Twins Studies

- Monozygotic (MZ; “identical”)
  - Fertilization of a single egg by a single sperm
  - Share 100% of their genetic material.
- Dizygotic (DZ, “fraternal” or “non-identical”)
  - Result from the independent fertilization of two eggs by two sperm
  - Share on average 50% of their genes (just like full siblings).

A Natural Experiment

- Twins reared apart
  - They did not experience the same environment
  - Gives a much stronger test of genetic and non-shared environmental contributions
  - But separated MZs are rare

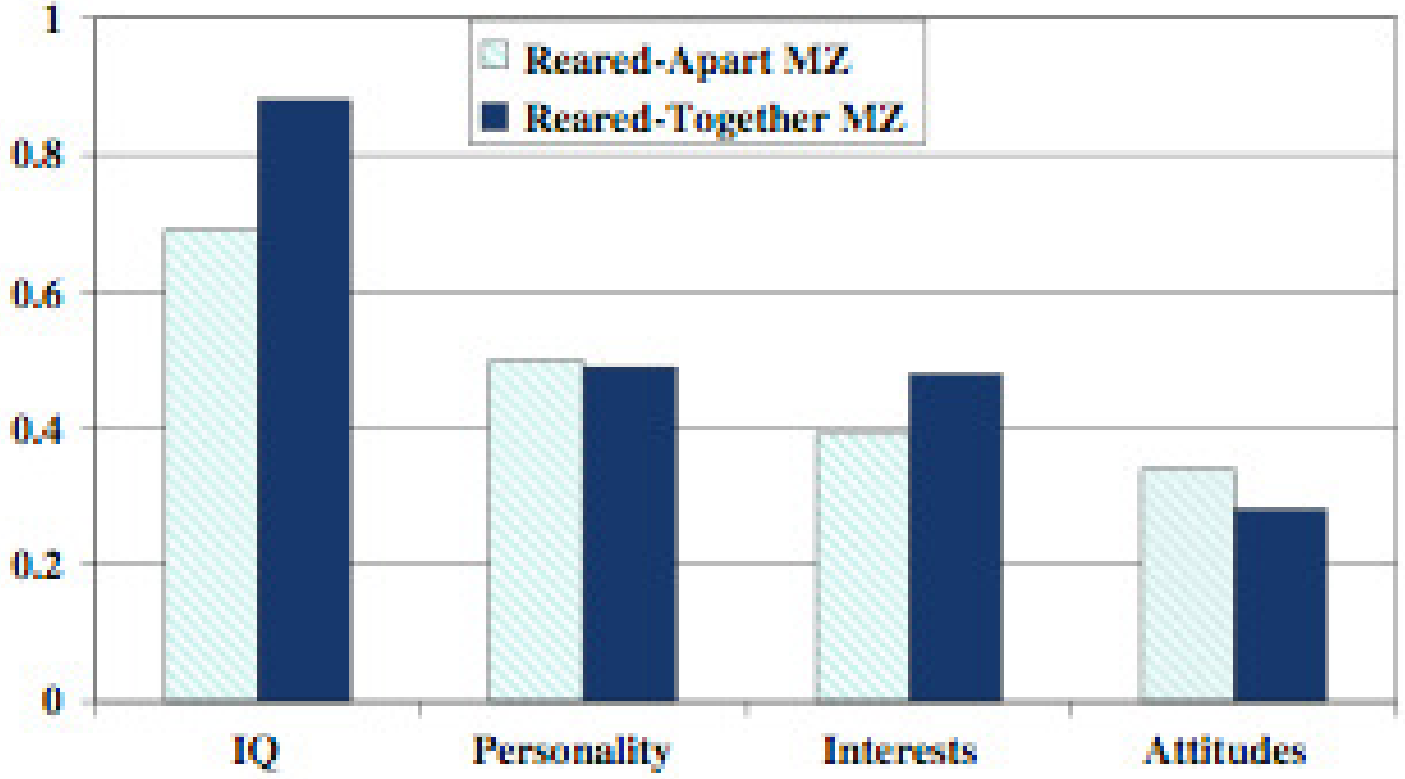


Fig. 2 Average reared-apart and reared-together monozygotic (MZ) twin correlations in four domains of psychological functioning. Adapted from Bouchard et al. (1990)

Variance Components

- Heritability (narrow-sense, A or a^2; broad-sense, H or h^2): phenotypic variance in a sample that can be attributed to genotypic variance.
- Shared or common environment (C or c^2): experiences that makes individuals more similar to one another, regardless of genetic similarity
- Non-shared or Unique environment and Error (E or e^2): What is left over

IDENTICAL TWINS

- MONOZYGOTIC:
- Have IDENTICAL genes (A)
- Come from the same family (C)
- Have unique experiences during life (E)

FRATERNAL TWINS

- DIZYGOTIC: Have DIFFERENT genes (A)
- Come from the same family (C)
- Have unique experiences during life (E)

