Model building & assumptions

Matt Keller, Sarah Medland, Hermine Maes

TC21

March 2008

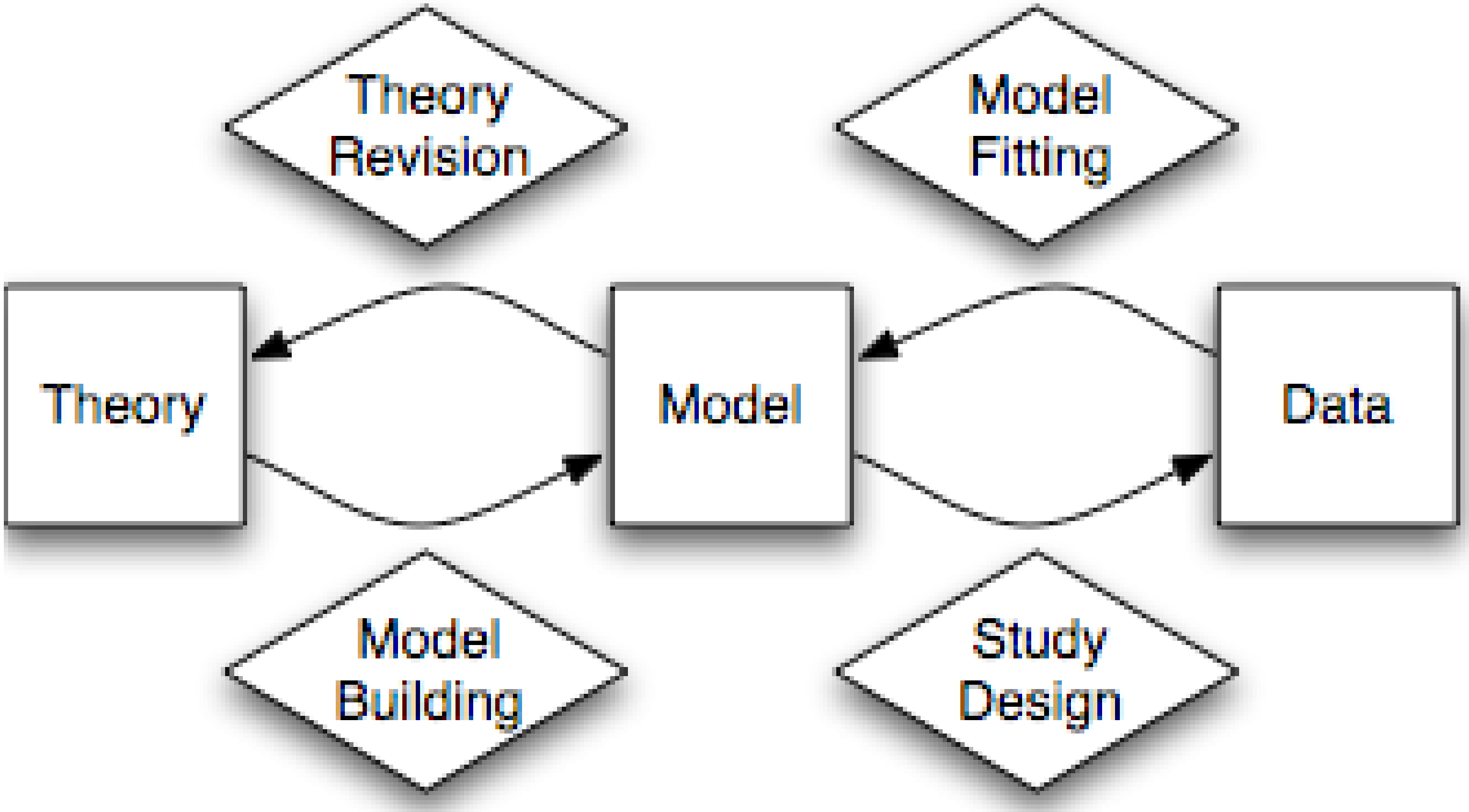
Acknowledgments

- John Jinks
- David Fulker
- Robert Cloninger
- John DeFries
- Lindon Eaves
- Nick Martin
- Dorret Boomsma
- Michael Neale

Behavior Genetic Methods

- For thirty years, BG studies have revolved around BUILDING and TESTING MODELS
- In particular, modeling means, variances, and covariances of genetically informative relatives
- Such models are based on our understanding of why relatives are similar to one another

Models mediate between theory and data



Example Model Building with MZ & DZ twins only

- Theory: variance due to E & A, & either D or C
- 3 unknowns > 3 independently informative equations needed to solve for VE, VA & VC (ACE) or VE, VA & VD (ADE)
- Mx arrives at much the same conclusion using a much different algorithm (ML). But for most BG models fitted today, no such easy close formed solutions exist. ML approach is better.

Algebra for VA & VC assuming the ACE model

- VP = VA + VC + VE = 1
- rMZ = VA + VC
- rDZ = .5VA +VC

- 1 - rMZ = VE
- rMZ - rDZ = .5VA
  - 2(rMZ - rDZ) = VA
- rMZ - 2rDZ = VA + VC - VA - 2VC = - VC
  - 2rDZ - rMZ = VC or
  - VP - VE - VA = VC

Practical I

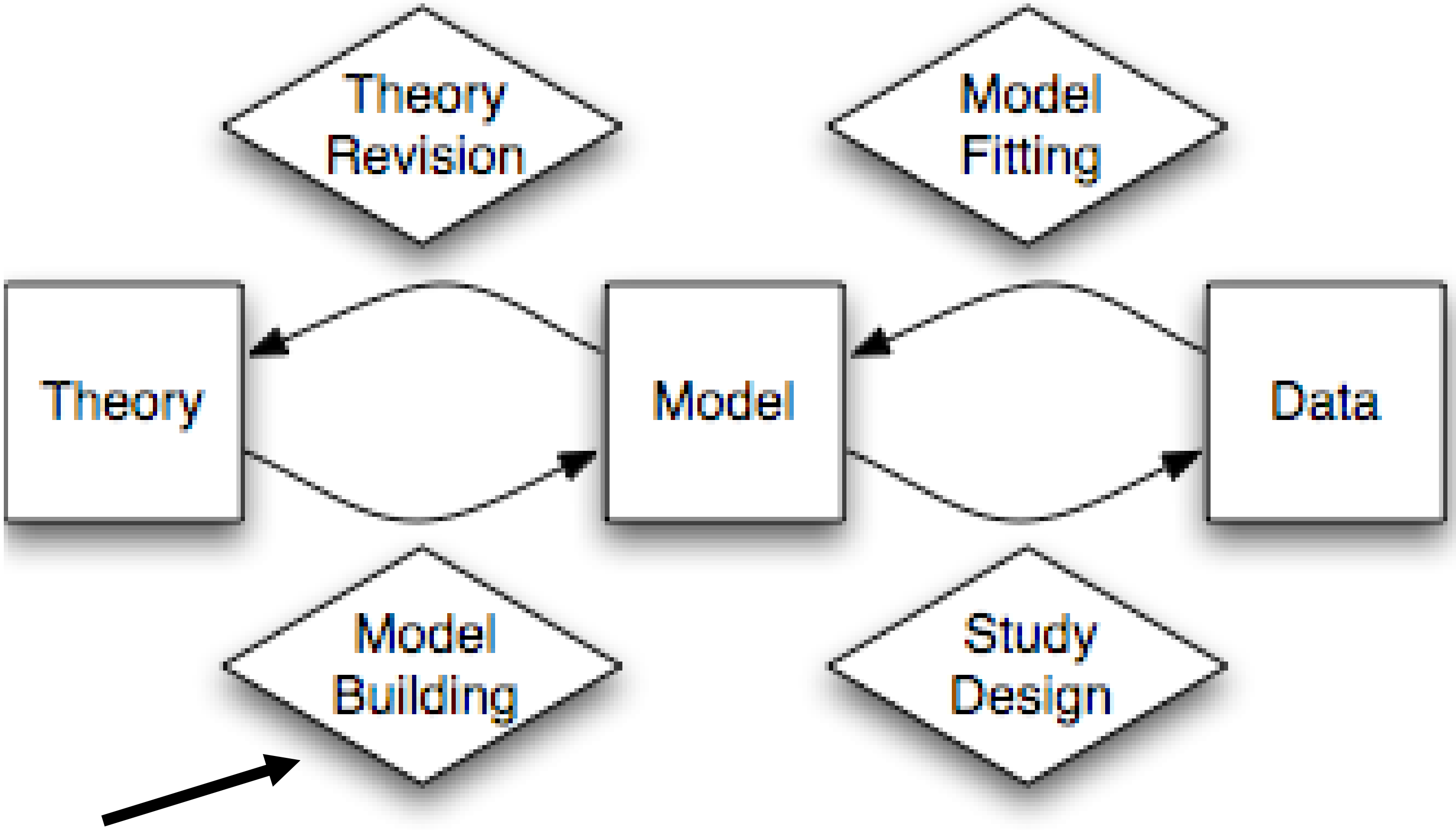
- Solve for VA & VD using the following equations:
  - VP = VA + VD + VE = 1
  - rMZ = VA + VD
  - rDZ = .5 VA +.25VD

Algebra for VA & VD assuming the ADE model

- VP = VA + VD + VE = 1
- rMZ = VA + VD
- rDZ = .5 VA +.25VD

- 1 - rMZ = VE
- rMZ - 4rDZ = VA + VD - 2VA - VD = -VA
  - 4rDZ - rMZ = VA
- rMZ - 2rDZ = VA + VD - VA - .5VD = .5VD
  - 2(rMZ - 2rDZ) = 2rMZ - 4rDZ = VD

In reality, models are constrained by data's ability to test particular theories



Model Assumptions

- All models must make simplifying assumptions. It is no different in BG, e.g., with MZ & DZ twins reared together:
- ACE Model
  - Assumes no D
- ADE Model
  - Assumes no C

Testable Assumptions with twin data

- Normality of residuals
- No AxC, AxE, CxE Interactions
- No AC, AE, CE Correlations
- Equal Environments Assumption
- No Sibling Interaction
- No Sex x A(CE) Interaction

Testable Assumptions with additional data

- Multivariate
  - Measurement Invariance
- Longitudinal
  - No Age x A(CE) Interaction
  - Measurement Error vs Unique Environment
- Other Relatives
  - No Assortative Mating
  - C&D, Cultural Transmission

Not Testable Assumptions

- No correlated errors
- No epistasis

Model Assumptions II

- Why must we make these simplifying assumptions? E.g., why can’t we estimate C & D at same time using twins only?
- Solve the following two equations for VA, VC, and VD:
  - $rMZ = VA + VD + VC$
  - $rDZ = 1/2VA + 1/4VD + VC$

Classical Twin Design Approach

- When rMZ < 2rDZ, we estimate VC and VA
  - if rMZ>2rDZ then VC negative or at 0 boundary
  - VA = 2(rMZ-rDZ)
  - VC = 2rDZ-rMZ
  - assumption: VD = 0
- When rMZ > 2rDZ, we estimate VD and VA
  - if rMZ<2rDZ then VD negative or at 0 boundary
  - VA = 4rDZ-rMZ
  - VD = 2rMZ-4rDZ
  - Assumption: VC = 0

Sensitivity Analysis

- What happens when our assumptions are wrong?
- Quantification of what happens when our models are wrong is a STRENGTH, not a weakness of model-based science!

