The Augmented Classical Twin Design

(Incorporating Genome-wide Identity by Descent Sharing into Twin Studies in Order to Model Violation of the Equal Environments Assumption)

David Evans[1,2,3]

1 Institute for Molecular Bioscience, University of Queensland

2 University of Queensland Diamantina Institute

3 MRC Integrative Epidemiology Unit, University of Bristol

The Classical Twin Design and the EEA

- The Classical Twin Design compares trait similarity between MZ twins and the similarity between DZ twins to estimate heritability

- Equal Environments Assumption (EEA) posits that trait relevant environmental covariation in MZ pairs is the same as that found within DZ pairs. Violation of this assumption will inflate heritability estimates.
    - MZ twins do experience more similar environments than DZ twins
    - The environmental factor must affect the trait
    - Twins must be "passive" recipients of more similar environments for violation of EEA to affect heritability estimates

- A possible reason for "missing heritability"?

Maher (2008) *Nature*

Methods for testing the EEA

- Measure environmental similarity between twins and examine whether this correlates with trait similarity
  - But need to measure the relevant environmental factor

- Use mistaken zygosity diagnosis to test the EEA. If parents treat MZ twins more similarly than DZ twins based on the preconceived notion that MZ twins are more similar than DZ twins, then trait similarity should be a function of perceived zygosity (tests all environmental factors simultaneously).
  - Has the advantage of not requiring the measurement of specific environmental factors
  - Requires mistaken zygosity diagnosis

OPEN ACCESS Freely available online PLoS GENETICS

Assumption-Free Estimation of Heritability from Genome-Wide Identity-by-Descent Sharing between Full Siblings

Peter M. Visscher*, Sarah E. Medland, Manuel A. R. Ferreira, Katherine I. Morley, Gu Zhu, Belinda K. Cornes, Grant W. Montgomery, Nicholas G. Martin

Genetic Epidemiology Group, Queensland Institute of Medical Research, Brisbane, Australia

The study of continuously varying, quantitative traits is important in evolutionary biology, agriculture, and medicine. Variation in such traits is attributable to many, possibly interacting, genes whose expression may be sensitive to the environment, which makes their dissection into underlying causative factors difficult. An important population parameter for quantitative traits is heritability, the proportion of total variance that is due to genetic factors. Response to artificial and natural selection and the degree of resemblance between relatives are all a function of this parameter. Following the classic paper by R. A. Fisher in 1918, the estimation of additive and dominance genetic variance and heritability in populations is based upon the *expected* proportion of genes shared between different types of relatives, and explicit, often controversial and untestable models of genetic and non-genetic causes of family resemblance. With genome-wide coverage of genetic markers it is now possible to estimate such parameters solely within families using the *actual* degree of identity-by-descent sharing between relatives. Using genome scans on 4,401 quasi-independent sib pairs of which 3,375 pairs had phenotypes, we estimated the heritability of height from empirical genome-wide identity-by-descent sharing, which varied from 0.374 to 0.617 (mean 0.498, standard deviation 0.036). The variance in identity-by-descent sharing per chromosome and per genome was consistent with theory. The maximum likelihood estimate of the heritability for height was 0.80 with no evidence for non-genetic causes of sib resemblance, consistent with results from independent twin and family studies but using an entirely separate source of information. Our application shows that it is feasible to estimate genetic variance solely from within-family segregation and provides an independent validation of previously untestable assumptions. Given sufficient data, our new paradigm will allow the estimation of genetic variation for disease susceptibility and quantitative traits that is free from confounding with non-genetic factors and will allow partitioning of genetic variation into additive and non-additive components.

Citation: Visscher PM, Medland SE, Ferreira MAR, Morley KI, Zhu G, et al. (2006) Assumption-free estimation of heritability from genome-wide identity-by-descent sharing between full siblings. PLoS Genet 2(3): e41.

Introduction

The theoretical basis for the resemblance between relatives due to genetic factors was developed by R.A. Fisher in a now famous and classic paper that reconciled Mendelian and biometrical genetics [1]. Following that theoretical basis, quantitative genetic parameters are estimated from the resemblance between different types of relatives by equating the observed phenotypic covariance to the degree of genetic relationship, which is estimated from pedigree data. The degree of relationship is usually expressed as the coefficient of kinship [2] or the additive coefficient of relationship [2,3]. In a non-inbred population, the coefficient of relationship is the expected proportion of alleles identical-by-descent (IBD) between relatives and determines the additive genetic covariance between a pair of relatives. Maximum likelihood (ML) methods and software have been developed to estimate genetic (co)variances in simple [4] and large complex pedigrees [5–7], for univariate and multivariate models. What all these methods have in common is that they estimate genetic parameters from observed variation between and within families, assuming an underlying model for causative components of variance [3]. For example, in twin studies it is commonly assumed that the variance between families is due to common environmental and additive genetic effects, and that the variance within families reflects individual environmental effects (for monozygotic [MZ] pairs) or both individual environmental and additive genetic effects (for dizygotic [DZ] pairs).

In human populations, the interplay of genetic, environmental, and cultural factors that cause family resemblance is complex; and crucially, the ultimate separation of nature and nurture effects can generally not be tested empirically

Editor: Trudy Mackay, North Carolina State University, United States of America

Received December 7, 2005; **Accepted** February 6, 2006; **Published** March 24, 2006

A previous version of this article appeared as an Early Online Release on February 6, 2006 (DOI: 10.1371/journal.pgen.0020041.eor).

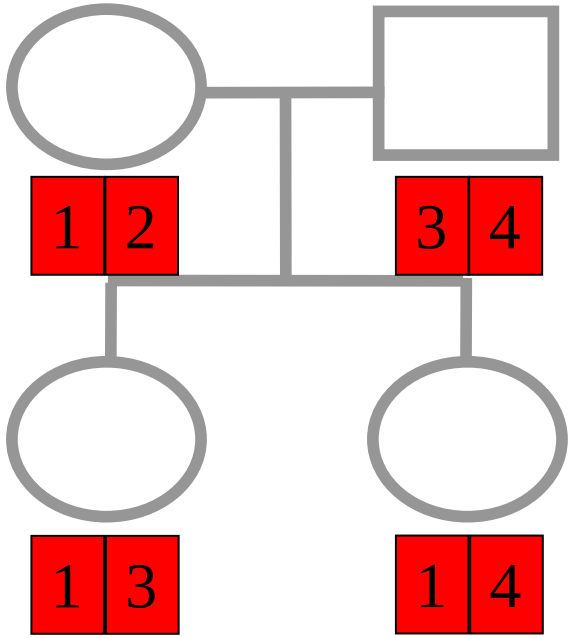
DOI: 10.1371/journal.pgen.0020041

Abbreviations: A, genome-wide additive effect; CI, confidence interval; DZ, dizygotic; E, residual error effect; F, non-genetic family effect; FAE, full model; FE, reduced model, containing F and E effects only; IBD, identical-by-descent; IBD2, identical-by-descent at two alleles; ML, maximum likelihood; MZ, monozygotic; SD, standard deviation; SE, standard error; QTL, quantitative trait loci

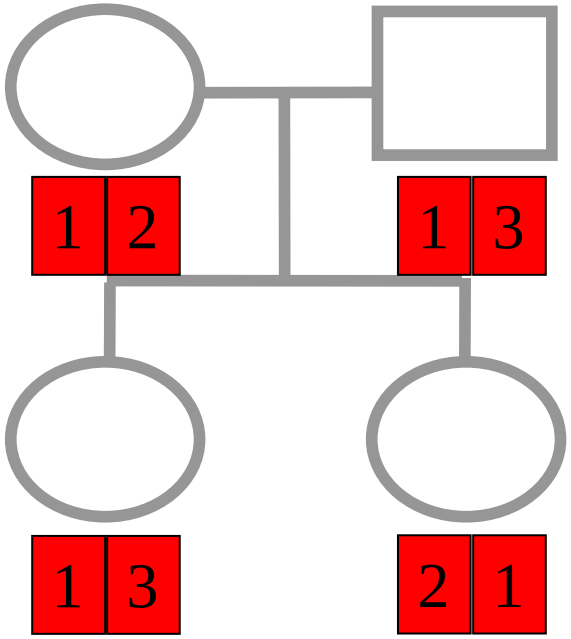
* To whom correspondence should be addressed. E-mail: peter.visscher@qimr.edu.au

- Phenotypic similarity between full siblings measured as a function of their genomic similarity
- Enables the estimation of trait heritability without many of the assumptions of the e.g. classical twin design
- Requires large numbers of sibling pairs to estimate heritability precisely
- Could this information be used in the Classical Twin Design to investigate violations of the EEA?

