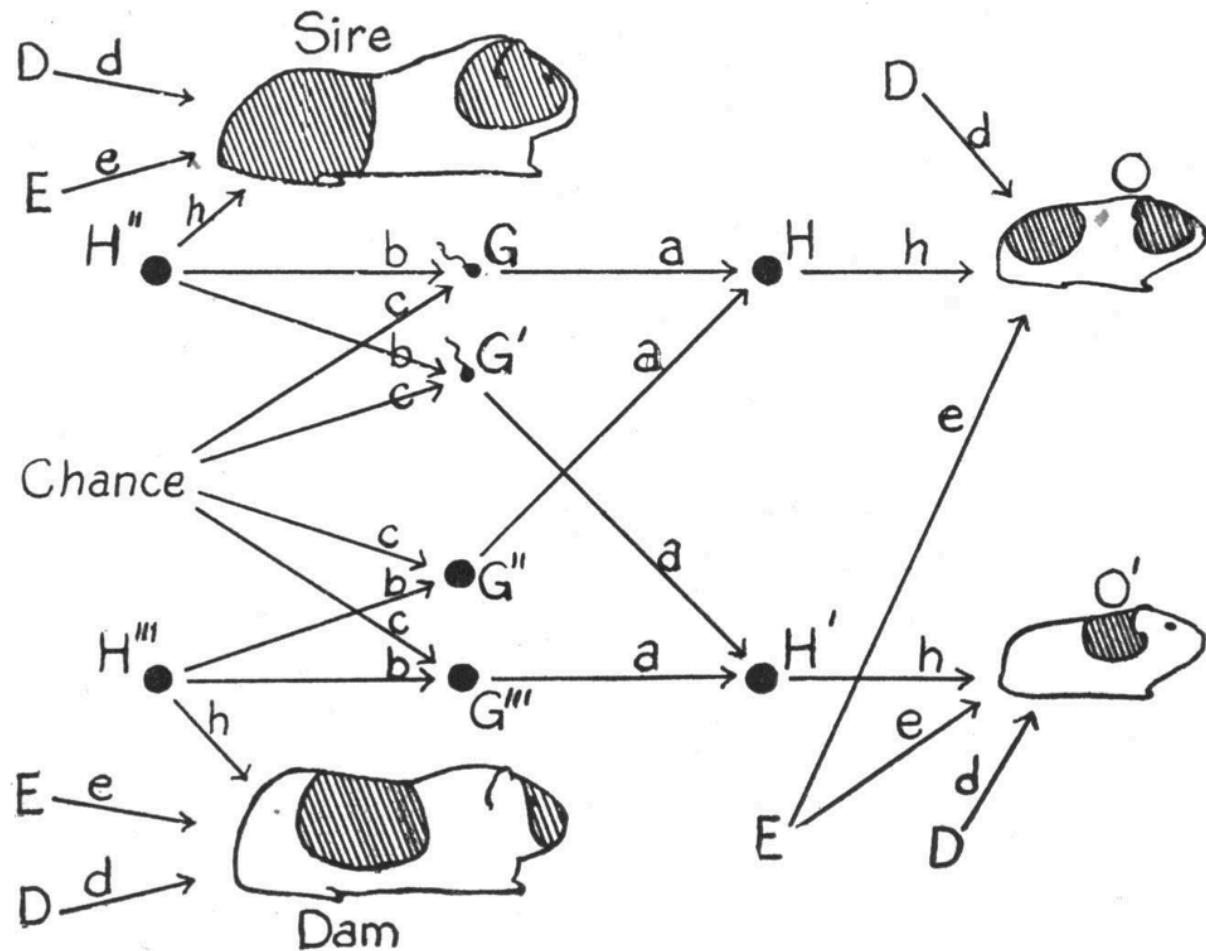


FIG. 5.

Diagram illustrating the causal relations between litter mates (O, O') and between each of them and their parents. H, H', H'', H''' represent the genetic constitutions of the four individuals, $G, G', G'',$ and G''' that of four germ cells. E represents such environmental factors as are common to litter mates. D represents other factors, largely ontogenetic irregularity. The small letters stand for the various path coefficients.