# Gene-Environment Correlation (and combined PGI + Twin model)

DANIEL GUSTAVSON

MARCH 7, 2024

INTERNATIONAL STATISTICAL GENETICS WORKSHOP

# Genes and Environments are not Independent

Gene-Environment Interaction (Tuesday)

- Impact of the environment depends on genetic risk

Gene-Environment Correlation

- Genetic propensities are correlated with environmental niches
- “Genetic Nurture” (parental genotype influences offspring outcomes by shaping the environments parents provide; e.g., Kong, et al. 2018)

# Gene Environment Correlation (rGE)

Passive rGE

- Parents pass on genes and also provide an environment correlated with those genes, both of which influence the child's development

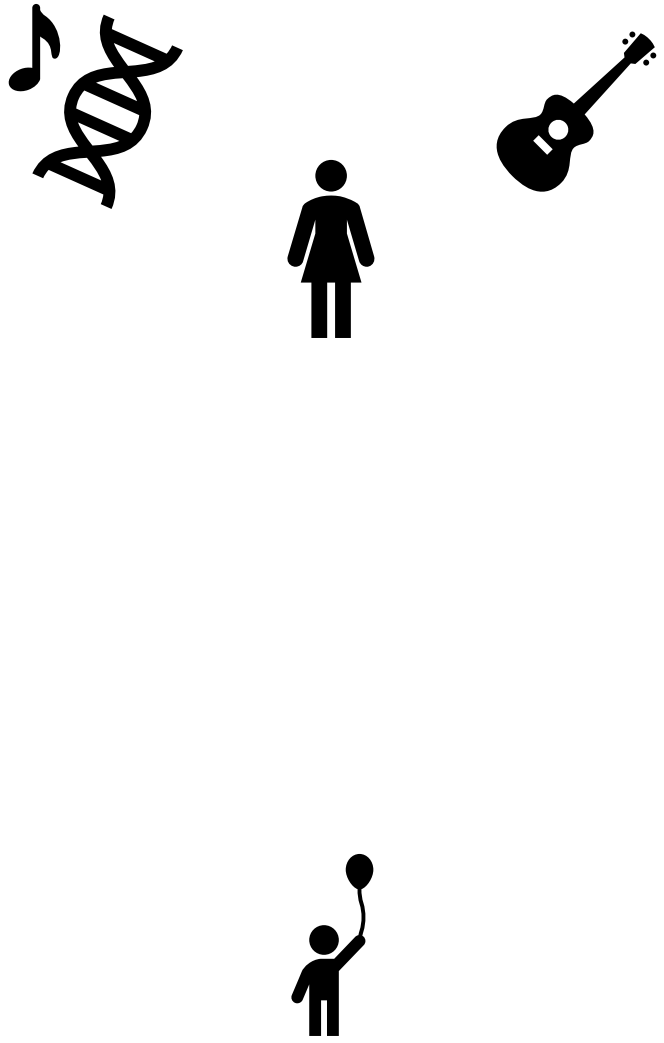
Active rGE

- A child's genetic background influences their choice of environment

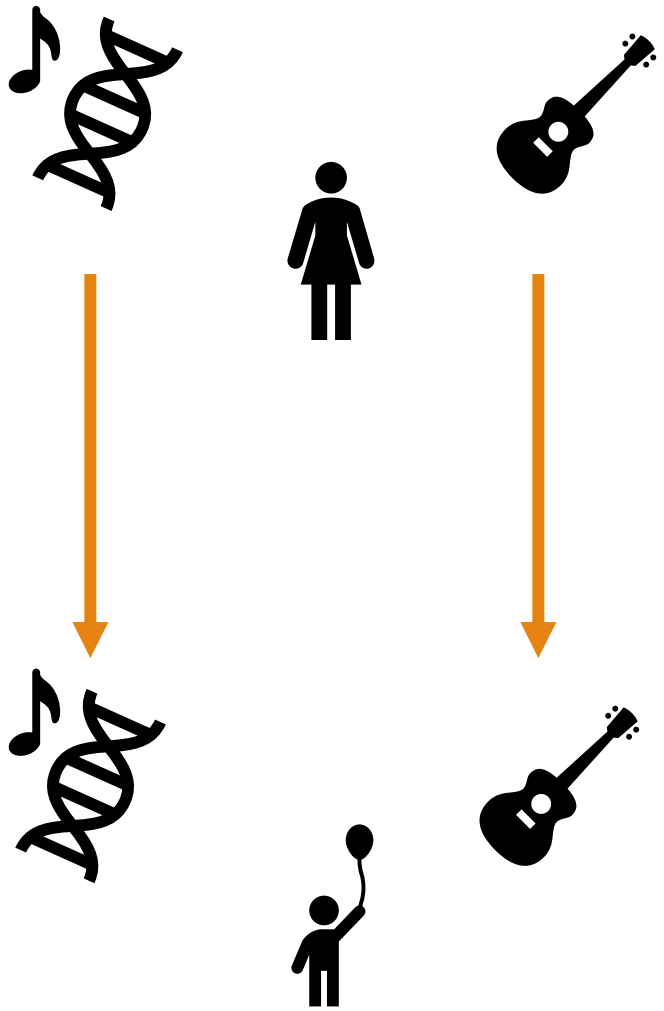
Evocative/reactive rGE

- A child's genetic background traits influences the reaction of others (and the environment provided by them)

Passive rGE



Passive rGE



# Consequences of unmodeled rGE

Passive rGE

- Upwardly biases shared environmental influences (e.g., ⇧ rMZ & ⇧ rDZ)

# Gene Environment Correlation (rGE)

Passive rGE

- Parents pass on genes and also provide an environment, both of which influence the child's development