IBD vs IBS



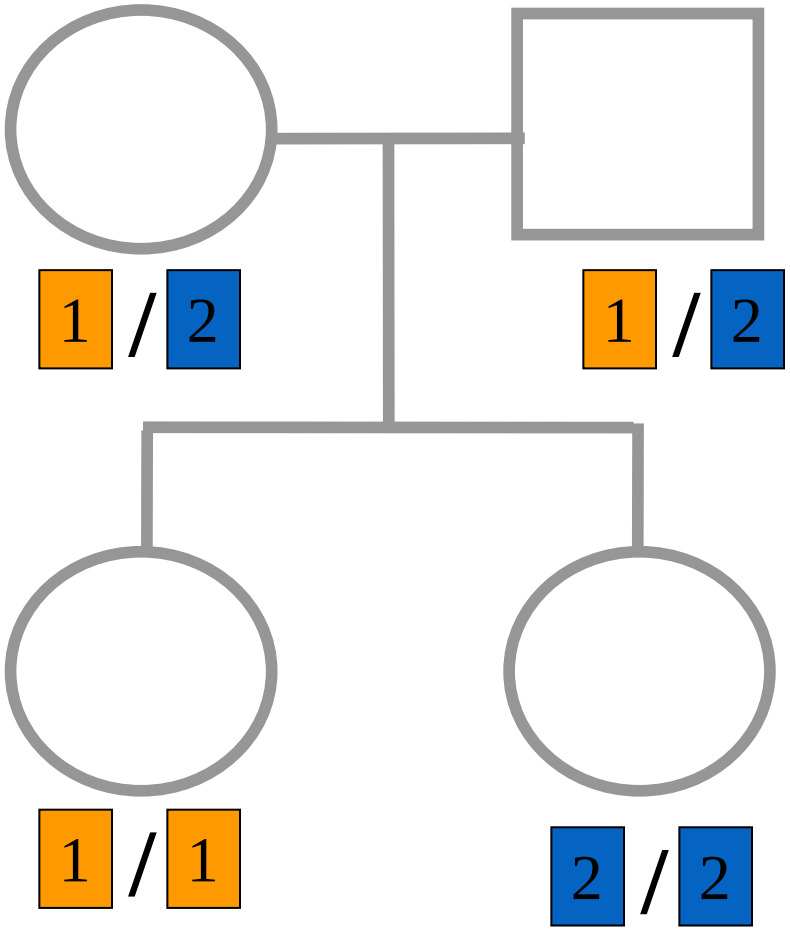
Identical by Descent



Identical by state only

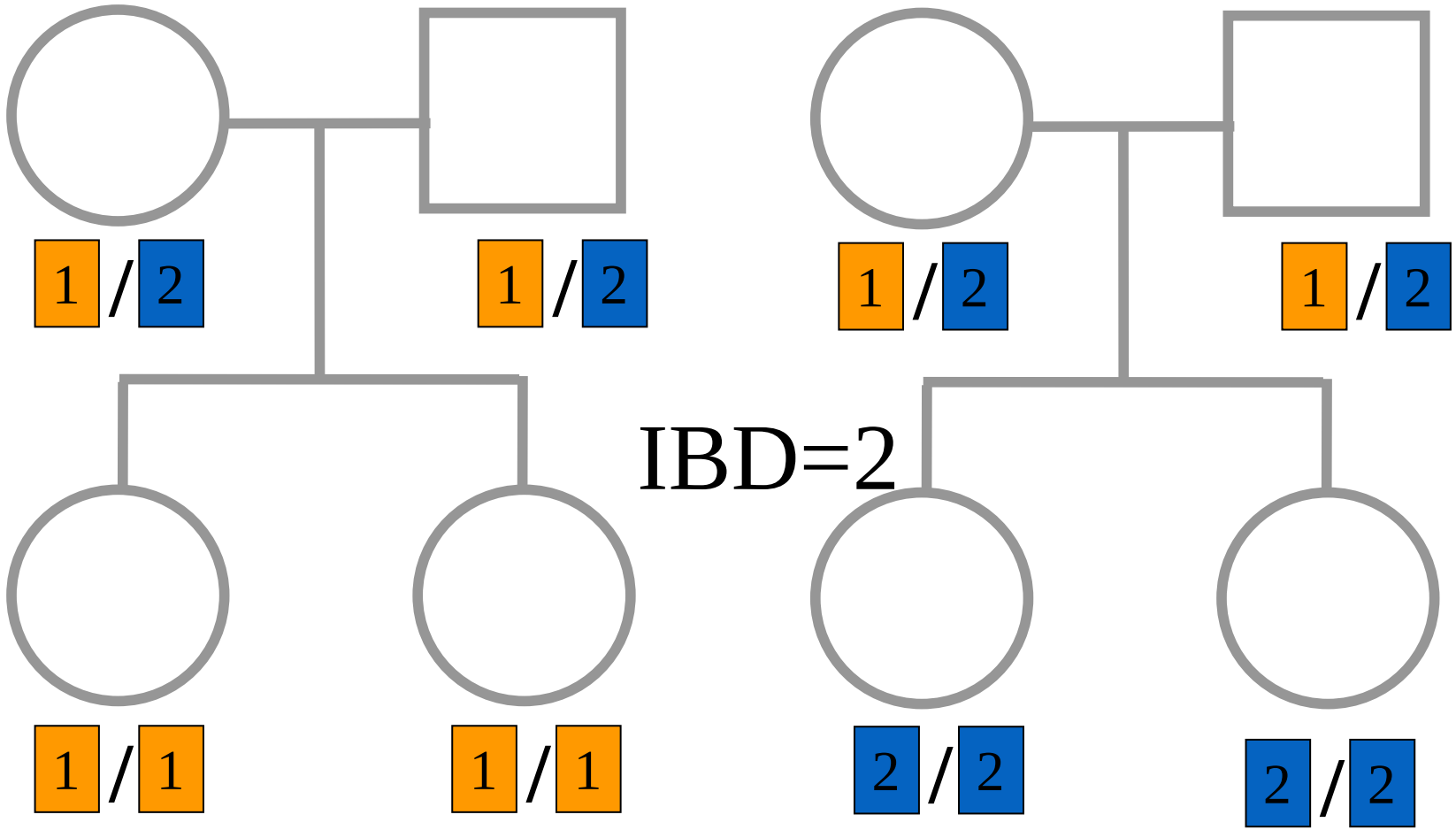
Two alleles are IBD if they are descended from the same ancestral allele