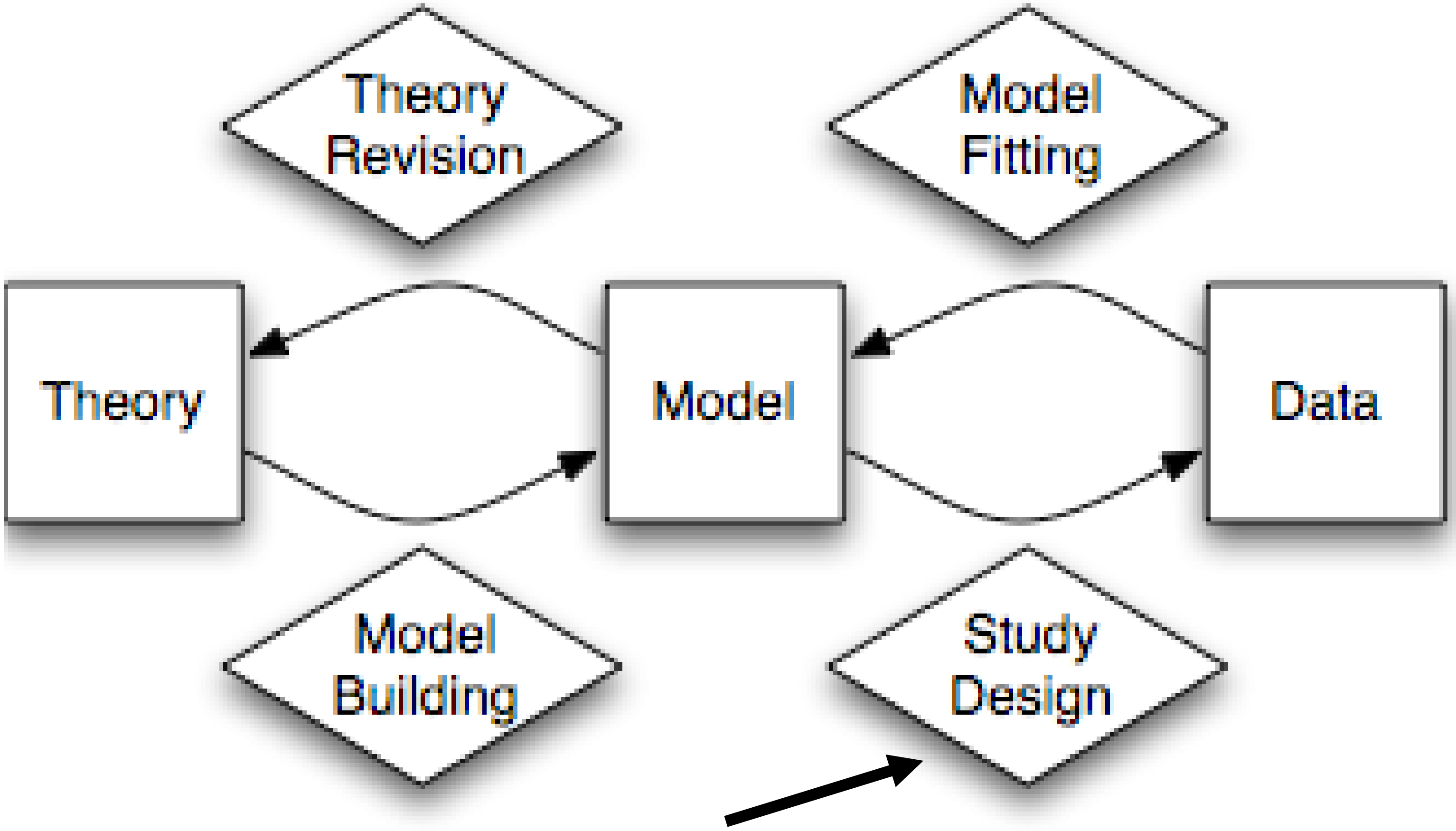
Practical II

- Given rMZ = .80 & rDZ = .30
- Solve algebraically for VA and VD when VC is assumed to be 0 (the normal case):
  - VA = 4rDZ - rMZ
  - VD = 2rMZ - 4rDZ
- Solve algebraically for VA and VD when VC is assumed to be .05 (a possibility after all) implies rMZ-VC = .75 & rDZ-VC = .25

Sensitivity Analysis Practical

- Given rMZ = .80 & rDZ = .30
- Under normal assumptions (VC=0):
  - VA = .4
  - VD = .4
- Under alternative assumptions (VC=.05):
  - VA = .25 = 4(.25) - .75 [4rDZ - rMZ]
  - VD = .5 = 2(.75) - 4(.25) [2rMZ - 4rDZ]
  - had we run a twin-only model, we would have underestimated VD & VC and overestimated VA

In reality, models determine study designs needed to test them



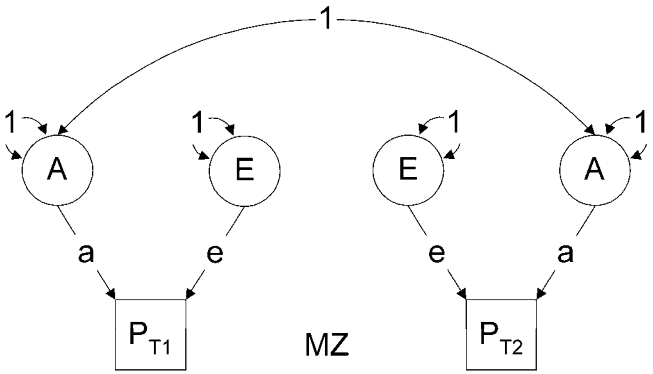
GeneEvolve

- Given a model, a study design is chosen (and assumptions made)
- Data are simulated under complex true world with GeneEvolve
- We can evaluate how biased results are given the chosen design/assumptions.

Simulation: True World

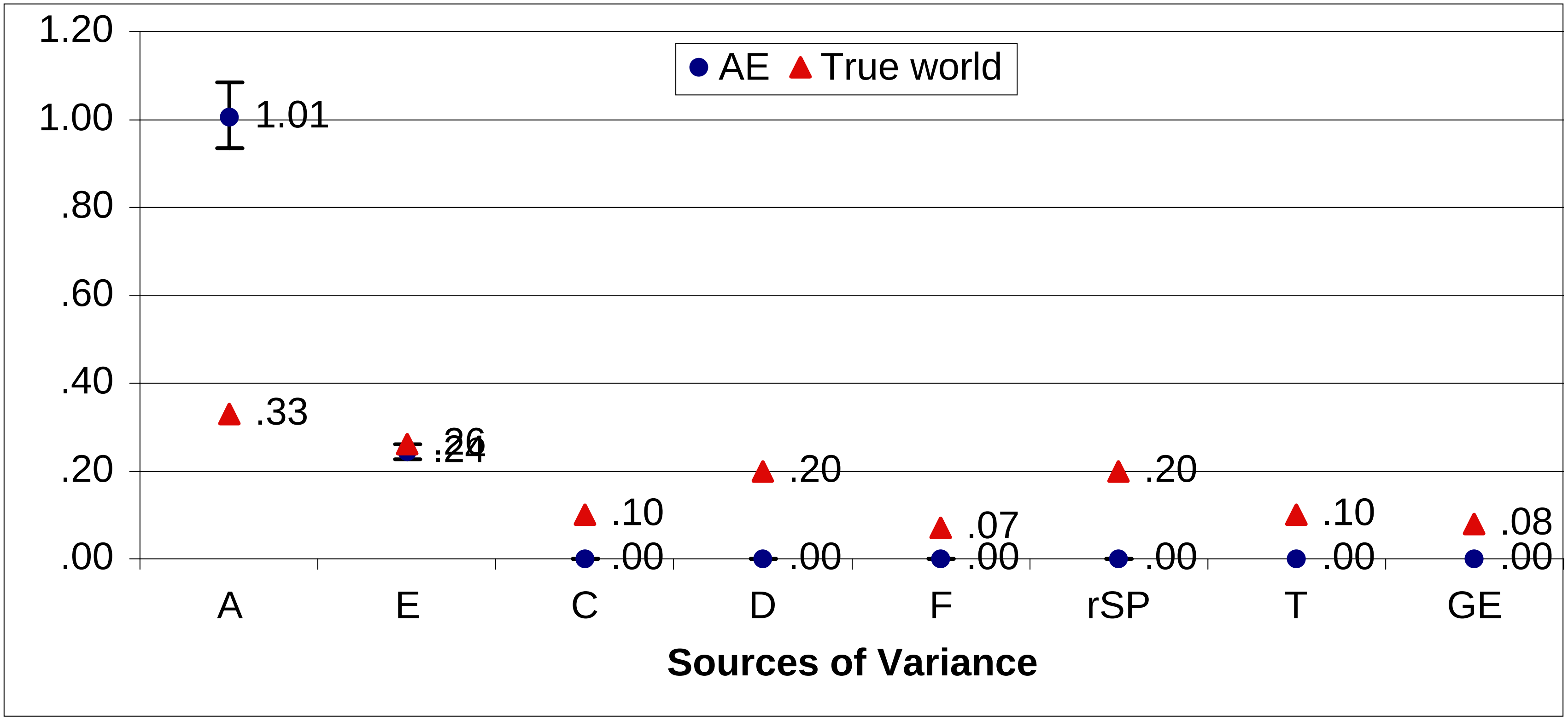
- VA=.33
- VC=.1
- VE=.36
- VD=.2
- VF=.07 cultural transmission
- COVAF=.08 GE covariance
- rSP=.2 spousal correlation