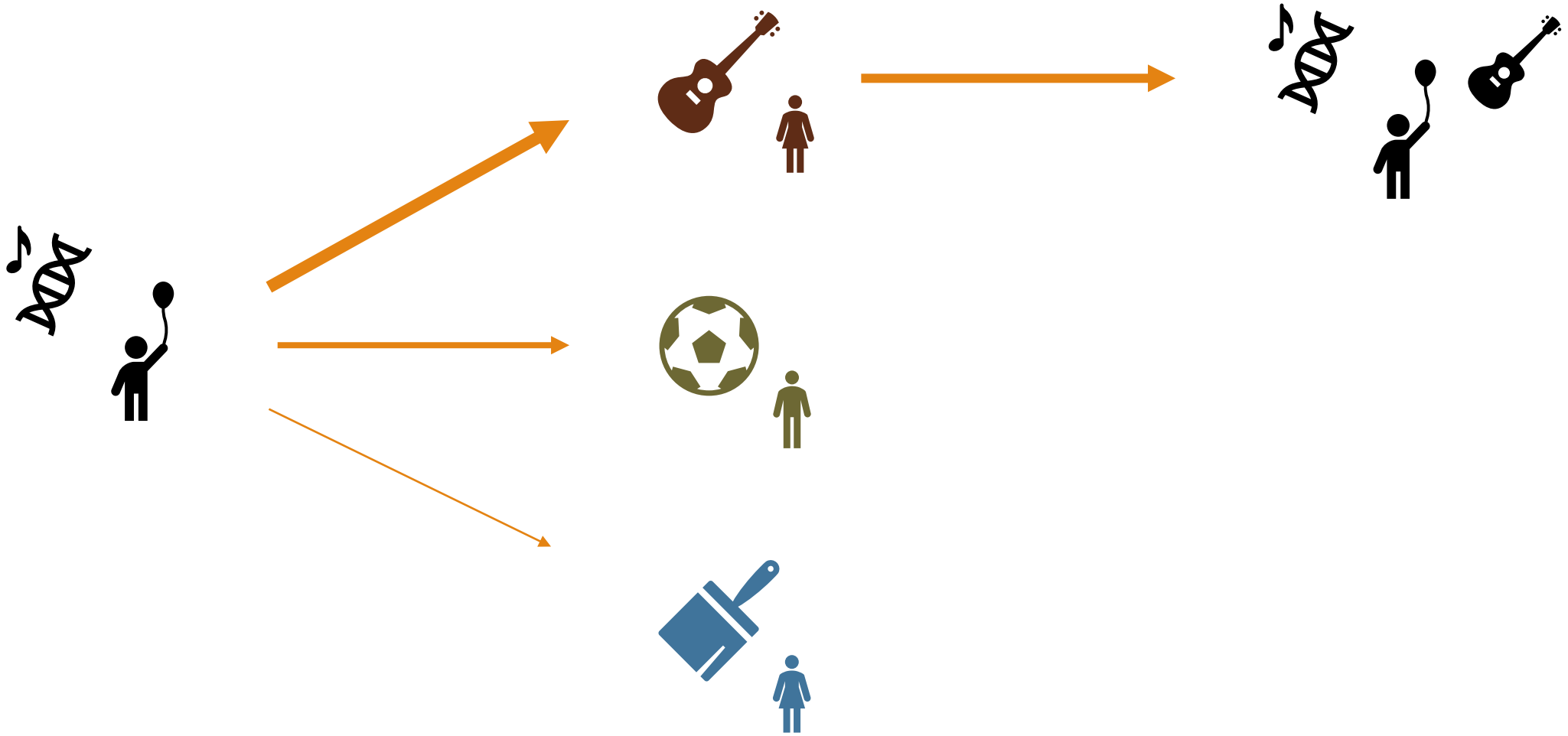
Active rGE

- A child's genetic background influences their choice of environment

Evocative/reactive rGE

- A child's genetic background traits influences the reaction of others (and the environment provided by them)

Active rGE



# Gene Environment Correlation (rGE)

Passive rGE

- Parents pass on genes and also provide an environment, both of which influence the child's development