IBD can be trivial…

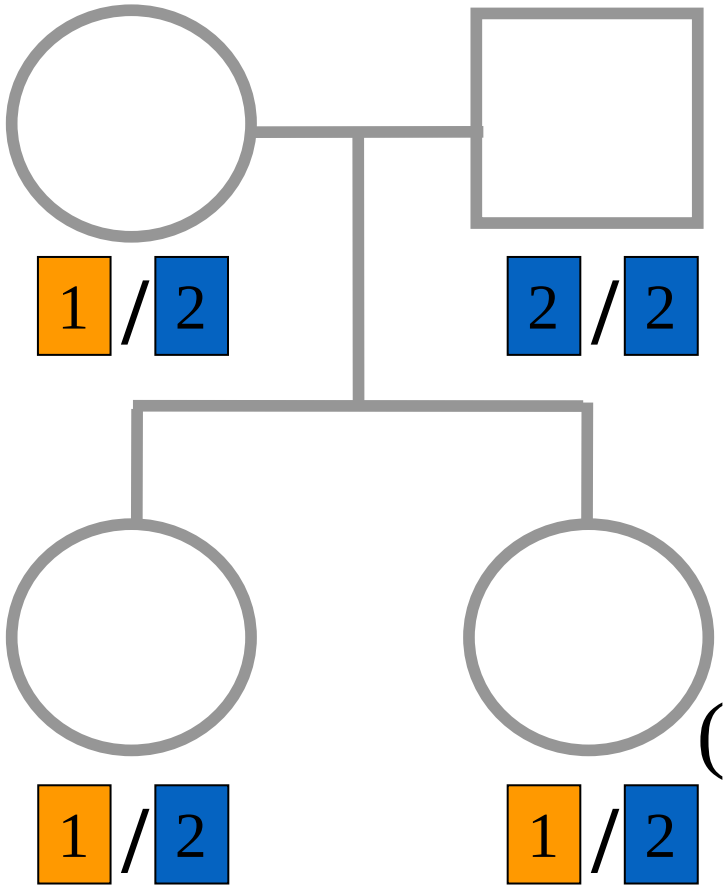


IBD=0

Two Other Simple Cases…



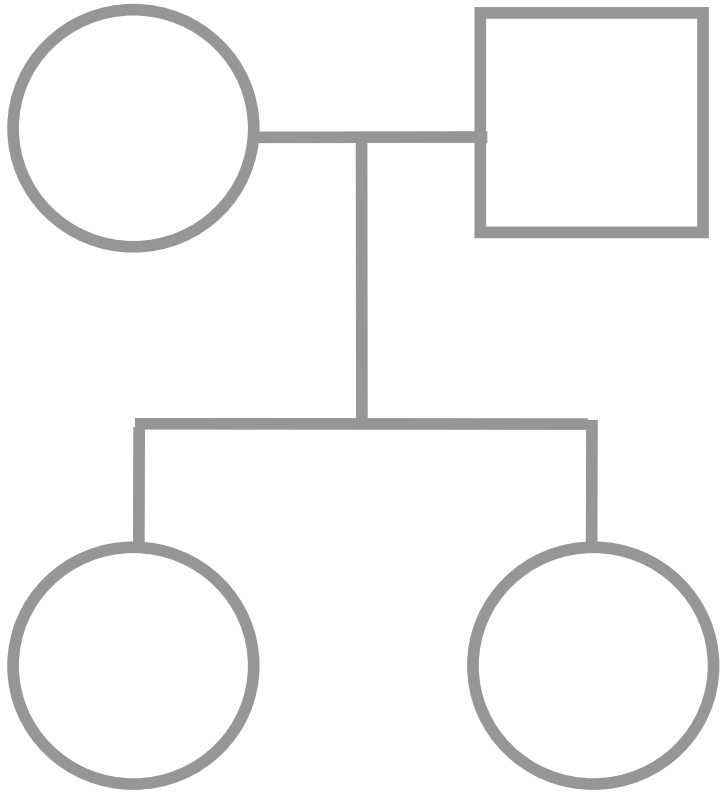
A little more complicated…



IBD=1
(50% chance)

IBD=2
(50% chance)

And even more complicated…



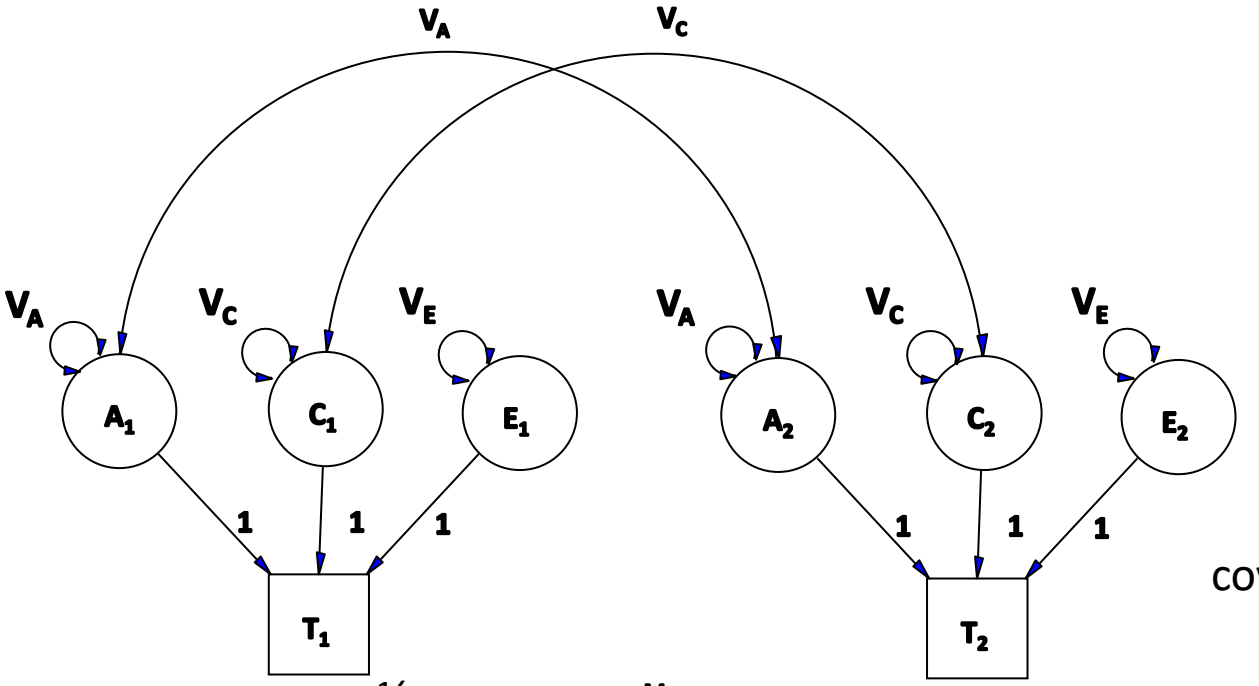
IBD=?