AE Model

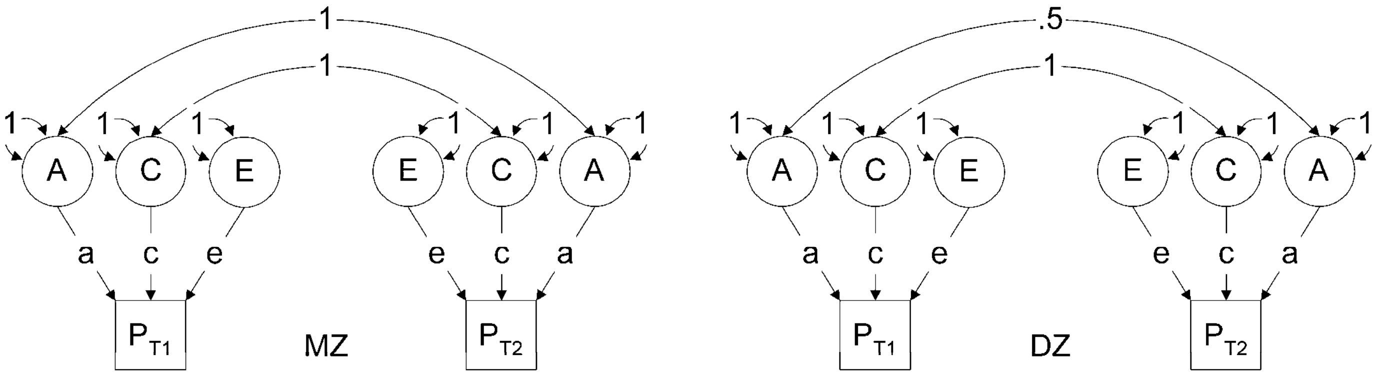


- Design: rMZ
- Assumptions: no C, D, assortment

AE Model vs True World

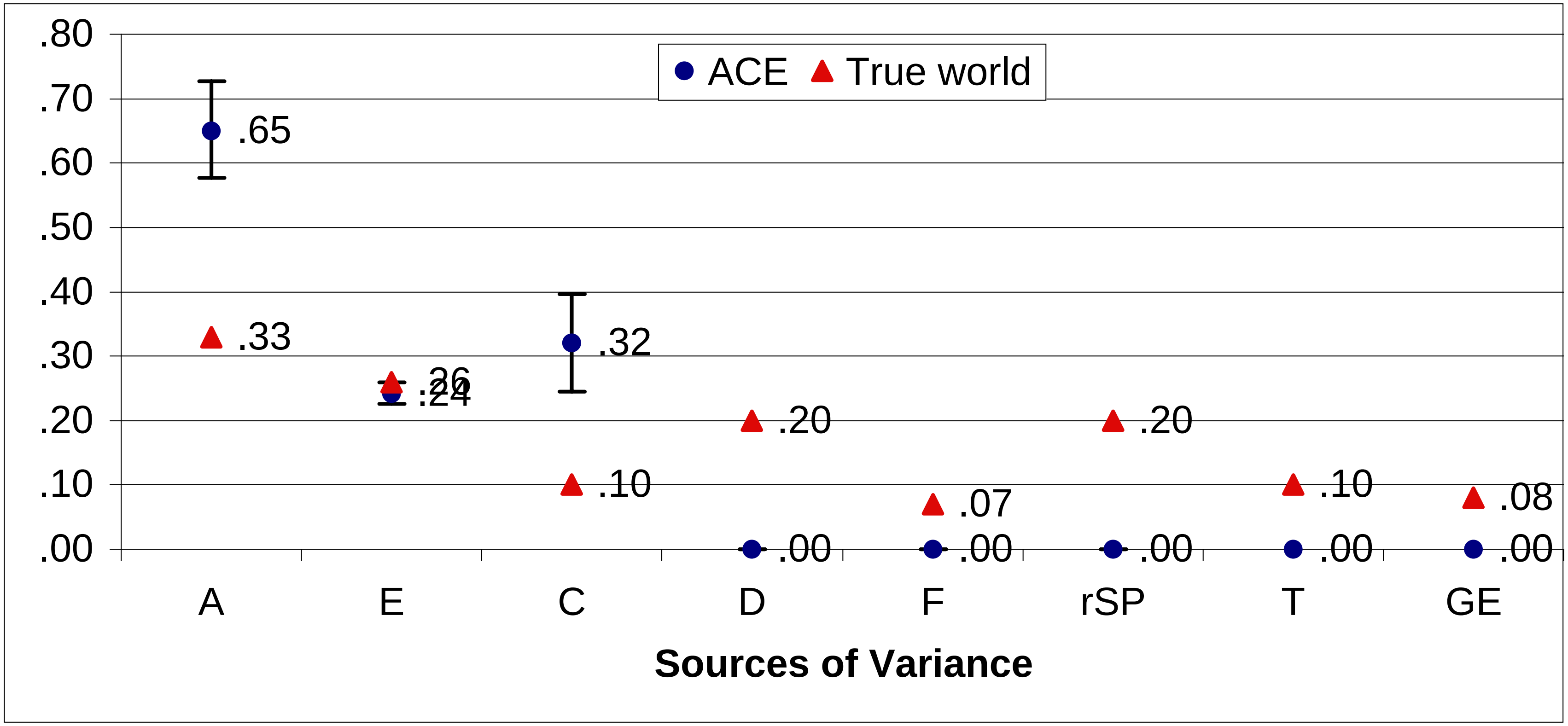


ACE Model

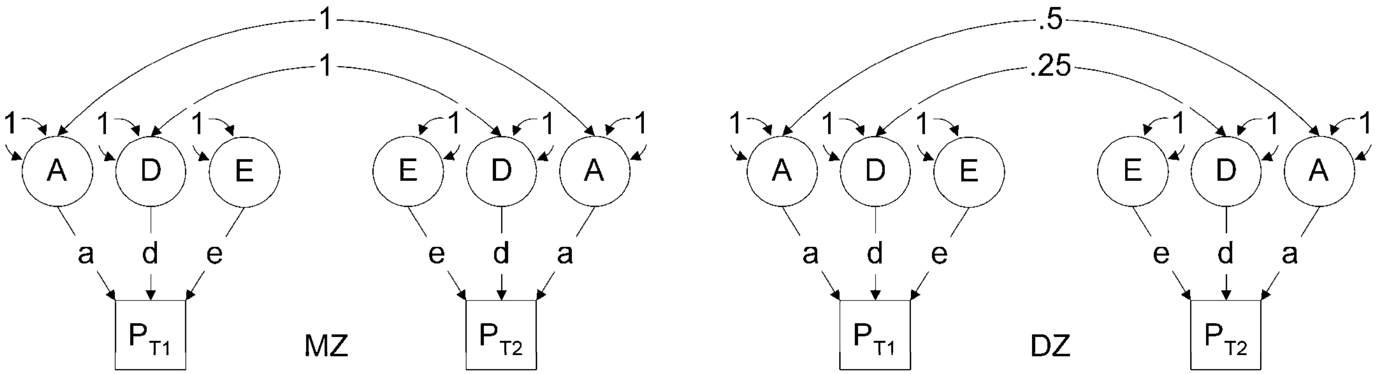


- Design: rMZ & rDZ
- Assumptions: no D, assortment

ACE Model vs True World

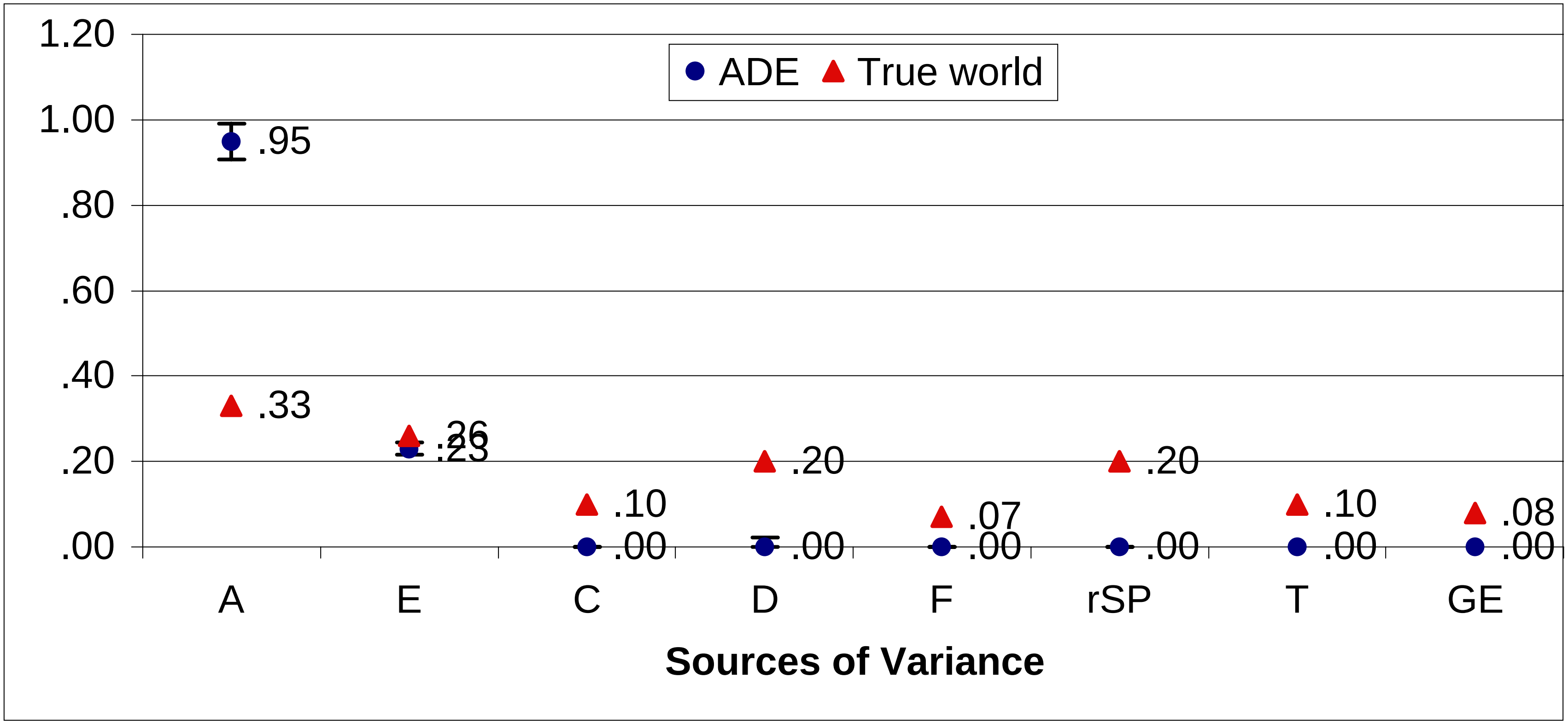


ADE Model

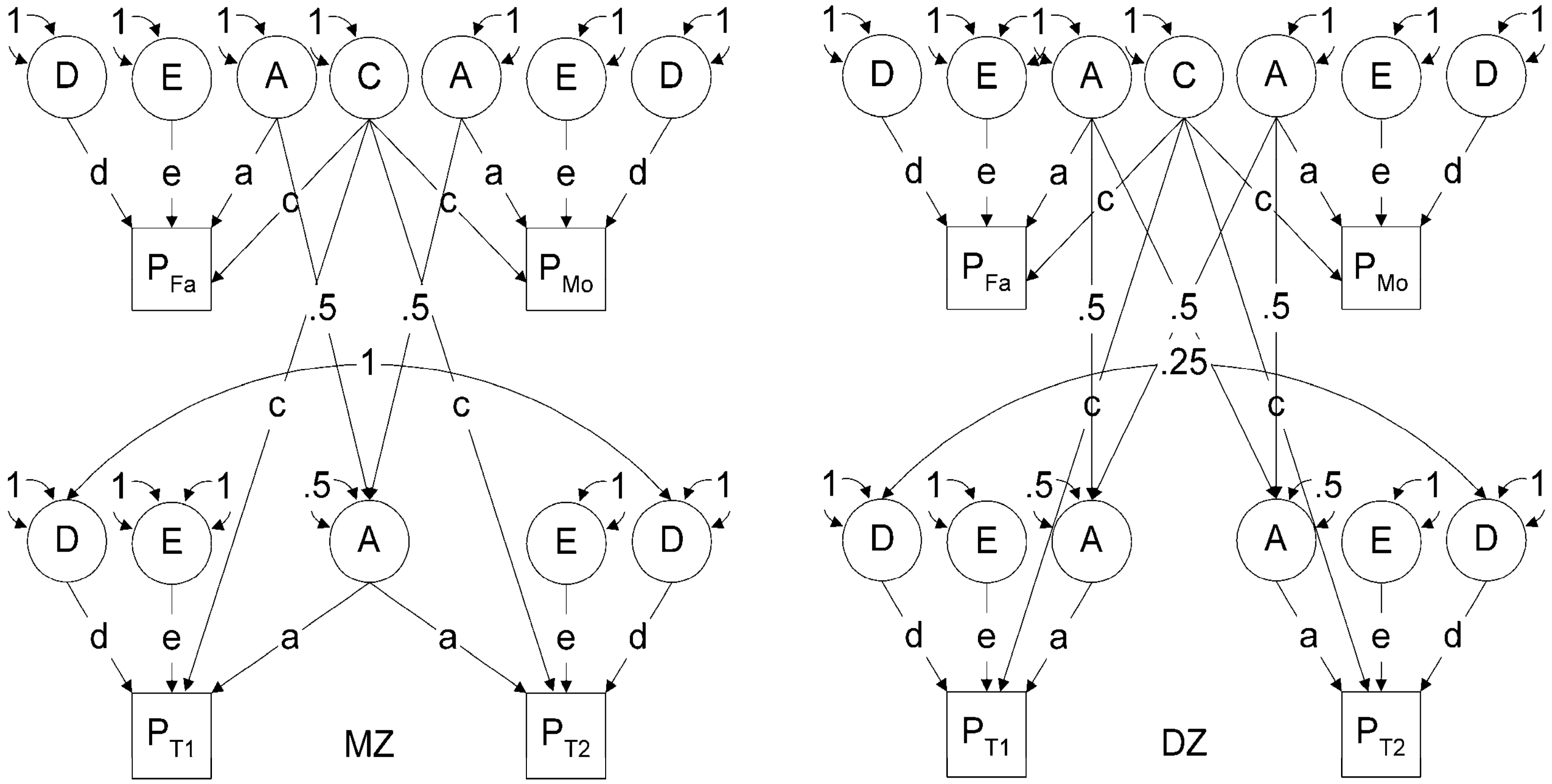


- Design: rMZ & rDZ