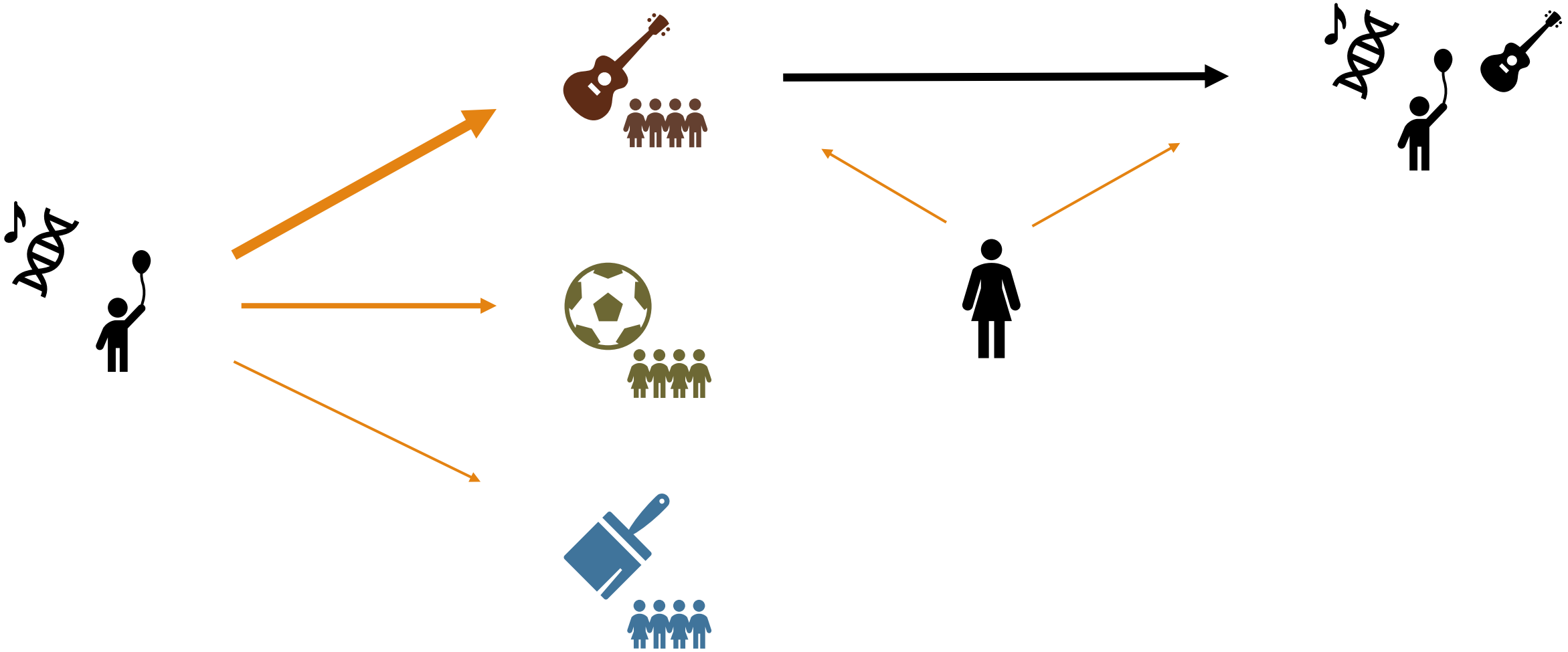
Active rGE

- A child's genetic background influences their choice of environment

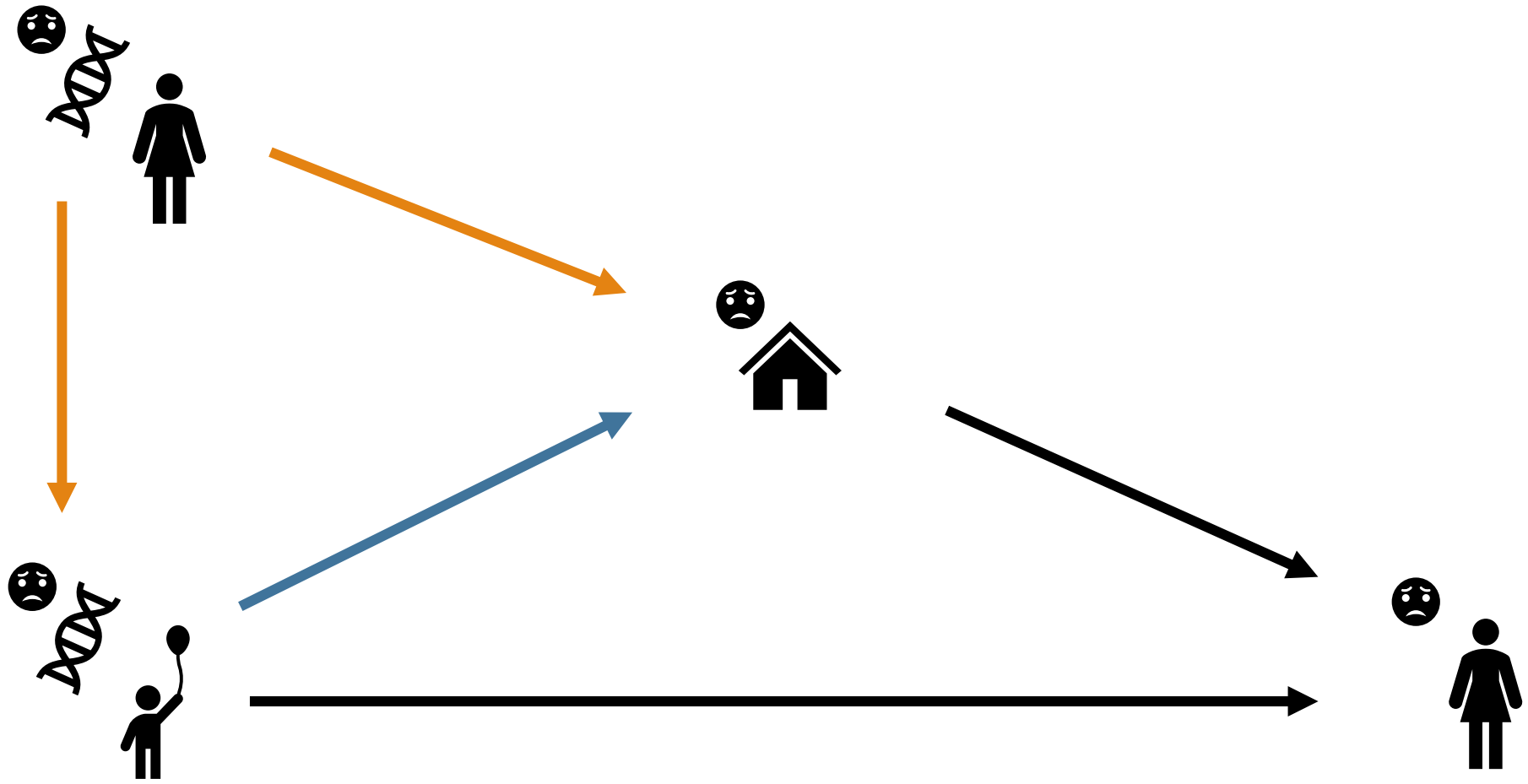
Evocative/reactive rGE

- A child's genetic background influences the reaction of others (and the environment provided by them)

Evocative rGE



## Passive vs. Evocative rGE



# Consequences of unmodeled rGE

Passive rGE (positive)