Falconer's model

Assumes all genetic effects are additive ($h^2 = a^2$)

$$r_{MZ} = a^2 + c^2$$

$$r_{DZ} = 0.5a^2 + c^2$$

$$1.0 = a^2 + c^2 + e^2$$

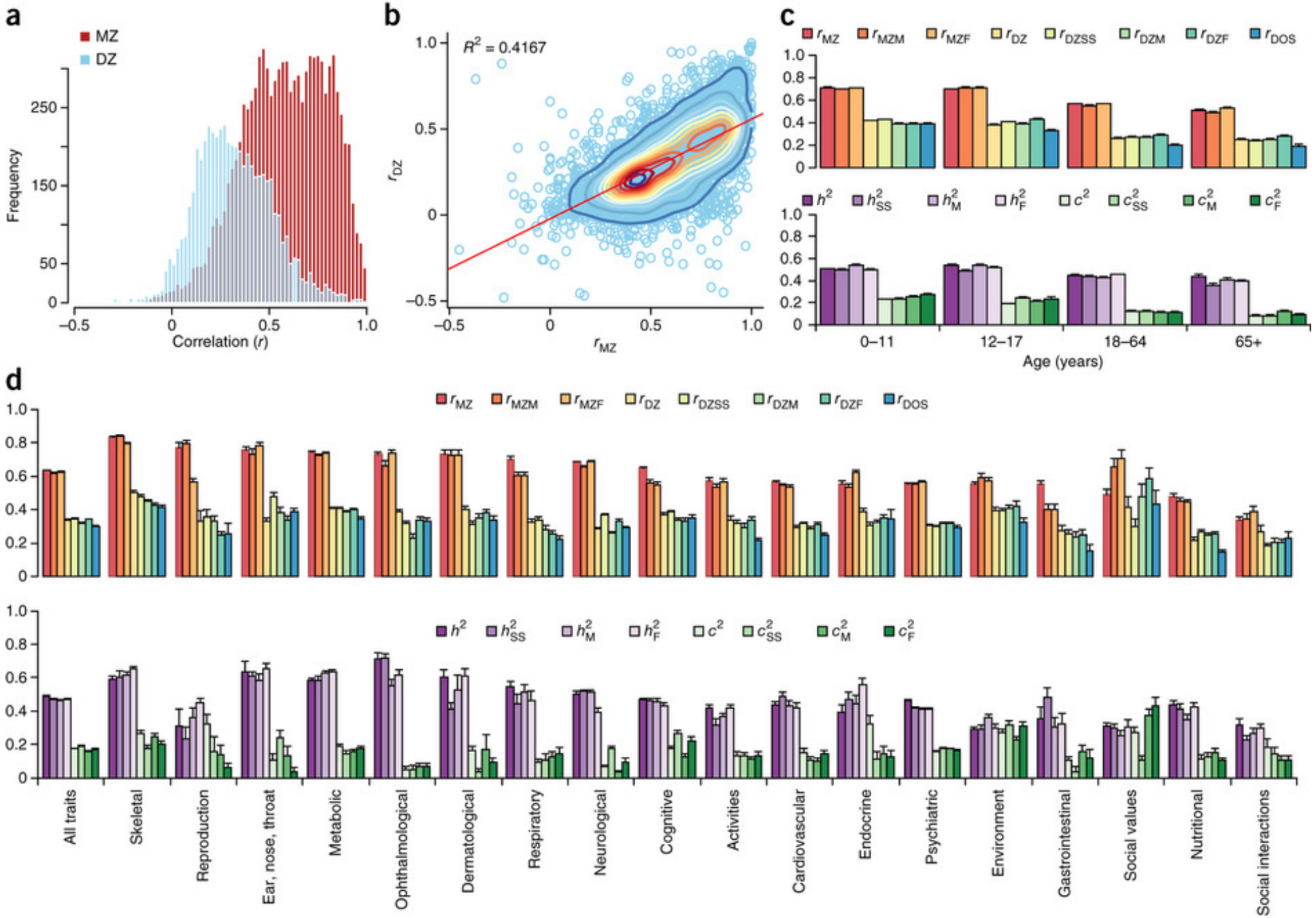
Falconer estimates

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

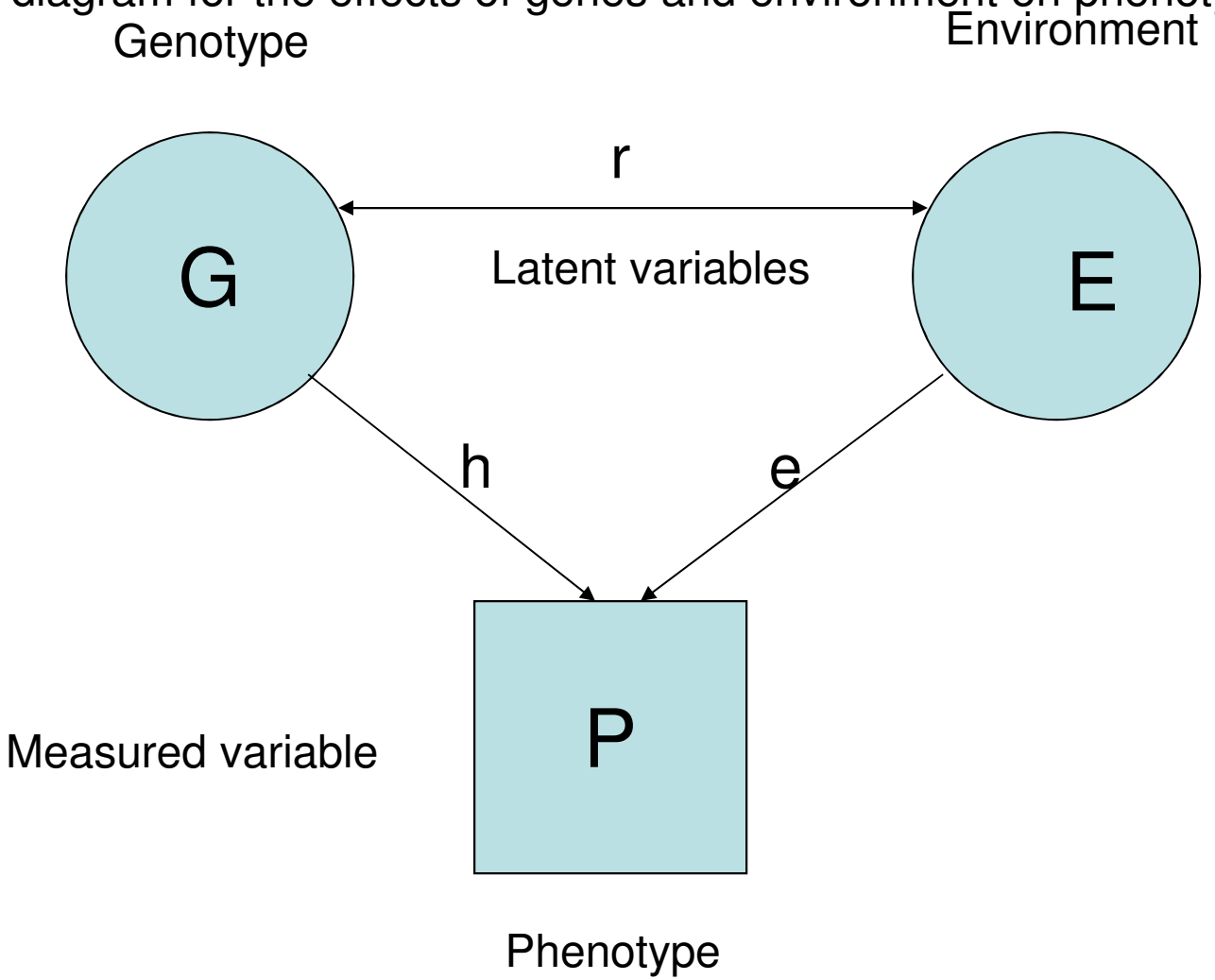
$$c^2 = 2r_{DZ} - r_{MZ}$$

$$e^2 = 1 - r_{MZ}$$

Polderman et al. 2015, *Nature Genetics*



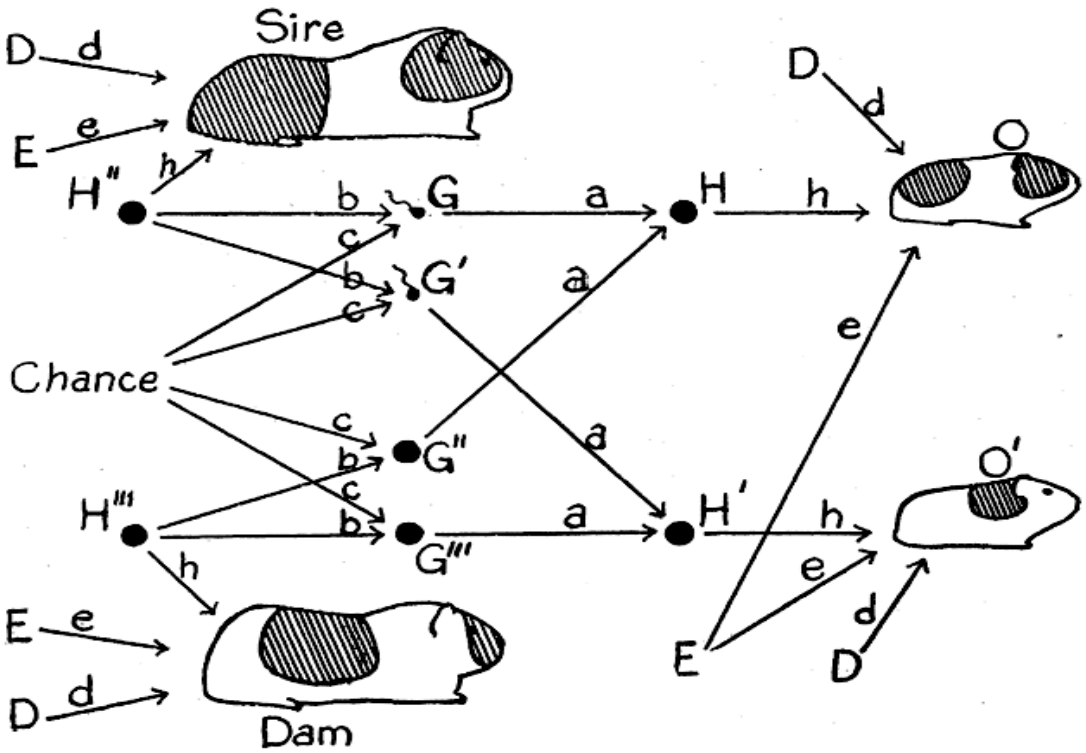
Path diagram for the effects of genes and environment on phenotype



Path Analysis

- Derive predictions for the variances and covariances of the variables under the specified model
  - Present relationships between variables using **diagrams**
  - The relationships can also be represented as **structural equations** and **covariance matrices**
  - Structural equation modelling (SEM) represents a unified platform for path analytic and variance components models