- Assumptions: no C, assortment

ADE Model vs True World



ACED Model



1
D
E
A
C
A
E
D
d
e
a
c
P_Fa
P_Mo
.5
.5
1
P_T1
P_T2
MZ
1
D
E
A
C
A
E
D
d
e
a
c
P_Fa
P_Mo
.5
.5
.5
.25
P_T1
P_T2
DZ

ACED II

- Design: rMZ & rDZ & parents
- Assumptions: common C, no assortment

- Alternative Designs:
  - rMZa & rDZa (twins reared apart)
  - rMZ & rDZ & half-sibs

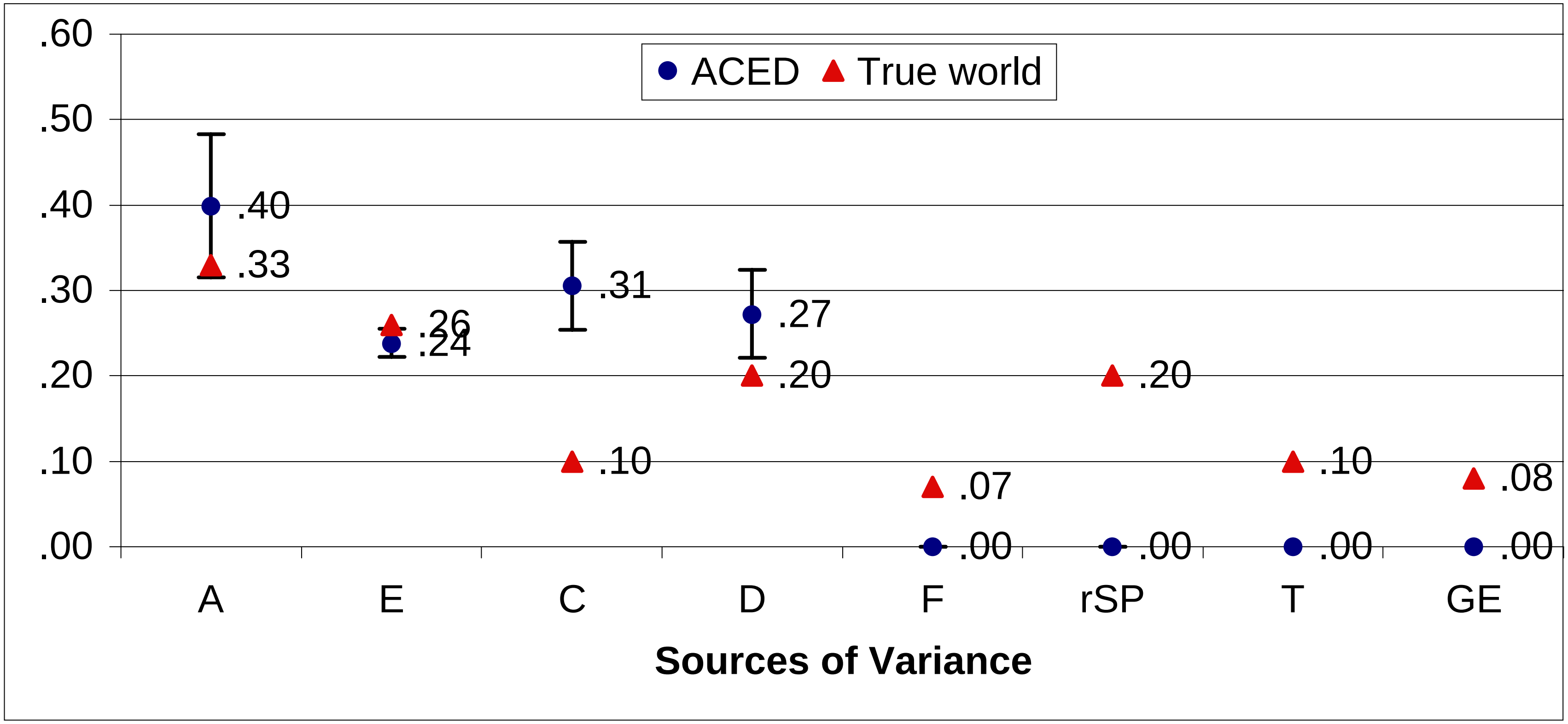
rMZa & rDZa
(twins reared apart)