1 / 1

1 / 1

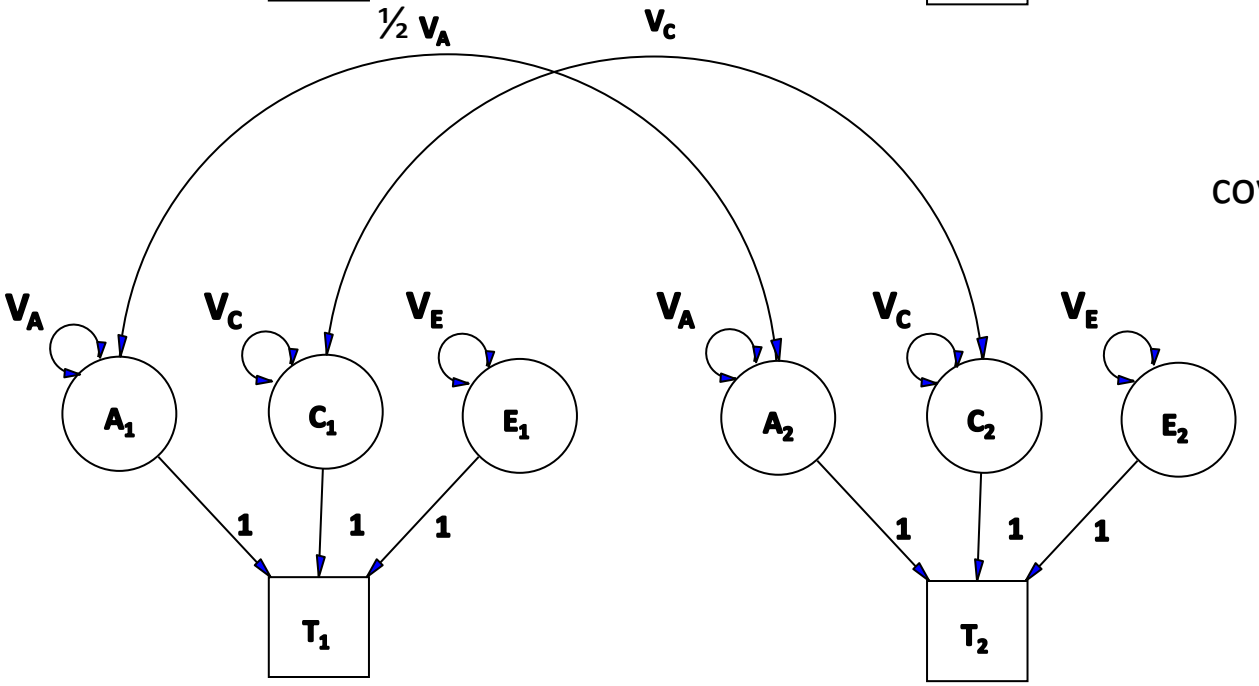
Path Model for the Classical Twin Design

MZ Twins:



$$cov_{MZ} = \begin{bmatrix} V_A + V_C + V_E & V_A + V_C \\ V_A + V_C & V_A + V_C + V_E \end{bmatrix}$$

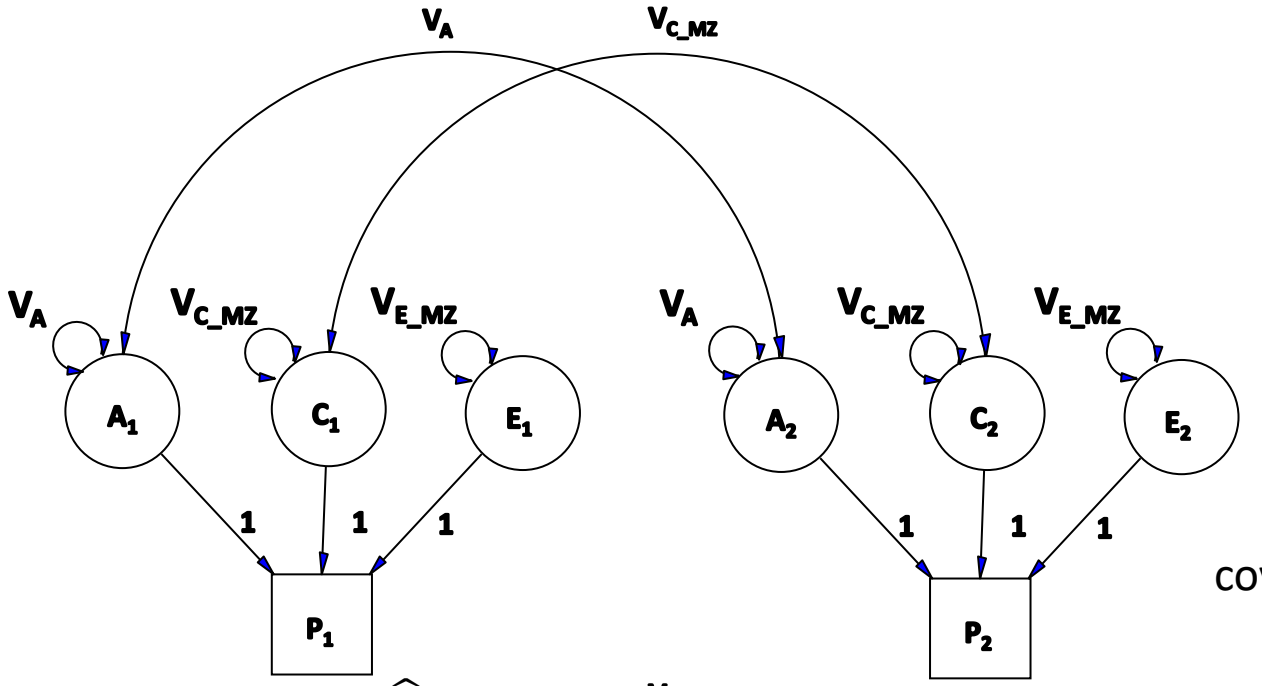
DZ Twins:



$$cov_{DZ} = \begin{bmatrix} V_A + V_C + V_E & \frac{1}{2} V_A + V_C \\ \frac{1}{2} V_A + V_C & V_A + V_C + V_E \end{bmatrix}$$

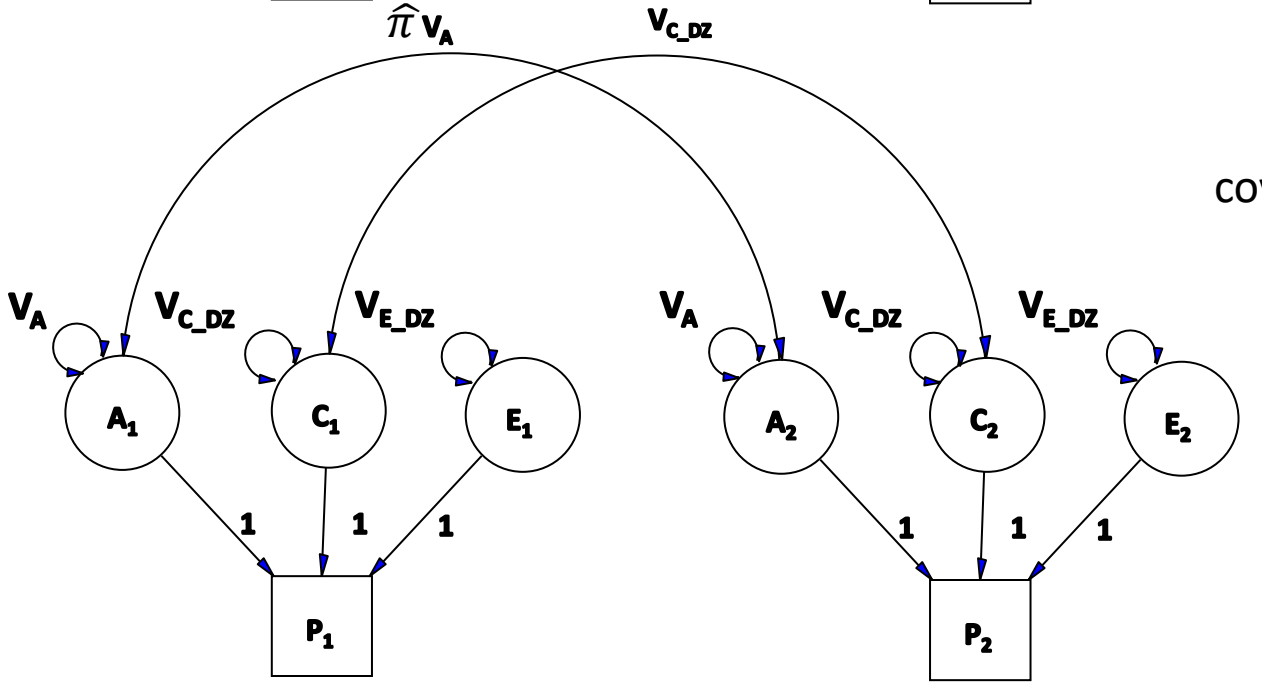
Path Model for the Augmented Classical Twin Design