Path Diagram



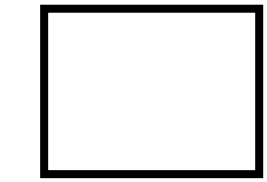
Maximum Likelihood Estimation

- Likelihood: probability that an observation (data point) is predicted by specified model
- For MLE, determine most likely values of population parameter value (e.g, μ, σ, β) given observed sample value
  - define model
  - define probability of observing a given event conditional on a particular set of parameters
  - choose a set of parameters which are most likely to have produced observed results

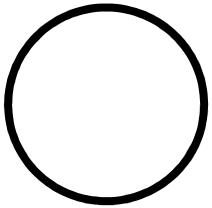
Classical Twin Study Assumptions

- Equal means/variances in Twin 1 and Twin 2
- Equal means/variances in MZ and DZ twins
- Random Mating
- Equal Environments of MZ and DZ pairs
- No GE Correlation
- No G x E Interaction
- No Sex Limitation

Path Diagram Conventions

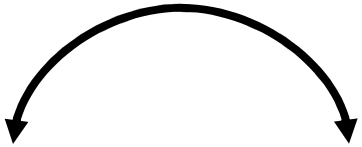


Observed Variable

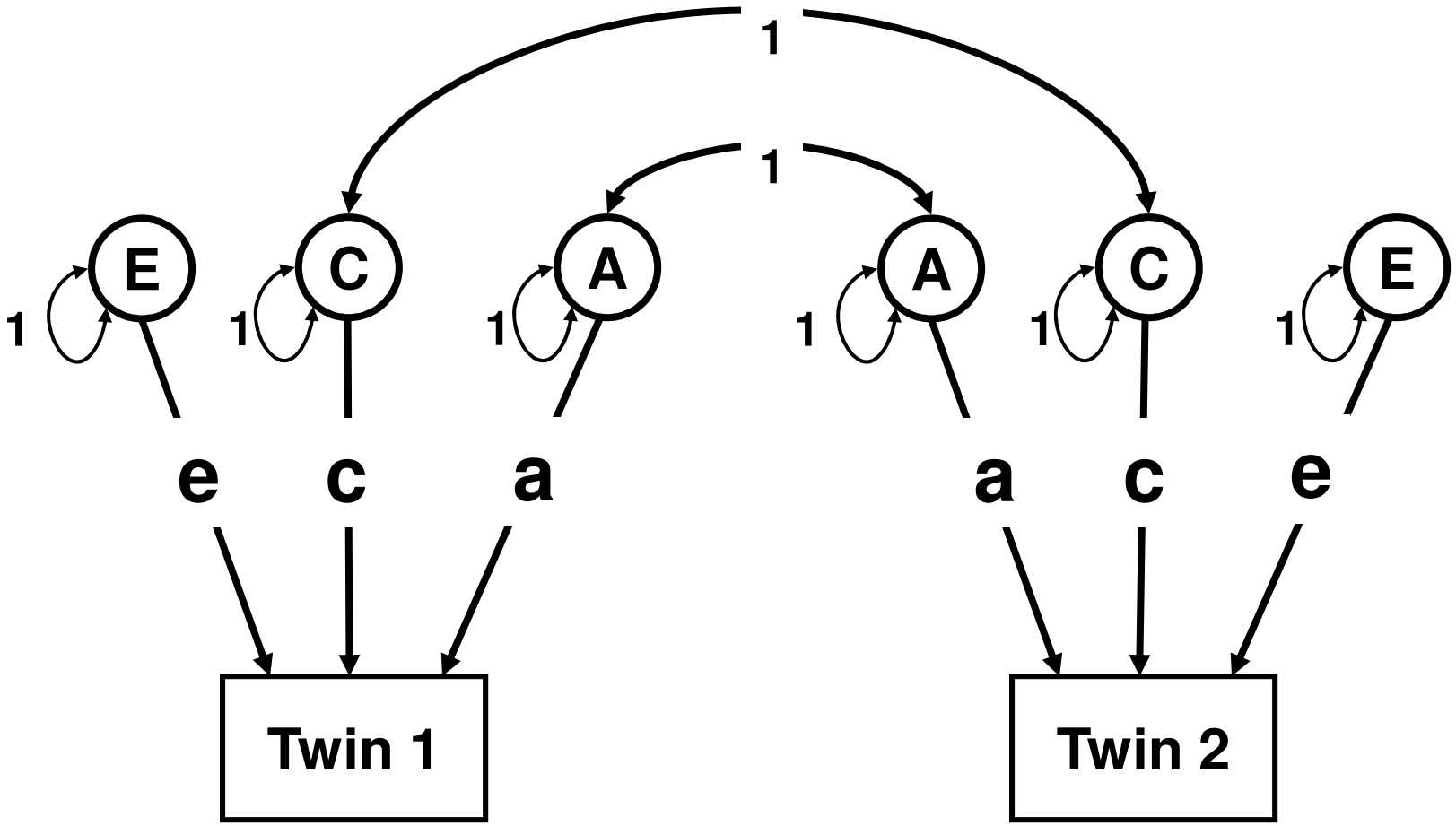


Latent Variable

Causal Path

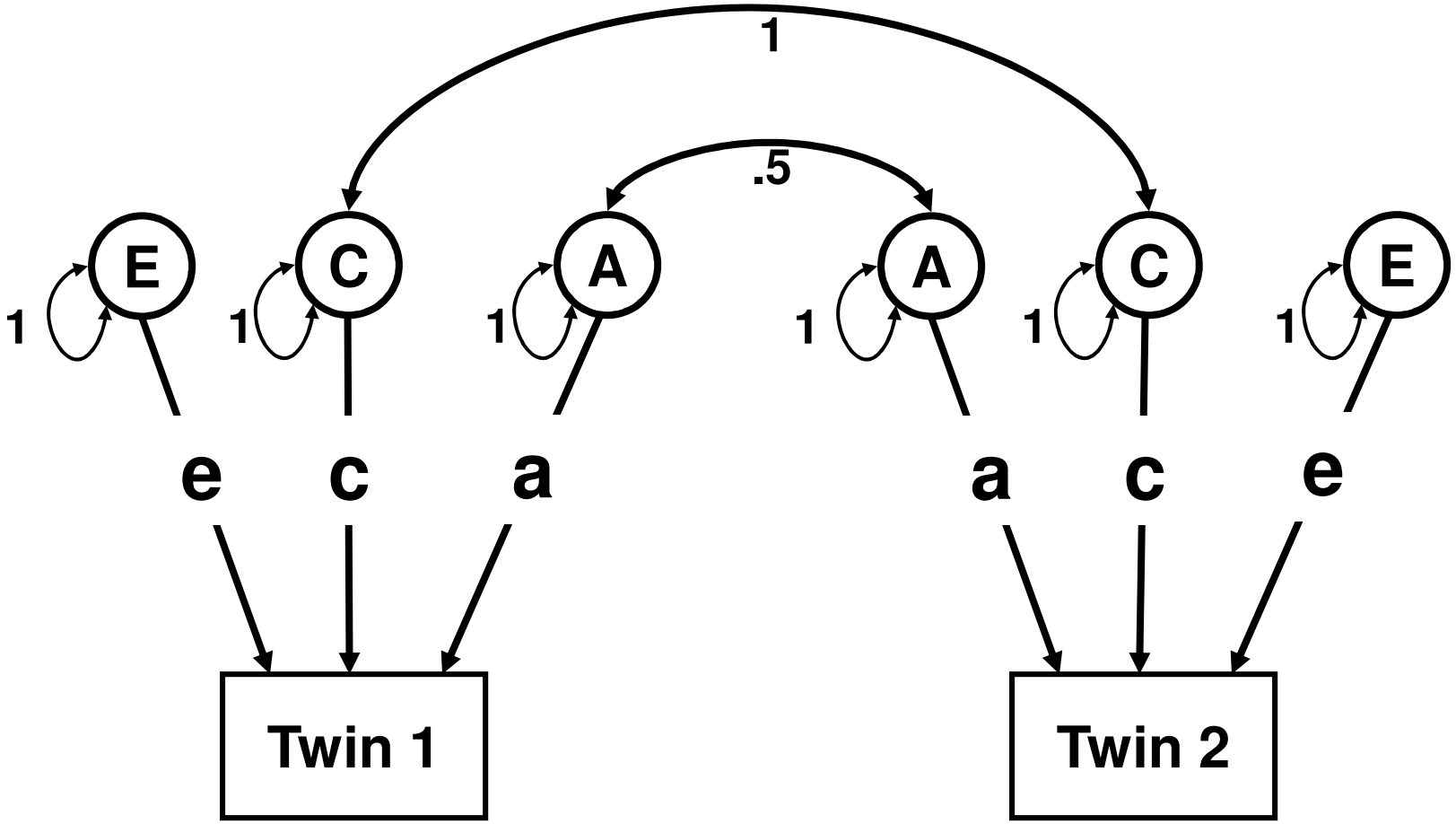


Covariance Path



Model for an MZ PAIR

Note: a, c and e are the same cross twins



Model for a DZ PAIR

Note: a, c and e are also the same cross groups

$$V_P = a^2 + c^2 + e^2$$

What about “dominance”* effects?

$$V_P = a^2 + d^2 + c^2 + e^2$$

* Dominance defined statistically as individuals being more similar on the basis of genetic overlap alone than would be predicted by a linear model – i.e., **any** non-linear genotype “dosage” effects

Twin Correlations → Sources of Variance

rMZ > rDZ	A
rMZ = 2 rDZ	only A (no C,D)
rMZ = rDZ	only C (no A,D)
rMZ < 2 rDZ	A & C
rMZ > 2 rDZ	A & D