- rMZ= VA + VC + VD
- rDZ= .5VA + VC + .25VD
- rMZA= VA + VD
- rDZA= .5VA + .25VD

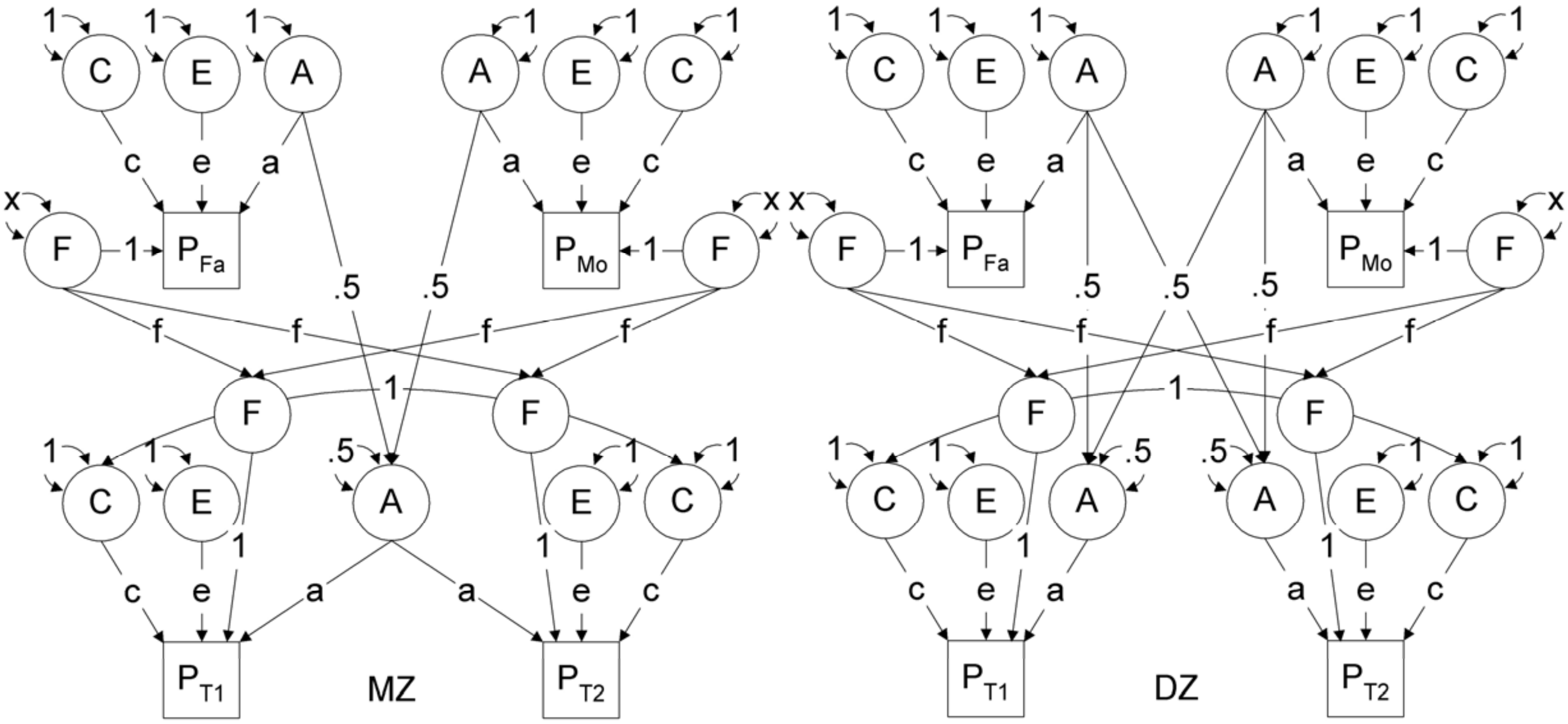
rMZ & rDZ & halfsibs rHS

- rMZ= VA + VC + VD
- rDZ= .5VA + VC + .25VD
- rHS= .25VA + VC

ACED Model vs True World



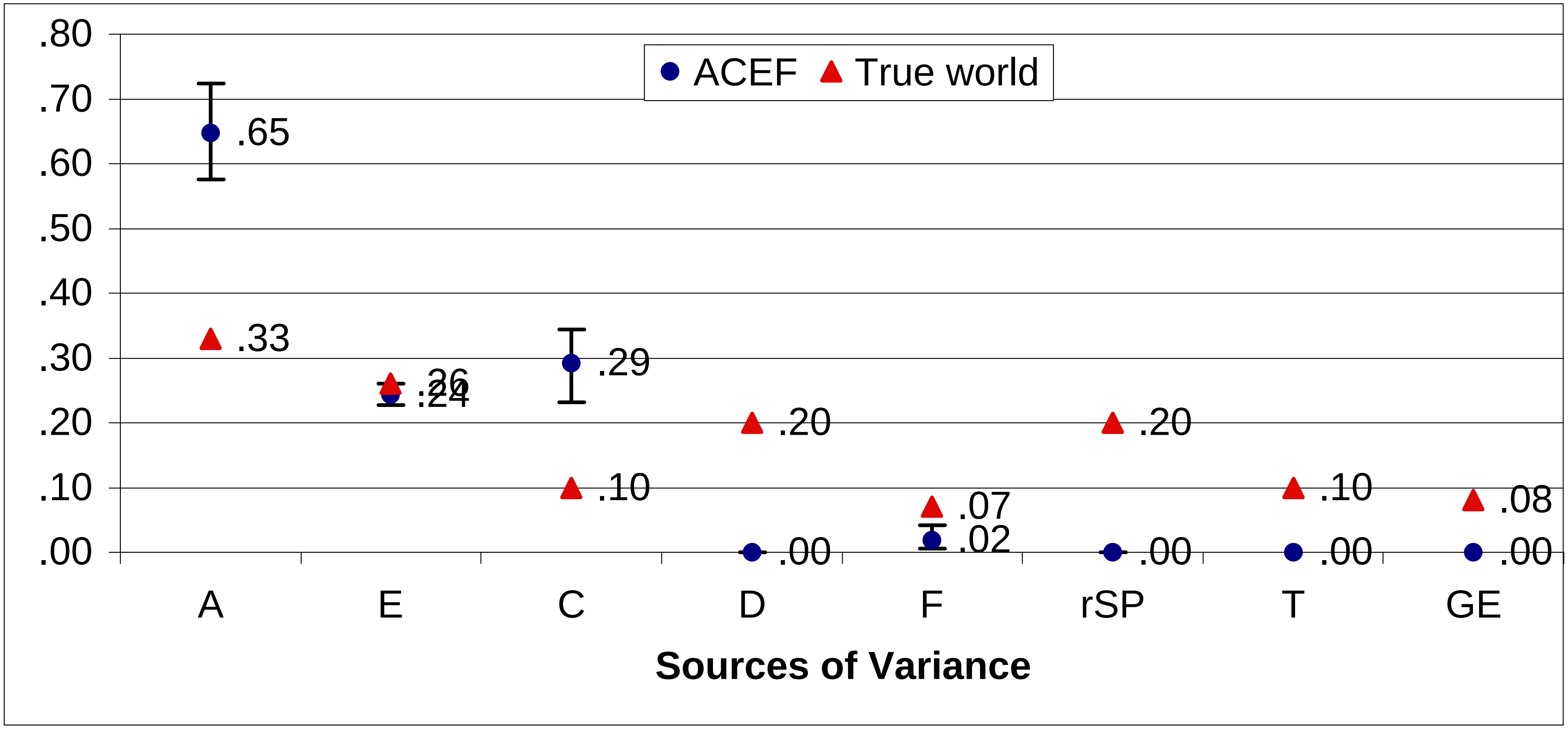
ACEF Model



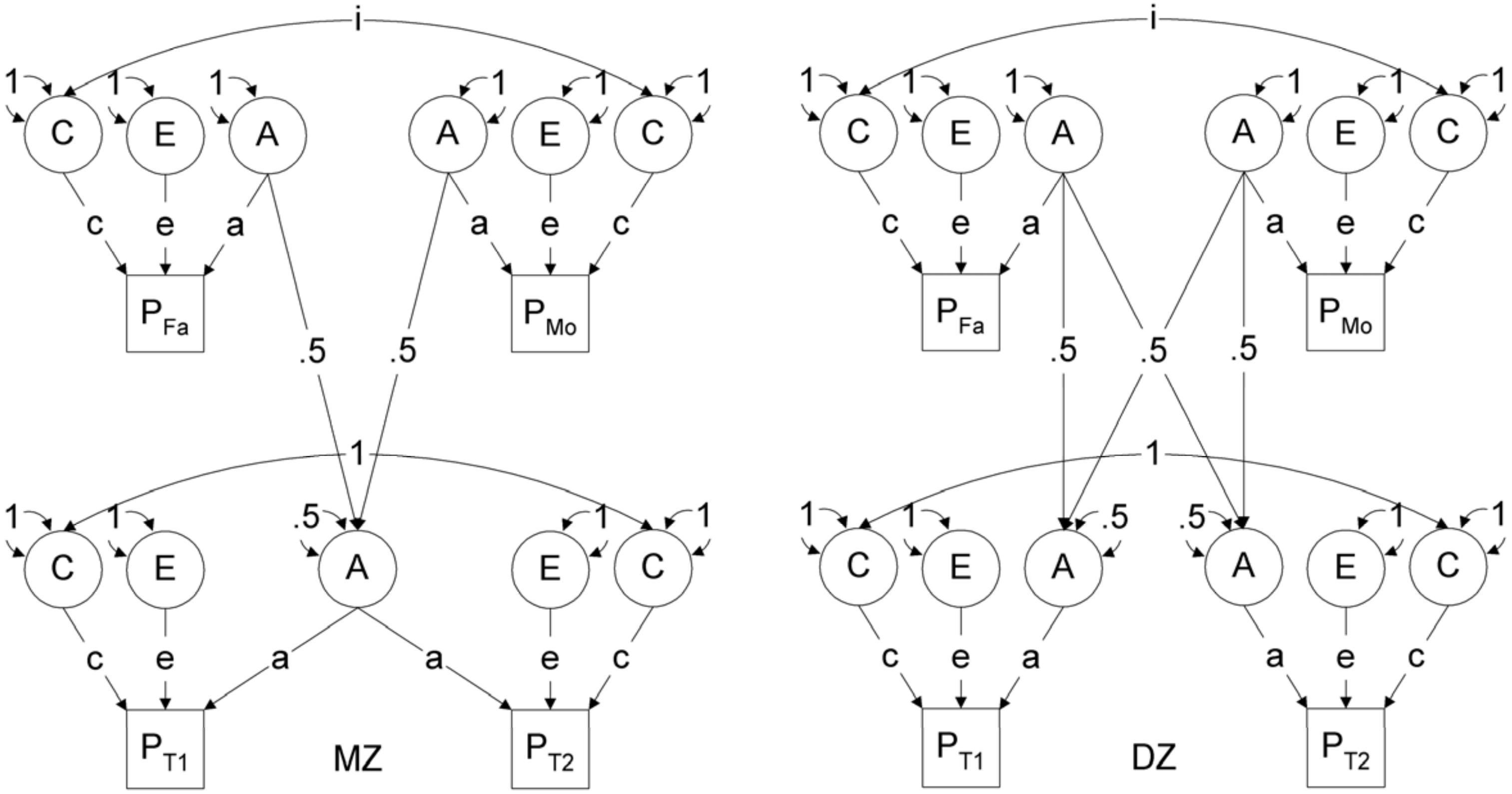
ACEF II

- Partition c^2 in cultural transmission and non-parental shared environment
- Design: rMZ & rDZ & parents
- Assumptions: no assortment, D

ACEF Model vs True World



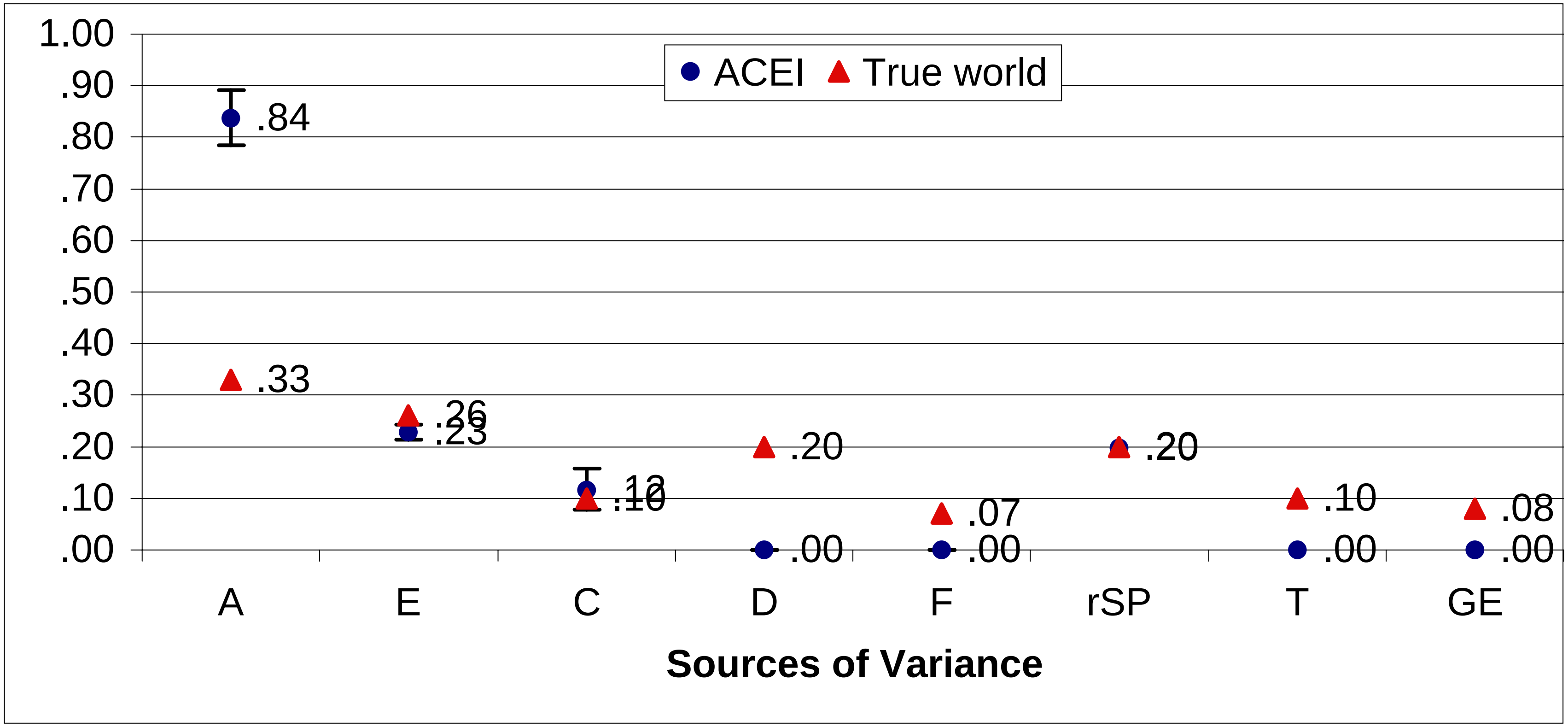
ACEI Model



ACEI II

- Account for assortative mating to correct estimates of a^2 and c^2
- Design: rMZ & rDZ & parents
- Assumptions: non-parental C, no D