MZ Twins:



$$cov_{MZ} = \begin{bmatrix} V_A + V_{C\_MZ} + V_{E\_MZ} & V_A + V_{C\_MZ} \\ V_A + V_{C\_MZ} & V_A + V_{C\_MZ} + V_{E\_MZ} \end{bmatrix}$$

DZ Twins:



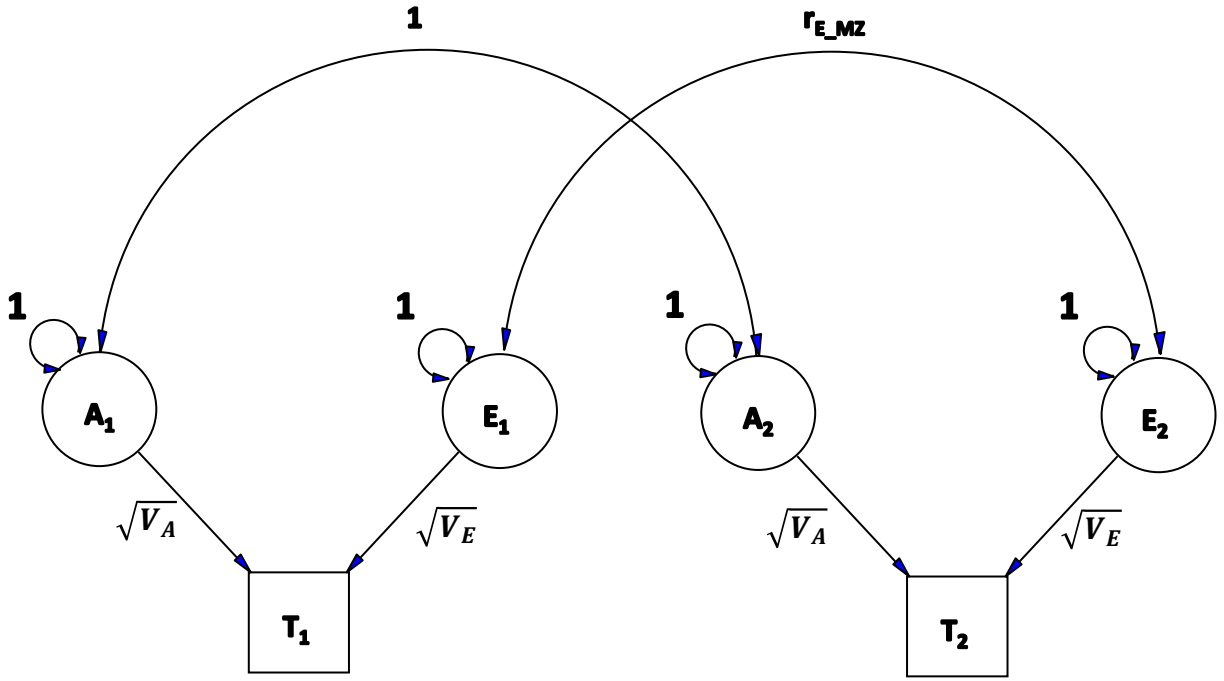
$$cov_{DZ} = \begin{bmatrix} V_A + V_{C\_DZ} + V_{E\_DZ} & \hat{\pi}\, V_A + V_{C\_DZ} \\ \hat{\pi}\, V_A + V_{C\_DZ} & V_A + V_{C\_DZ} + V_{E\_DZ} \end{bmatrix}$$

*Constraint: $V_{C\_MZ} + V_{E\_MZ} = V_{C\_DZ} + V_{E\_DZ}$

Test of EEA is whether $V_{C\_MZ} = V_{C\_DZ}$

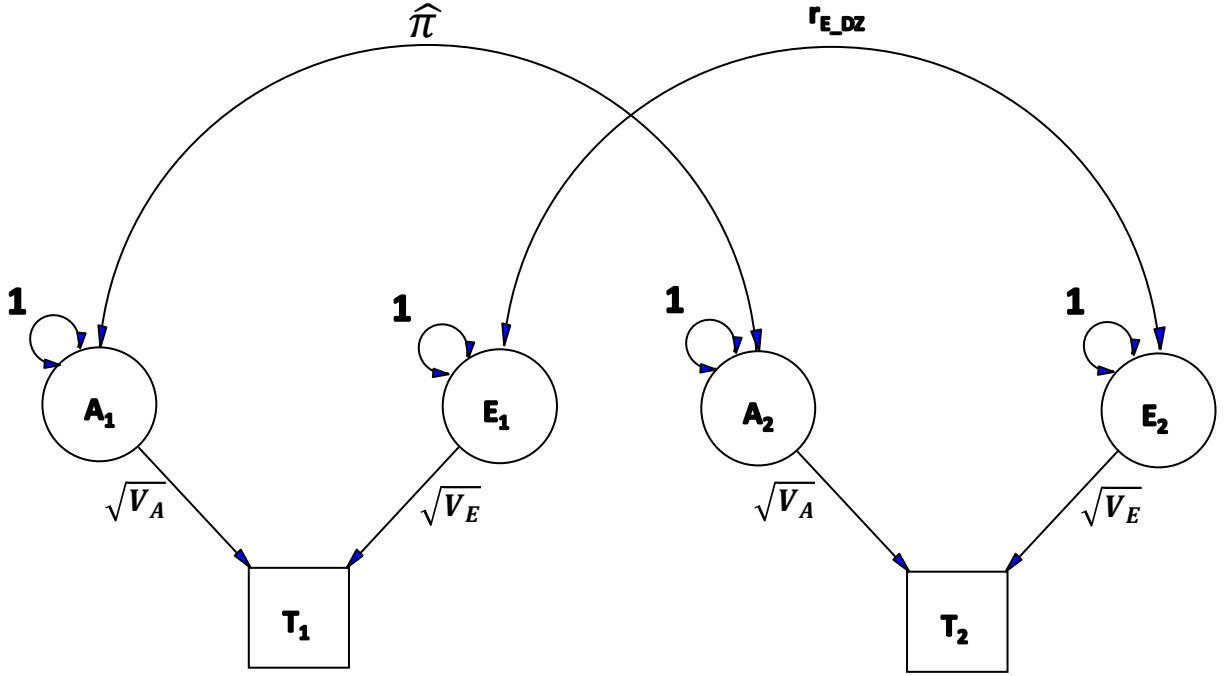
Path Model for the Augmented Classical Twin Design

MZ Twins:



$$cov_{MZ} = \begin{bmatrix} V_A + V_E & V_A + r_{E\_MZ}V_E \\ V_A + r_{E\_MZ}V_E & V_A + V_E \end{bmatrix}$$