ACE **or** ADE

$$\text{Cov(mz)} = a^2 + c^2 \ \textbf{or} \ a^2 + d^2$$

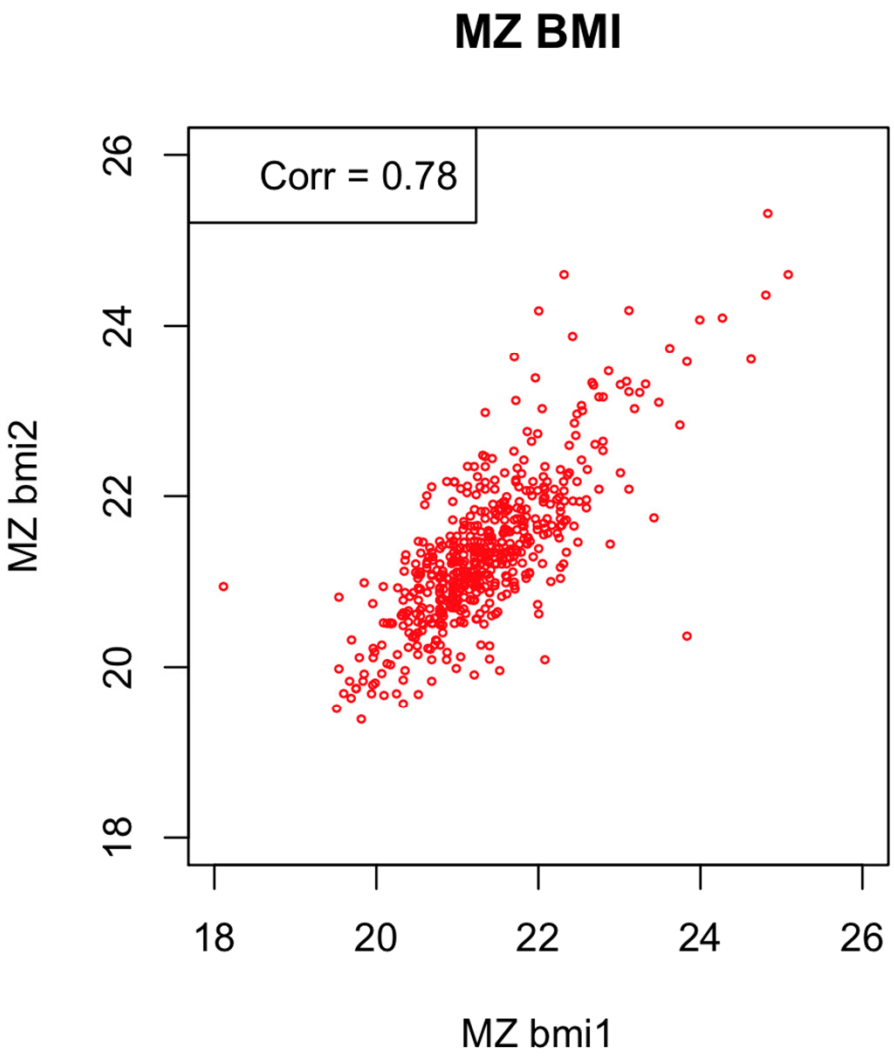
$$\text{Cov(dz)} = \tfrac{1}{2} a^2 + c^2 \ \textbf{or} \ \tfrac{1}{2} a^2 + \tfrac{1}{4} d^2$$

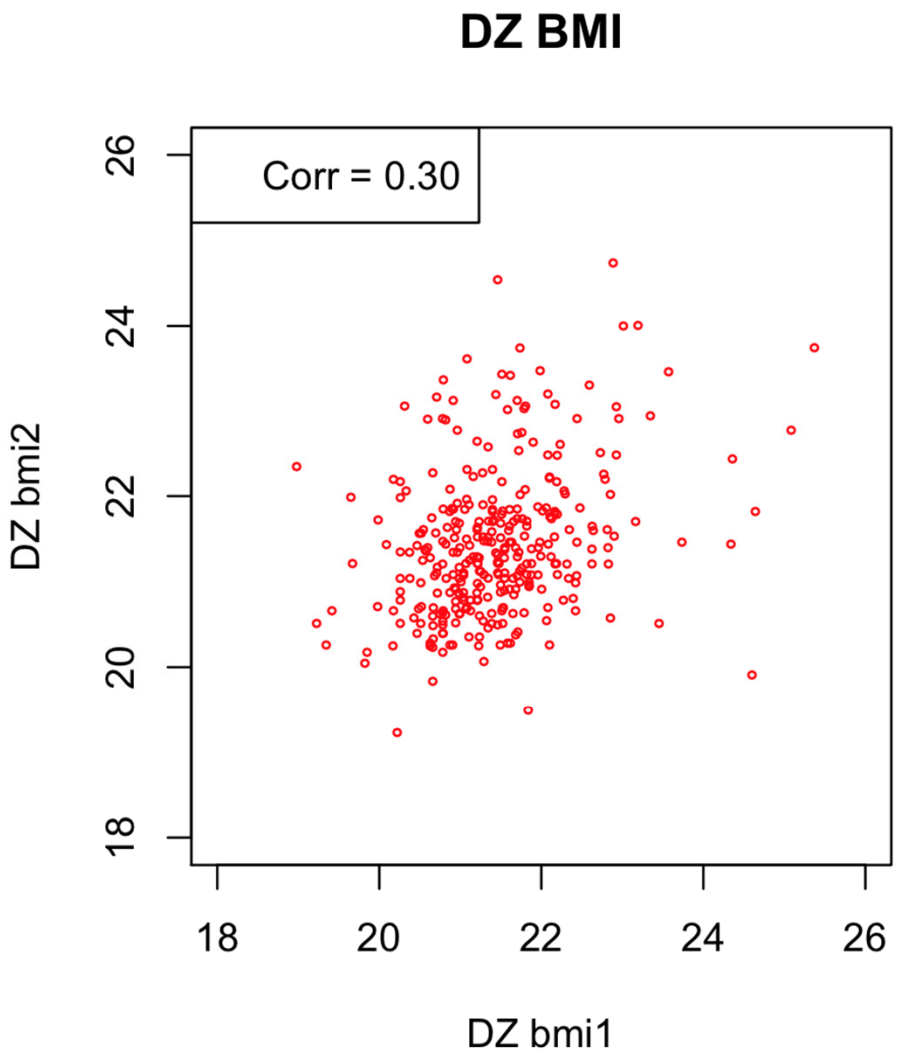
$$V_P = a^2 + c^2 + e^2 \ \textbf{or} \ a^2 + d^2 + e^2$$