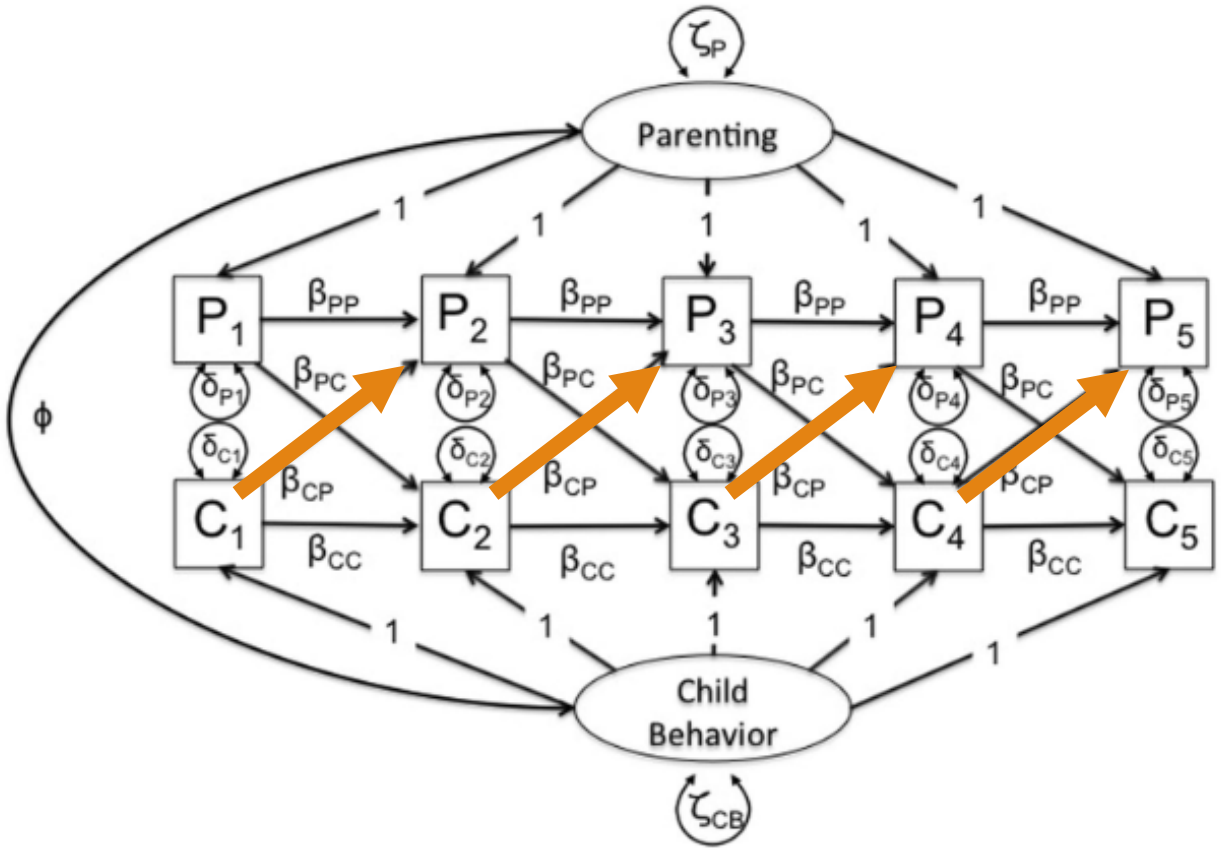
- Upwardly biases shared environmental influences on parenting (e.g., ⇧ rMZ & ⇧ rDZ)
- Downwardly biases genetic influences
- As A or C approach zero (and/or E increases), the impact of bias is minimized

Active or Evocative rGE (positive)

- Upwardly biases genetic influences on parenting (e.g., ⇧ rMZ & ⇧ rDZ)
- As A or C approach zero (and/or E increases), the impact of bias is minimized

Neiderhiser et al. (2004) *Developmental Psychology*
Verhulst & Hatemi (2013) *Political Analysis*