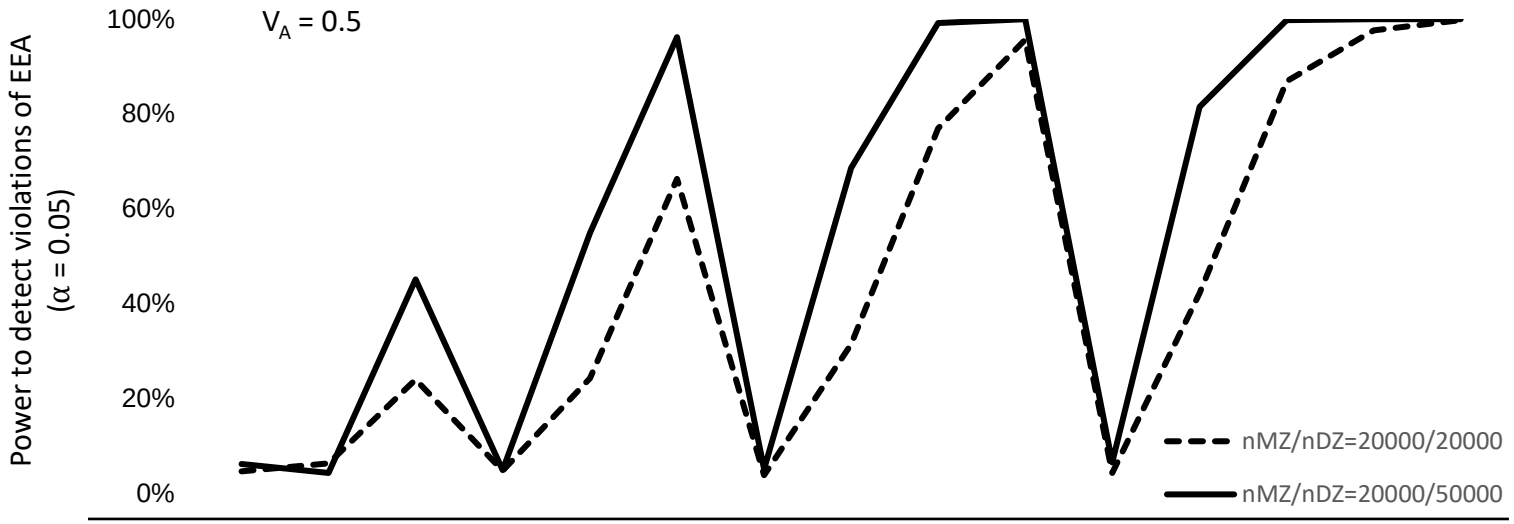
DZ Twins:



$$cov_{DZ} = \begin{bmatrix} V_A + V_E & \hat{\pi}V_A + r_{E\_DZ}V_E \\ \hat{\pi}V_A + r_{E\_DZ}V_E & V_A + V_E \end{bmatrix}$$

Test of EEA is whether $r_{MZ} = r_{DZ}$

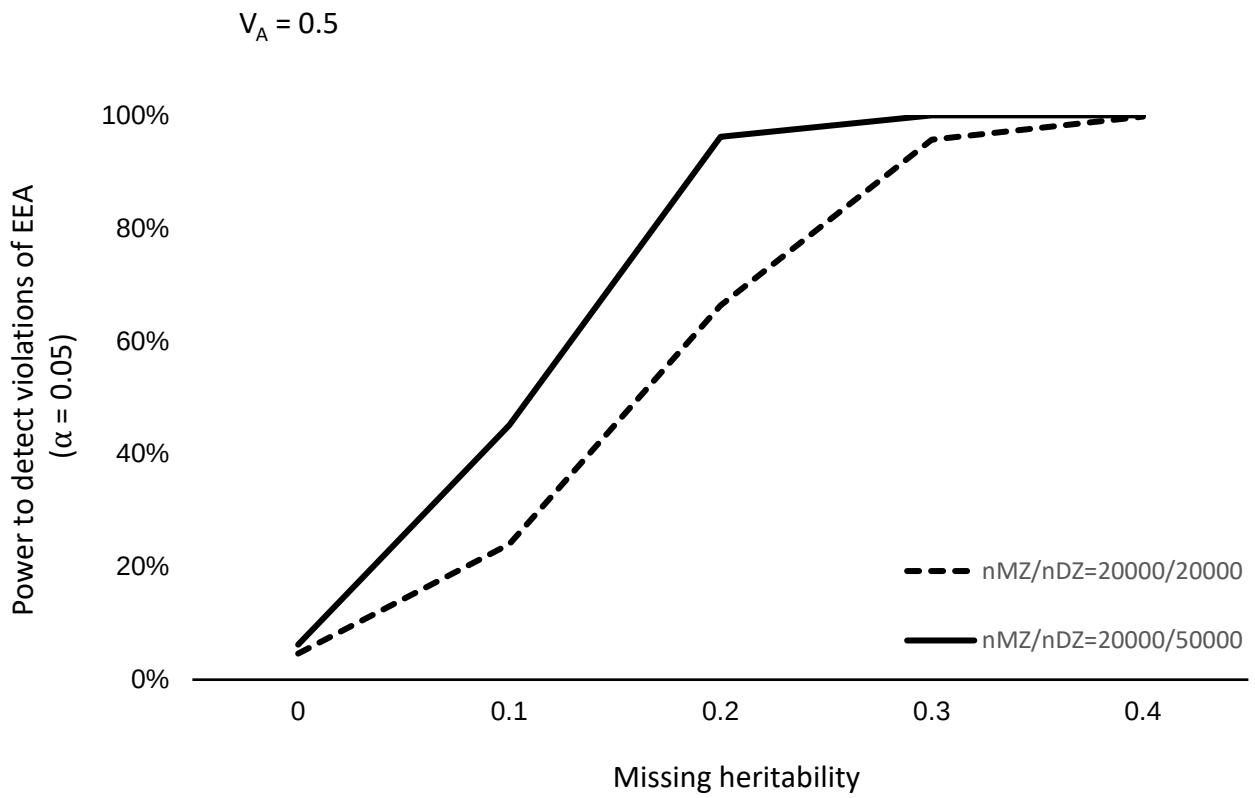
Power



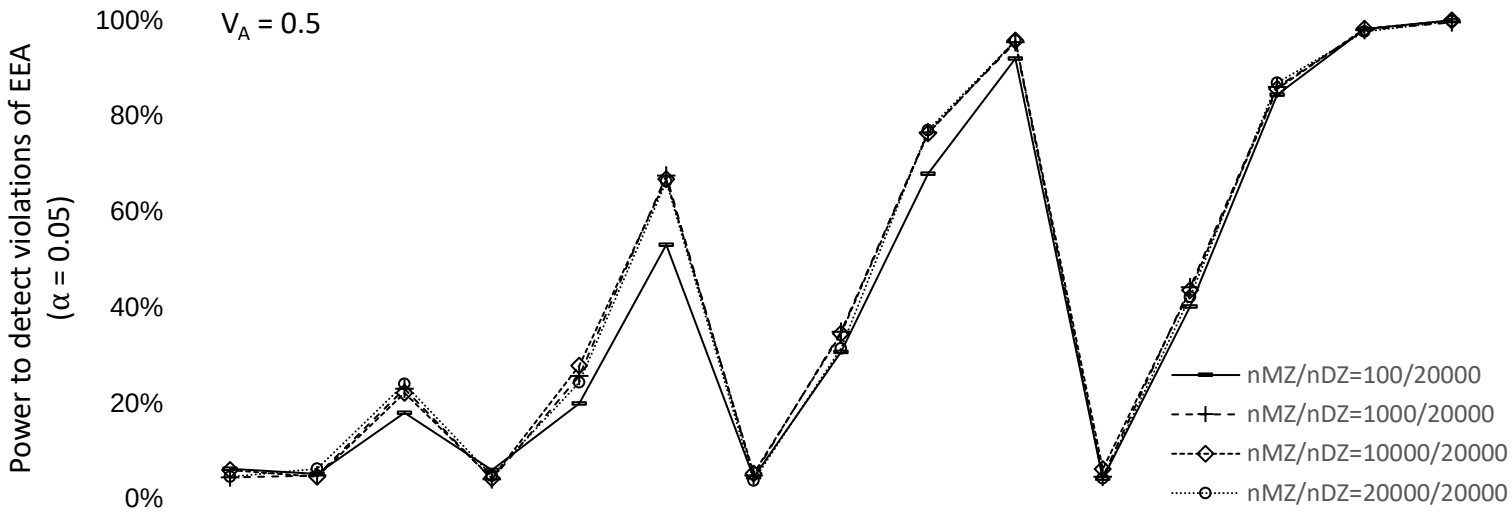
V_{C_MZ}	0	0.1	0.1	0.2	0.2	0.2	0.3	0.3	0.3	0.3	0.4	0.4	0.4	0.4	0.4
V_{C_DZ}	0	0.1	0	0.2	0.1	0	0.3	0.2	0.1	0	0.4	0.3	0.2	0.1	0
r_{E_MZ}	0	0.2	0.2	0.4	0.4	0.4	0.6	0.6	0.6	0.6	0.8	0.8	0.8	0.8	0.8
r_{E_DZ}	0	0.2	0	0.4	0.2	0	0.6	0.4	0.2	0	0.8	0.6	0.4	0.2	0