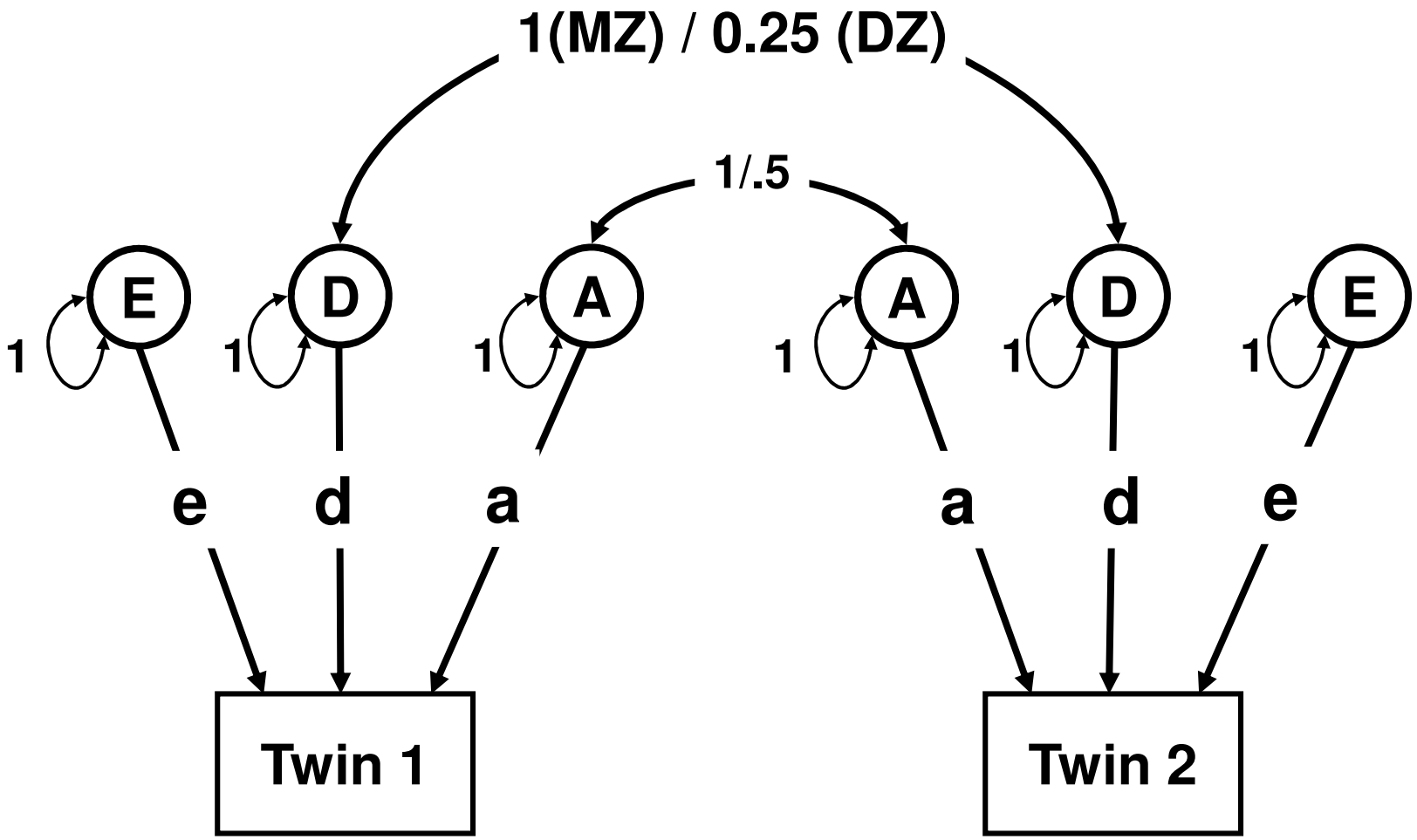
3 unknown parameters (a, c, e or a, d, e), and only **3** distinctive predicted statistics:

Cov MZ, Cov DZ, Vp)