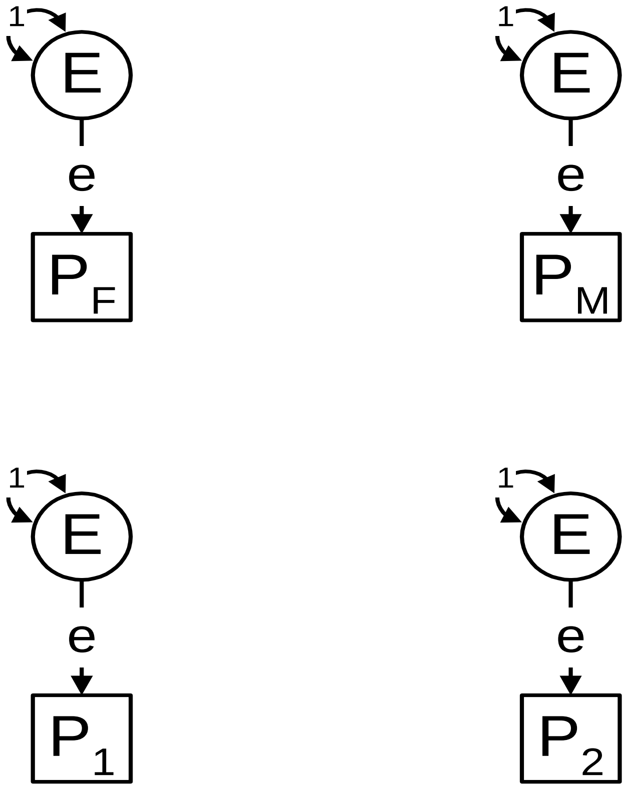
ACEI Model vs True World



Twins & Parents Design

- ACE +
- Dominance OR
- Cultural Transmission
  - Different mechanisms
- Assortative Mating
  - Different mechanisms

Path Diagram Twins & Parents



P_F : Phenotype Father

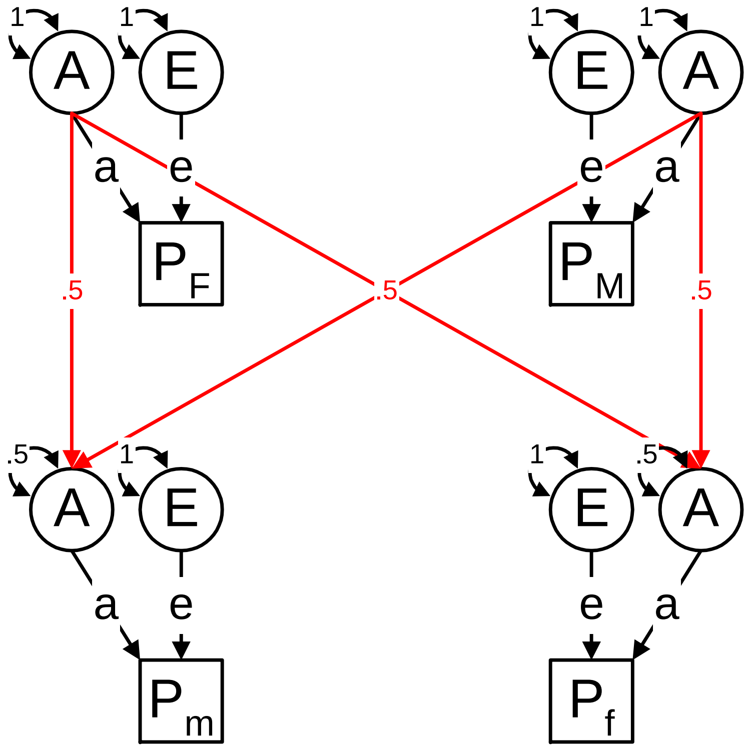
P_M : Phenotype Mother

P_1 : Phenotype Twin 1

P_2 : Phenotype Twin 2

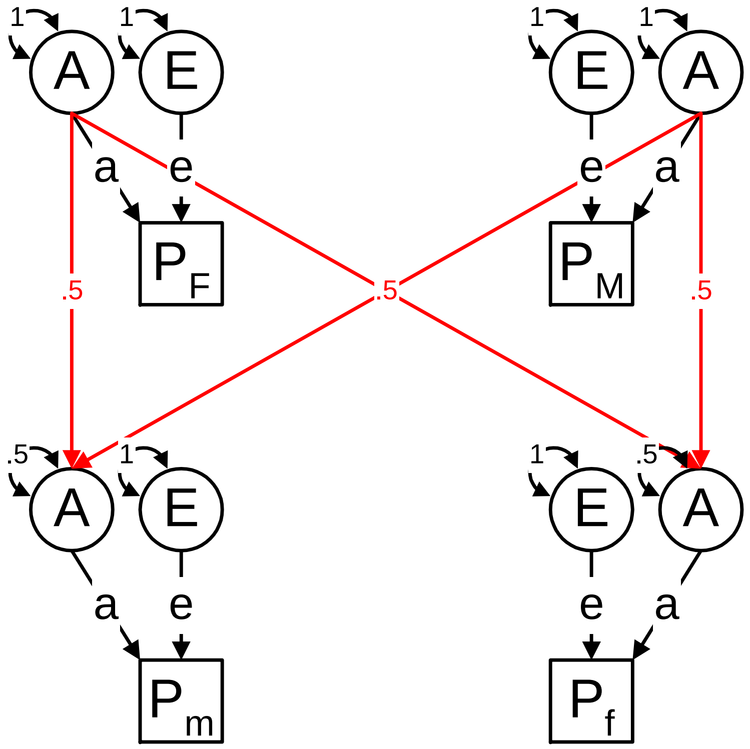
Genetic Transmission Model

- Genetic transmission
  - Fixed at .5
- Residual Genetic Variance
  - Fixed at .5
  - Equilibrium of variances across generations



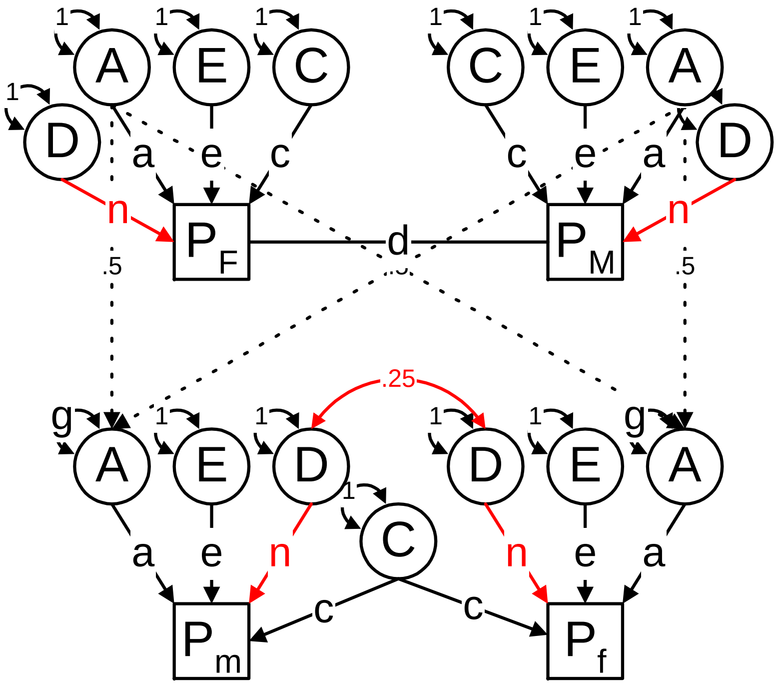
Genetic Transmission Expectations

- rSP= 0
- rPO= .5a^2
- rMZ= a^2
- rDZ= .5a^2
- Var= a^2+e^2



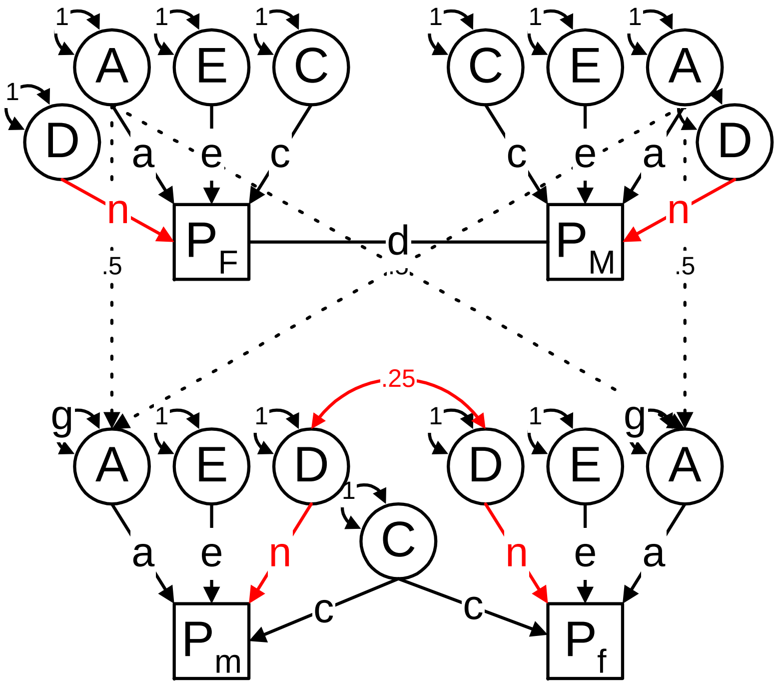
Dominance Model

- Dominance
- Common environment
    - ☐ Non-parental
- Assortment



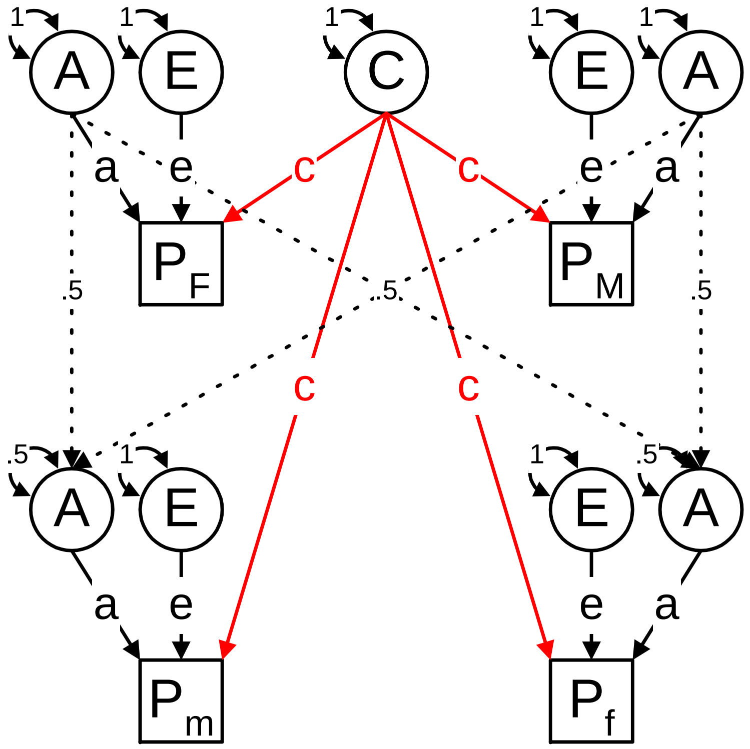
Dominance Expectations

- rSP= d
- rPO= $.5a^2$
- rMZ= $a^2+c^2+n^2$
- rDZ= $.5a^2+c^2+.25n^2$
- Var= $a^2+c^2+e^2+n^2$