- Power increases with increased number of DZ twins, differences in environmental similarity between MZ and DZ twins, and increased proportion of variance due to the shared environment

Power as a Function of Missing Heritability



Effect of Number of MZ Twins on Power



V_{C_MZ}	0	0.1	0.1	0.2	0.2	0.2	0.3	0.3	0.3	0.3	0.4	0.4	0.4	0.4	0.4
V_{C_DZ}	0	0.1	0	0.2	0.1	0	0.3	0.2	0.1	0	0.4	0.3	0.2	0.1	0
r_{E_MZ}	0	0.2	0.2	0.4	0.4	0.4	0.6	0.6	0.6	0.6	0.8	0.8	0.8	0.8	0.8
r_{E_DZ}	0	0.2	0	0.4	0.2	0	0.6	0.4	0.2	0	0.8	0.6	0.4	0.2	0

Intuition:

V_{E_MZ} **MZ twins**

V_A **DZ twins**

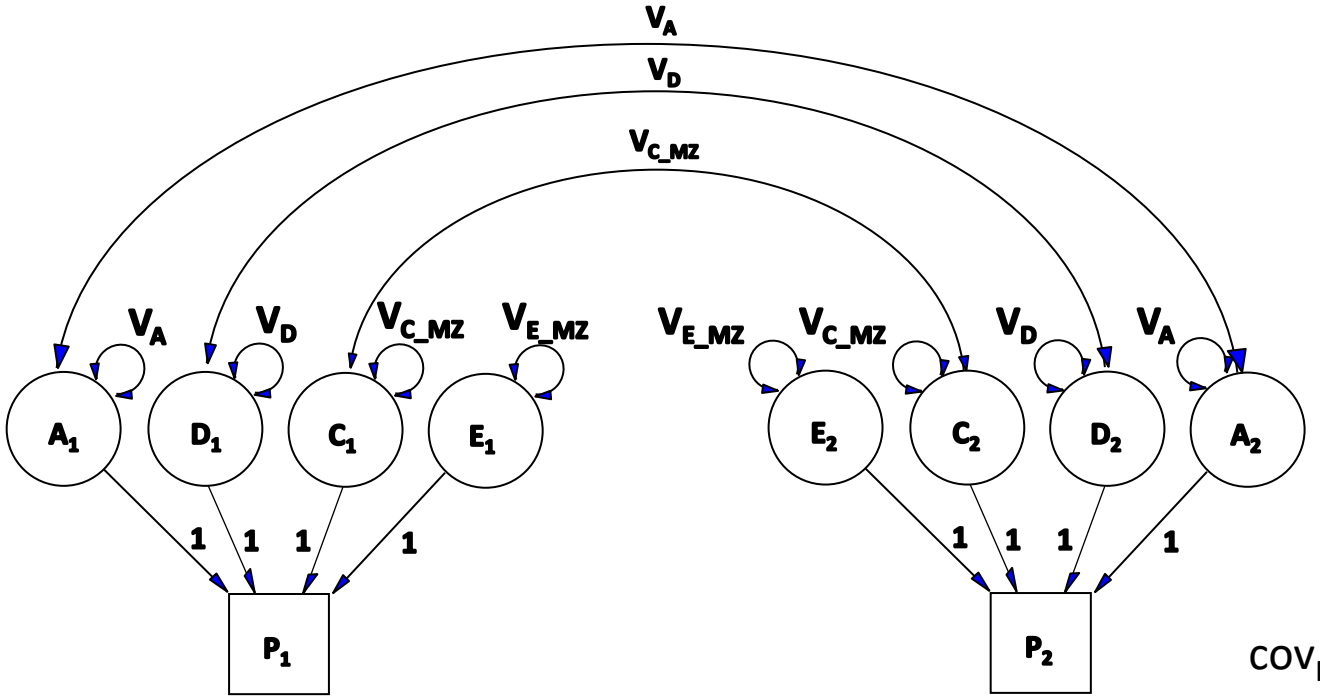
→ **V_{C_MZ}**

V_{C_DZ} **DZ twins**

V_{E_DZ} **DZ twins**

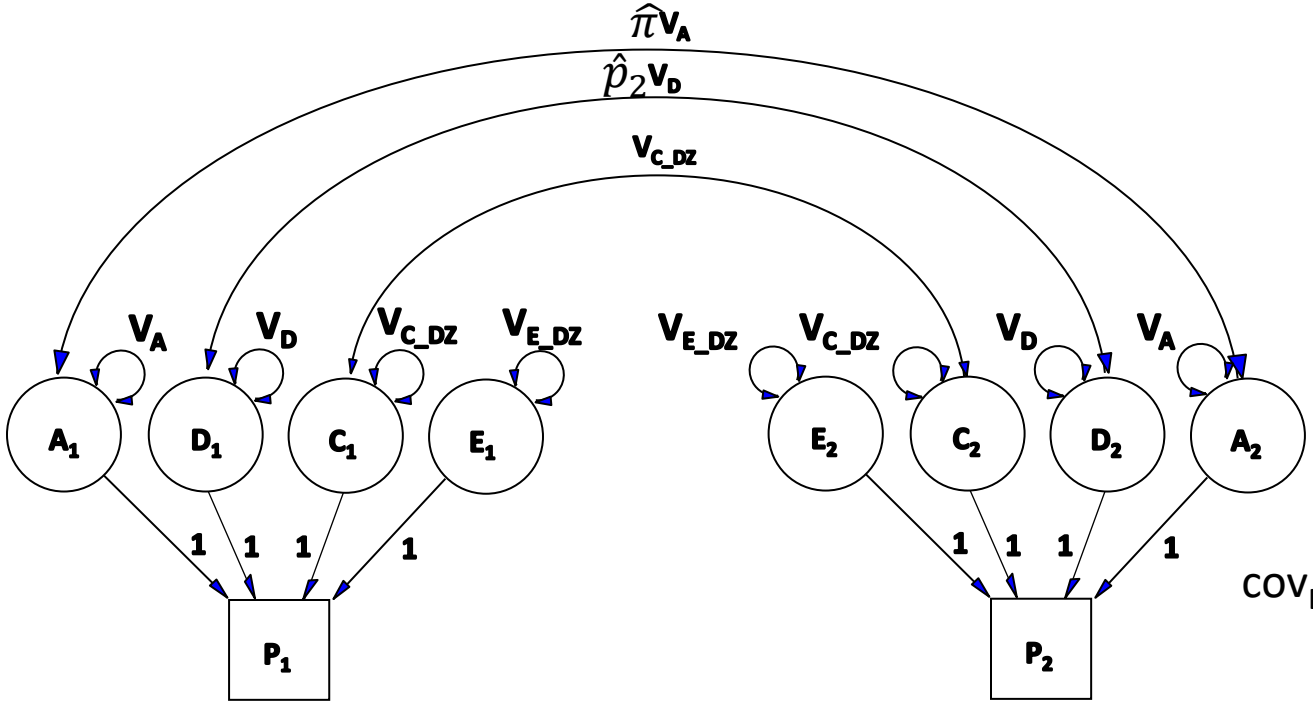
Including Dominance in the Augmented Classical Twin Design

MZ Twins:



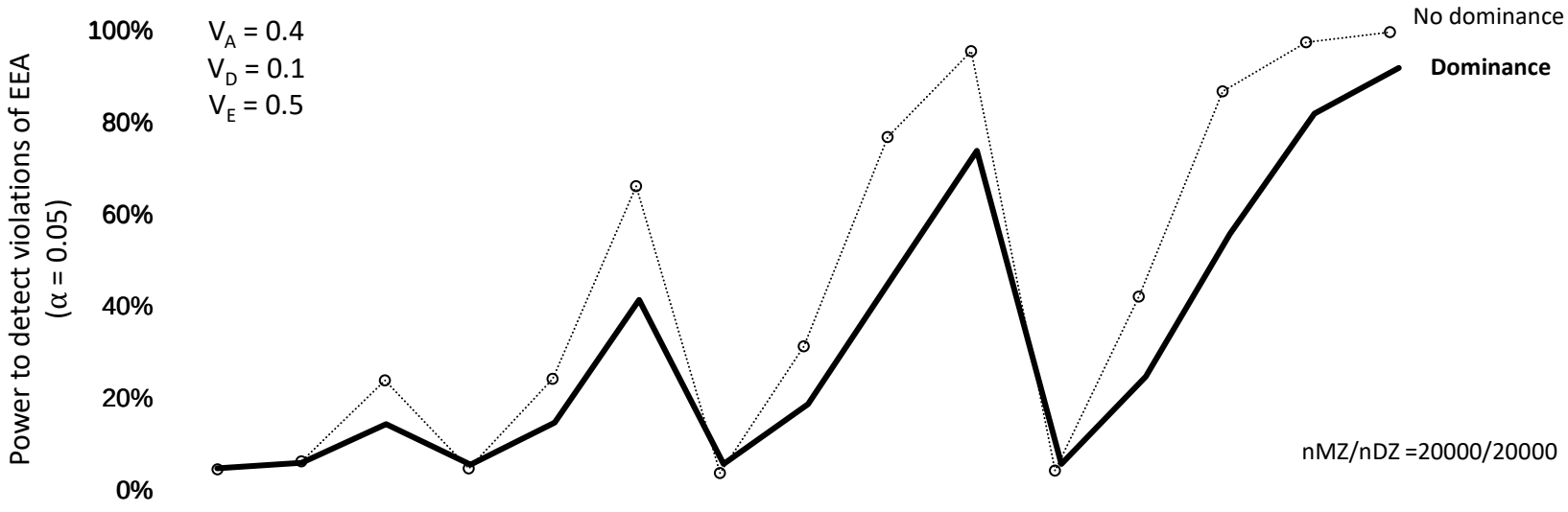
$$COV_{MZ} = \begin{bmatrix} V_A+V_D+V_{C\_MZ}+V_{E\_MZ} & V_A+V_D+V_{C\_MZ} \\ V_A+V_D+V_{C\_MZ} & V_A+V_D+V_{C\_MZ}+V_{E\_MZ} \end{bmatrix}$$

DZ Twins:



$$COV_{DZ} = \begin{bmatrix} V_A+V_D+V_{C\_DZ}+V_{E\_DZ} & \hat{\pi}V_A+\hat{p}_2V_D+V_{C\_DZ} \\ \hat{\pi}V_A+\hat{p}_2V_D+V_{C\_DZ} & V_A+V_D+V_{C\_DZ}+V_{E\_DZ} \end{bmatrix}$$

Inclusion of Dominance



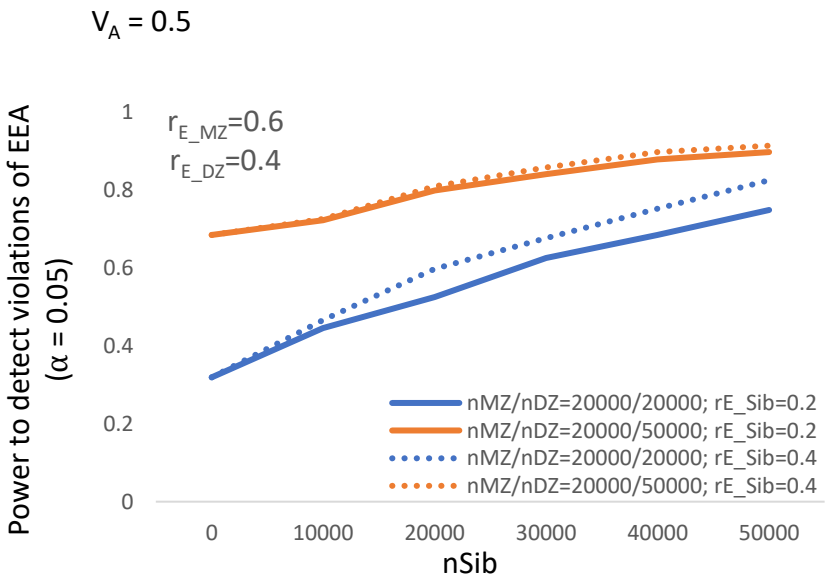
V_{C_MZ}	0	0.1	0.1	0.2	0.2	0.2	0.3	0.3	0.3	0.3	0.4	0.4	0.4	0.4	0.4
V_{C_DZ}	0	0.1	0	0.2	0.1	0	0.3	0.2	0.1	0	0.4	0.3	0.2	0.1	0
r_{E_MZ}	0	0.2	0.2	0.4	0.4	0.4	0.6	0.6	0.6	0.6	0.8	0.8	0.8	0.8	0.8
r_{E_DZ}	0	0.2	0	0.4	0.2	0	0.6	0.4	0.2	0	0.8	0.6	0.4	0.2	0

- Presence of unmodelled dominance gives spurious evidence for violation of EEA (not shown)
- Modelling dominance component provides appropriate Type 1 error rates for test of EEA
- Presence of modelled dominance results in slightly reduced power relative to the no dominance situation

Model Assumptions

- The “usual” (no GE-cov, GxE, assortment, bivariate normality etc)
- ANY unmodelled factor that increases MZ pair similarity relative to DZ pair similarity will increase estimates of rE_MZ relative to rE_DZ and provide spurious evidence for violation of the EEA
- Genetic non-additivity (Epistasis!)
- Errors in IBD calculation in DZ pairs
- Gene x Age interactions (Balance age across MZ and DZ twins)
- Gene x Sex interactions (Same sex twins only!)
- Important to model the X chromosome!

Inclusion of non-twin Siblings?



- Power increases with the number of sibling pairs
- Power increases with sibling common environmental sharing

Conclusions

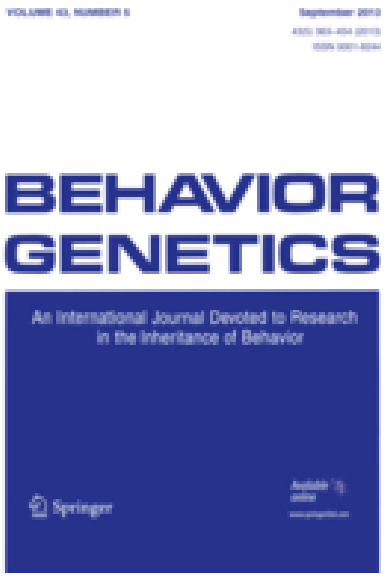
- Augmented Classical Twin Design can identify violations of the EEA
- Requires very large numbers of (DZ) pairs for power
- Model extremely sensitive to unmodelled factors that increase MZ compared to DZ similarity (e.g. epistasis)
- Inclusion of non-twin siblings in the analysis may improve power, but would limit application of the method because of gene x age concerns

Acknowledgments

- The Organizers
- Liang-Dar (“Daniel”) Hwang
- Brittany Mitchell
- Sarah Medland
- Nick Martin
- Mike Neale

Daniel Hwang

Brittany Mitchell



The Augmented Classical Twin Design: Incorporating Genome-Wide Identity by Descent Sharing Into Twin Studies in Order to Model Violations of the Equal Environments Assumption. **Hwang LD**, Mitchell BL, Medland SE, Martin NG, Neale MC, **Evans DM.** Behav Genet. 2021 Feb 13. doi: 10.1007/s10519-021-10044-0.