this model is **just identified**


MZ BMI
Corr = 0.78
MZ bmi2
MZ bmi1
18
20
22
24
26


DZ BMI
Corr = 0.30
DZ bmi2
DZ bmi1
18
20
22
24
26

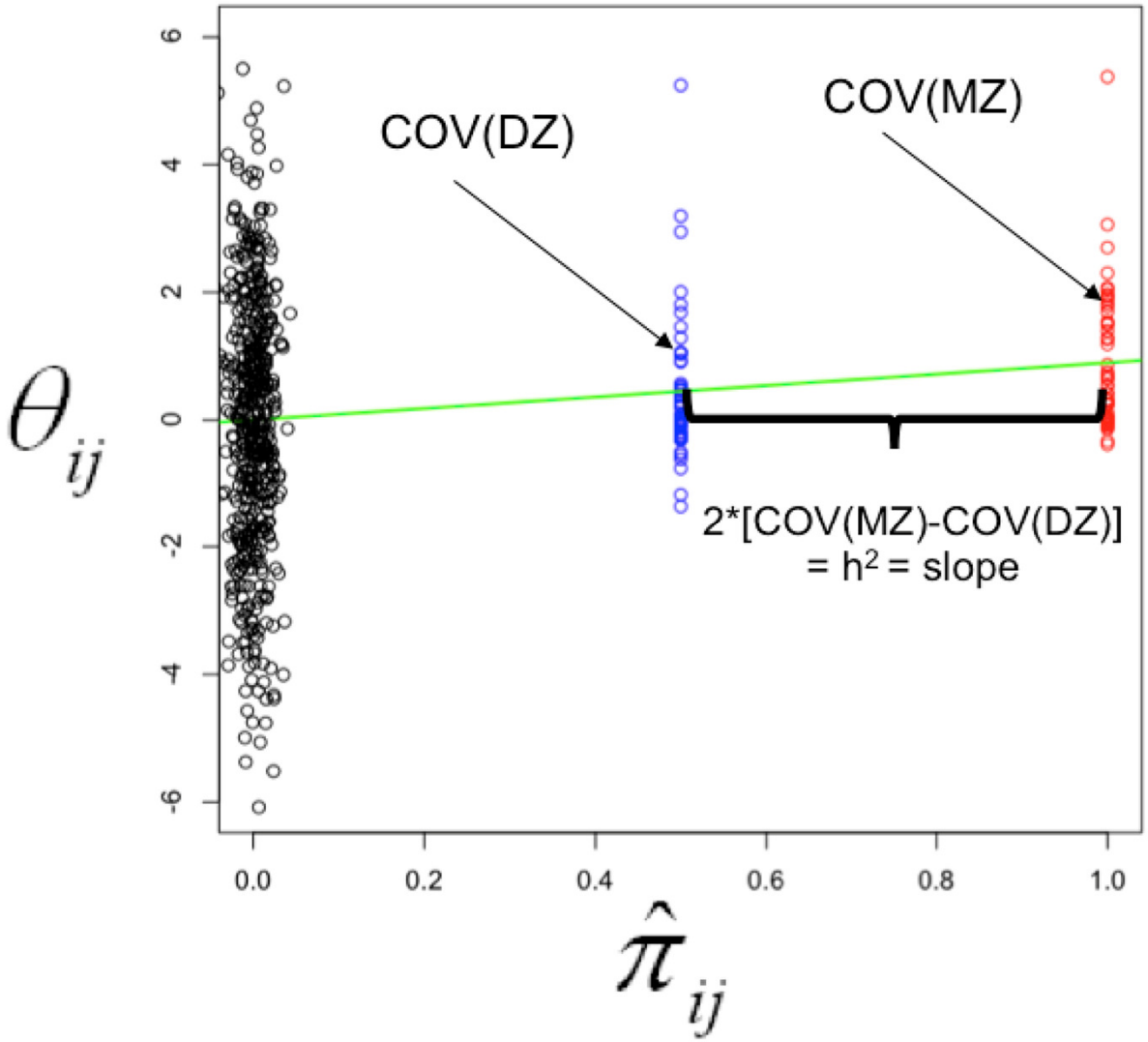

1(MZ) / 0.25 (DZ)
1/.5
E
D
A
A
D
E
1
1
1
1
1
1
e
d
a
a
d
e
Twin 1
Twin 2

Can use the same approach for other relationships

Contributions of V_A and V_D to covariances between relatives (ignoring environment)

Relationship	Contribution to Covariance	
	V_A	V_D
Total variance	1	1
Sibling (DZ twin)	½	¼
MZ twin	1	1
Half-sibling	¼	0
First cousin	$^1/_8$	0
Parent-offspring	½	0
Avuncular	¼	0
Grand-parent	$^1/_8$	0
Unrelated	0 *	0*

Heritability among unrelateds



Heritability among unrelateds

- Estimate of V_A captured by genotyped & imputed variants
  - Upper bound of how much V_A GWAS can detect
- By not using relatives who also share environmental effects:
  - Does not rely on assumption that r(MZ) > r(DZ) for purely genetic reasons

Trait or Disease	h^2 Pedigree Studies	h^2 GWAS Hits[a]	h^2 All GWAS SNPs[b]
Type 1 diabetes	0.9[98]	0.6[99,c]	0.3[12]
Type 2 diabetes	0.3–0.6[100]	0.05-0.10[34]	
Obesity (BMI)	0.4–0.6[101,102]	0.01-0.02[36]	0.2[14]
Crohn's disease	0.6–0.8[103]	0.1[11]	0.4[12]
Ulcerative colitis	0.5[103]	0.05[12]	
Multiple sclerosis	0.3–0.8[104]	0.1[45]	
Ankylosing spondylitis	>0.90[105]	0.2[106]	
Rheumatoid arthritis	0.6[107]		
Schizophrenia	0.7–0.8[108]	0.01[79]	0.3[109]
Bipolar disorder	0.6–0.7[108]	0.02[79]	0.4[12]
Breast cancer	0.3[110]	0.08[111]	
Von Willebrand factor	0.66–0.75[112,113]	0.13[114]	0.25[14]
Height	0.8[115,116]	0.1[13]	0.5[13,14]
Bone mineral density	0.6-0.8[117]	0.05[118]	
QT interval	0.37–0.60[119,120]	0.07[121]	0.2[14]
HDL cholesterol	0.5[122]	0.1[57]	
Platelet count	0.8[123]	0.05–0.1[58]	

Variance explained by common SNPs for complex traits

Visscher PM, et.al. (2012) Am J Hum Genetics