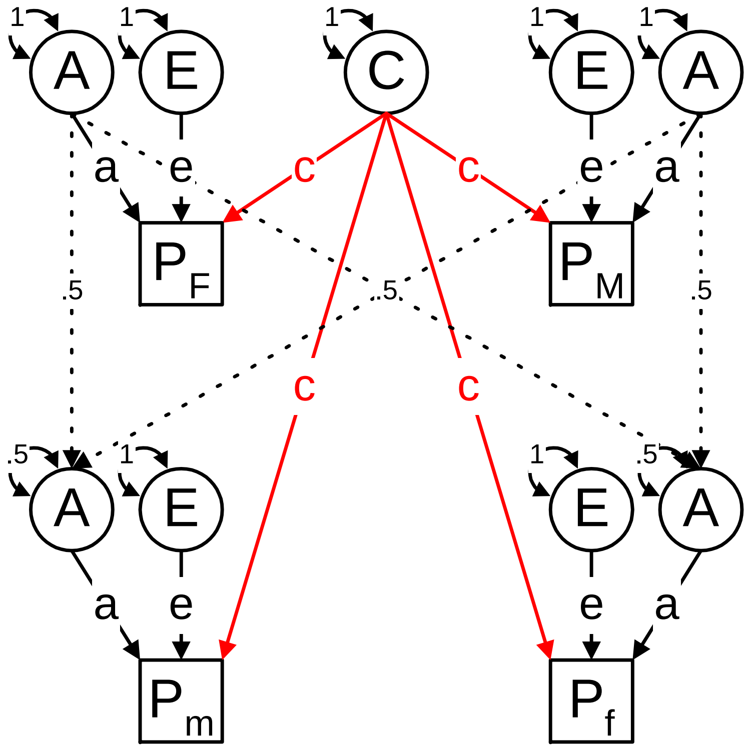
Common Environment Model

- Common environment
  - Same for all family members
- Assortment
  - Function of common environment



Common Environment Expectations

- rSP= c^2
- rPO= $.5a^2 + c^2$
- rMZ= a^2+c^2
- rDZ= $.5a^2+c^2$
- Var= $a^2+c^2+e^2$

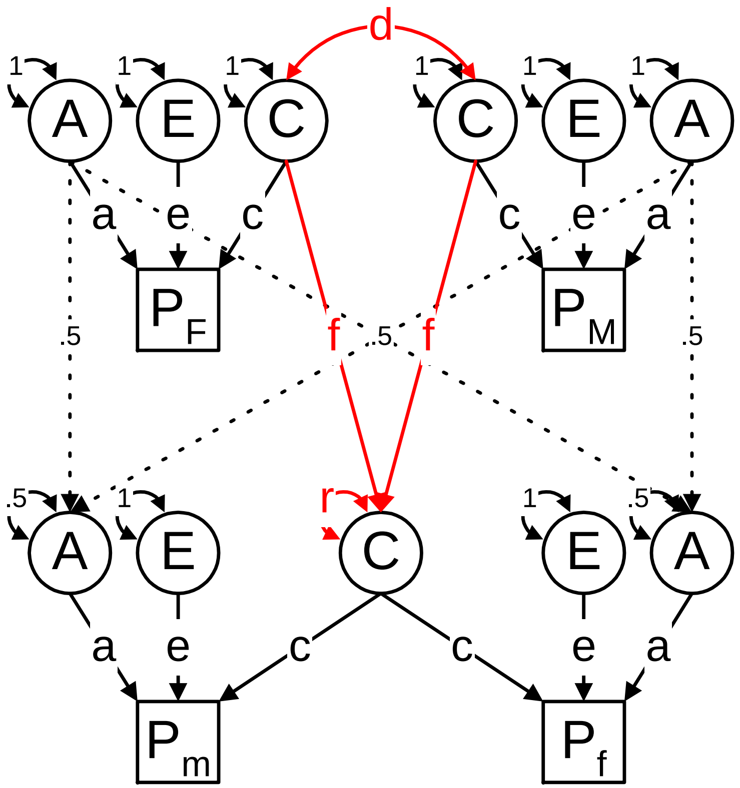


Model Assumptions IIr

- But if we make simplifying assumptions, we can estimate C & D at same time using twins and parents?
- Solve the following three equations for VA, VC, and VD:
  - $rMZ = VA + VD + VC$
  - $rDZ = 1/2VA + 1/4VD + VC$
  - $rPO = 1/2VA + \quad + VC$

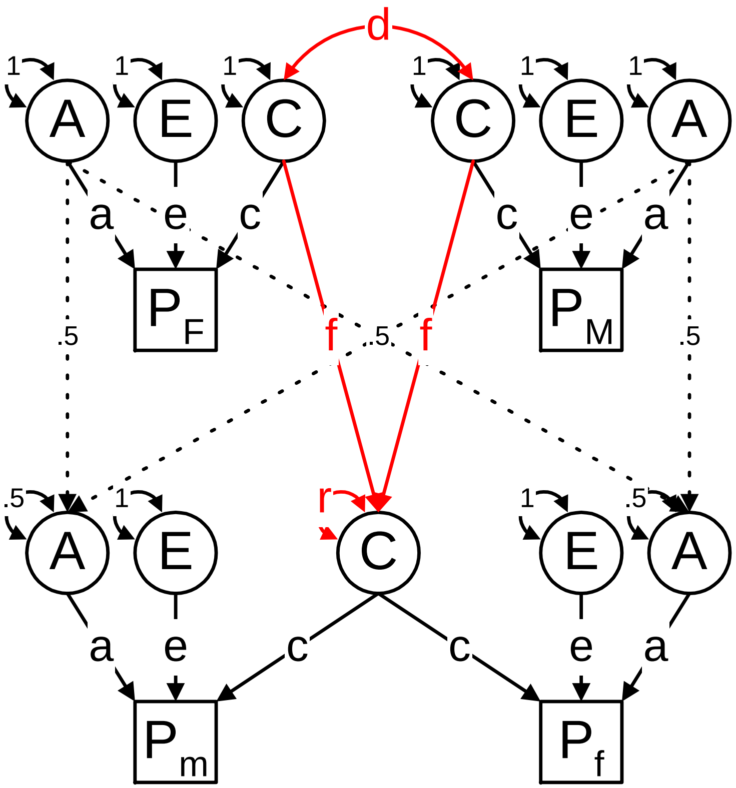
Social Homogamy Model

- Assortment
  - Social
- Cultural Transmission
  - From C to C
- Non-parental Shared Environment
  - Residual



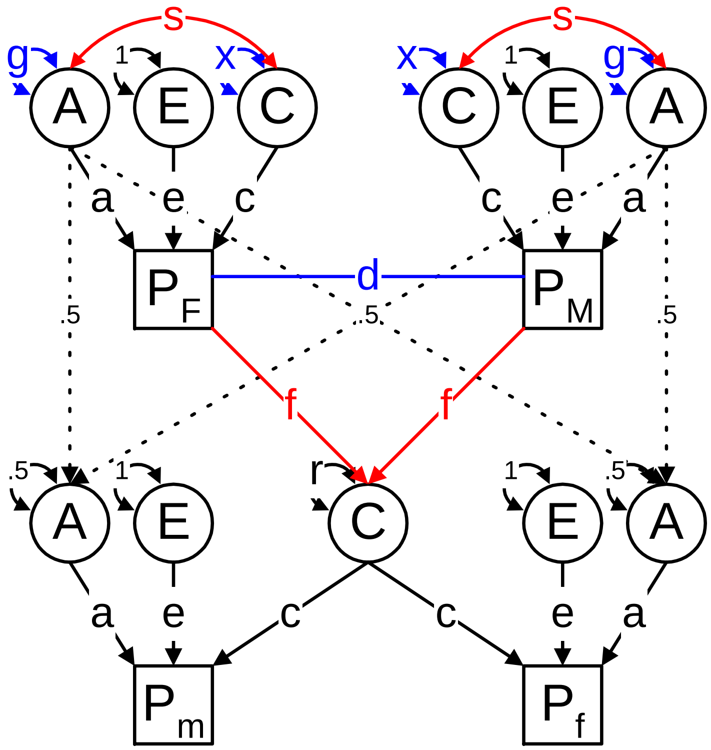
Social Homogamy Expectations

- $rSP = dc^2$
- $rPO = (f+df)c^2 + .5a^2$
- $x = 2f^2 + 2df^2 + r$
- $rMZ = a^2 + xc^2$
- $rDZ = .5a^2 + xc^2;$
- $Var = a^2 + xc^2 + e^2$



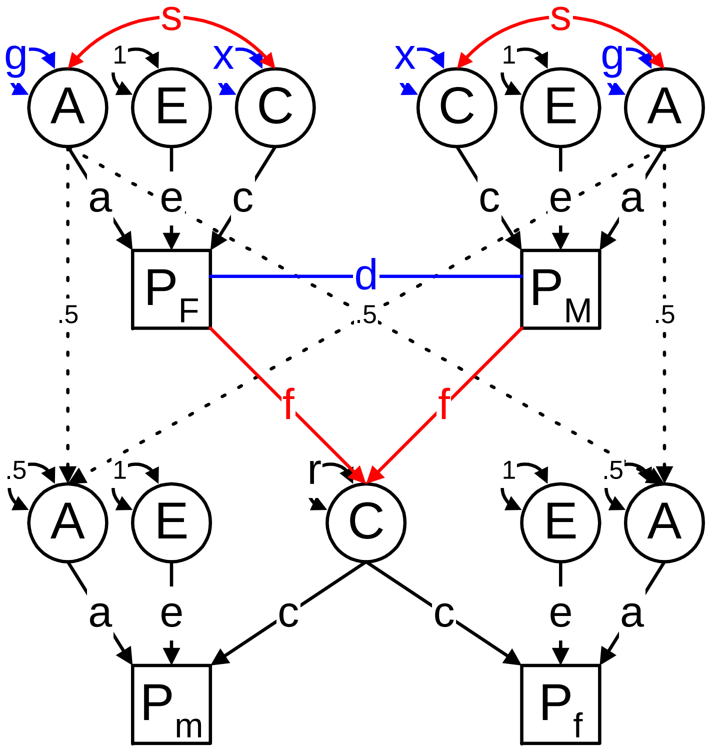
Phenotypic Assortment Model

- Assortment
  - Phenotypic
- Cultural Transmission
  - From P to C
- Non-parental Shared Environment
  - Residual
- Genotype-Environment Covariance

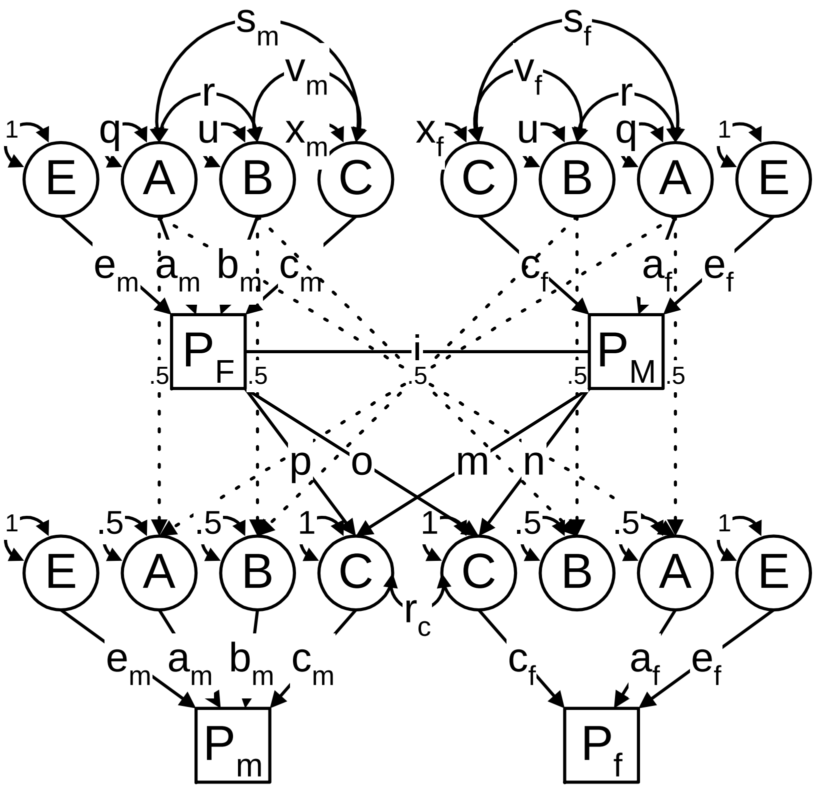


Phenotypic Assortment Expectations

- rSP= w= pdp
    - rGP= t= ga+sc
- rPO= (pf+wf)c+.5a(1+pd)t
    - rGE= j= asc+csa
- rMZ= ga2+xc2+j
    - pa= y= g+.5(t(d+d)t)
- rDZ= .5ya2+xc2+j
- + constraints

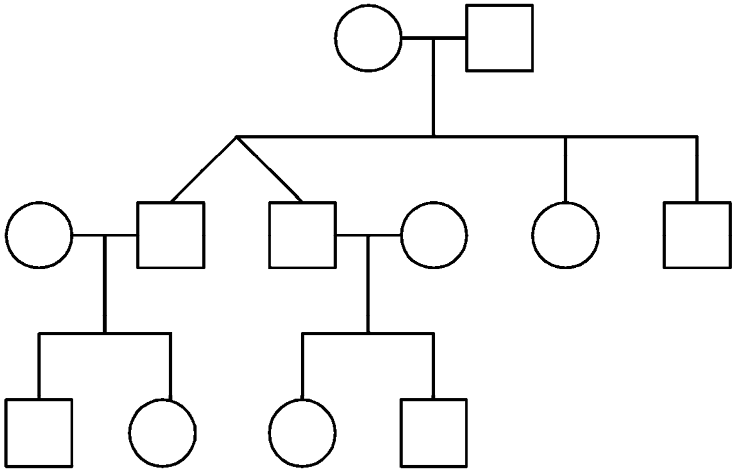


Sex Differences



ET (Stealth/Cascade) Design

- Extended Twin Kinships (ET model):
    - twins, their parents
    - siblings, spouses
    - and children

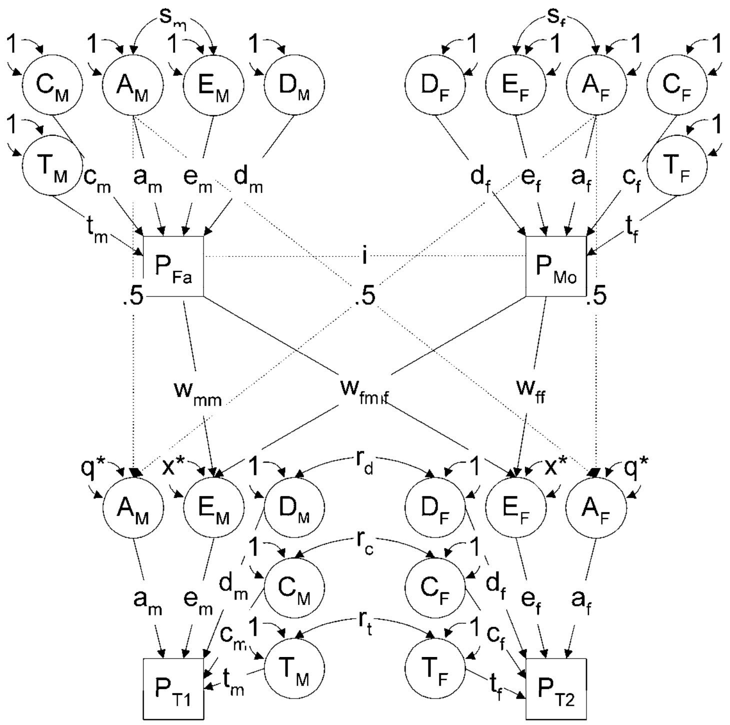


- 88 unique sex-specific biological/social relationships

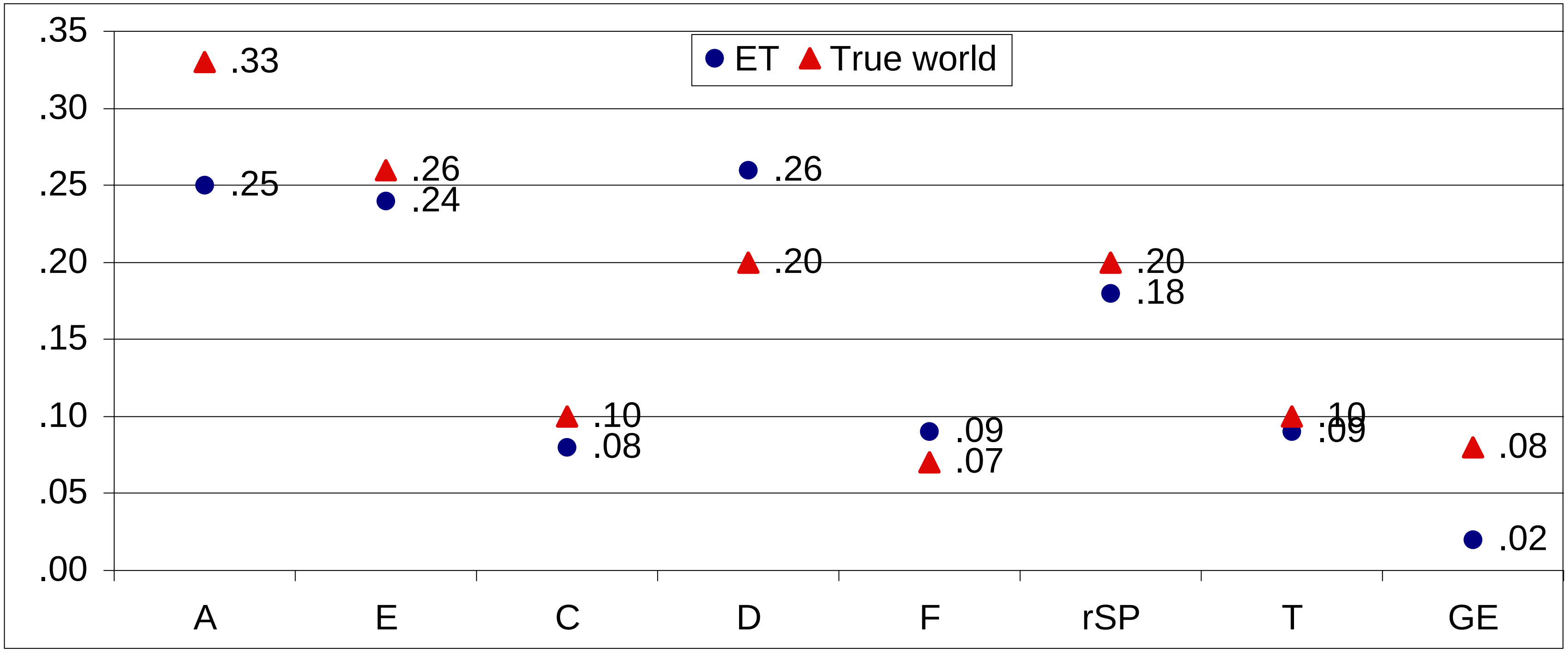
ET Model

(Eaves et al 1999, Maes et al 1999, 2006)

- Genetic
  - Additive (A)
  - Dominance (D)
- Environment
  - Specific (E)
  - Shared/Common (C)
  - Twin (T)
  - Cultural transmission (w)
- GE covariance (s)
- Assortative mating (i)



ET vs True World



Environmental Factors

- rMZ < 1 >> specific environmental factors

- rSIB = rDZ = rPO >> shared environmental factors
  - increase similarity between people living or having grown up in same home (first-degree relatives and MZs)
- rSIB > rPO >> non-parental environmental factors
  - in common for siblings, such as school environment, peers, friends
- rTWIN > rSIB >> special twin environment
  - additional twin similarity due to greater sharing of aspects of environment
- rSIB = rPO > 1/2 rMZ >> cultural transmission

Genetic Factors

- rMZ > first-degree relatives (rDZ, rSIB, rPO) > second-degree relatives (grandparents, half-siblings, avuncular pairs) > more distant relatives such as cousins >> additive genetic factors
- rSIB & rDZ < 1/2 rMZ (expectation: DZ=1/4MZ) and zero correlations for other pairs of relatives >> dominance
- phenotypic cultural transmission + genetic transmission >> GE covariance
- partner selection is based on phenotype >> non-random mating: source of similarity which may have both genetic and environmental implications

The 22th International Statistical Genetics Methods Workshop

August 11 - 15, 2008

Leuven, Belgium