# How do we identify active rGE?



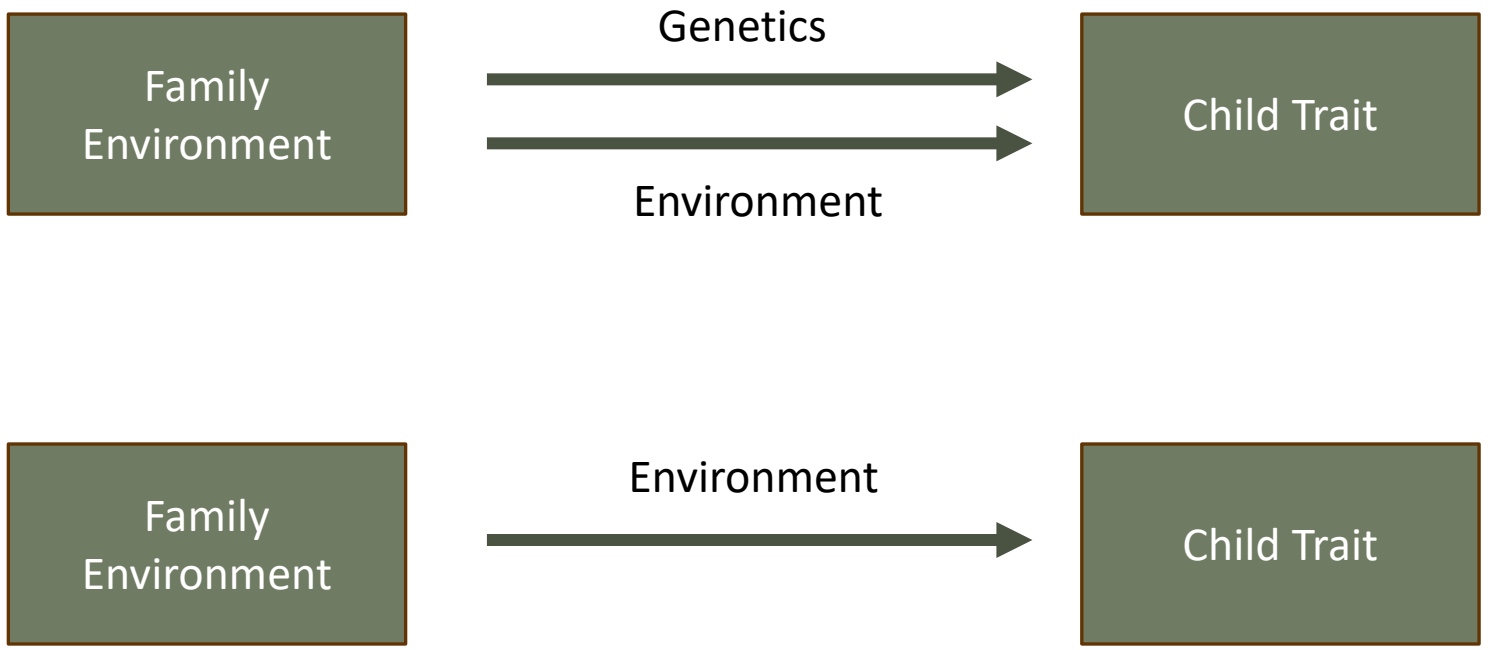

Moscati et al. (2016) *Behavior Genetics*

# How do we identify passive rGE?

Adoption Design

# How do we identify passive rGE?

Adoption Design



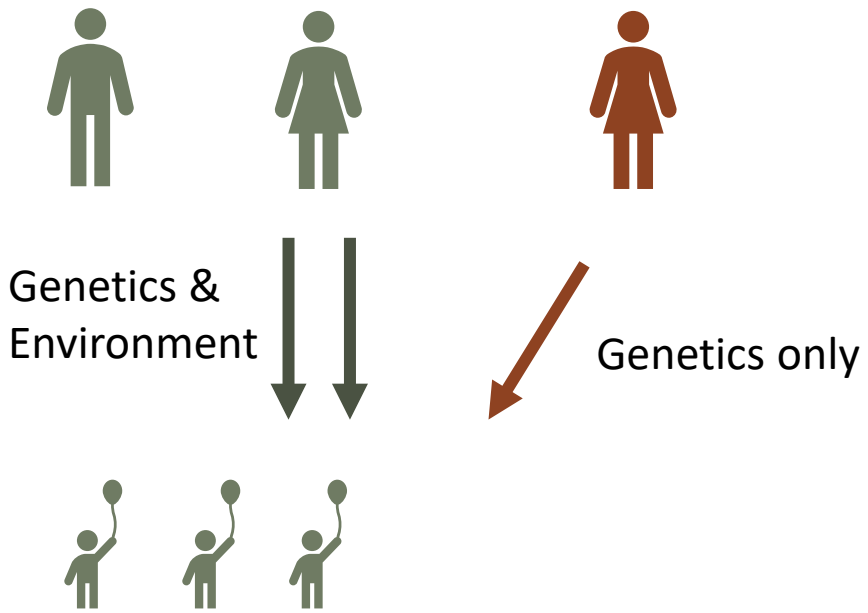

Knopik et al. (2017) *Behavioral Genetics (7th ed)*

# How do we identify passive rGE?

Adoption Design

Children of Twins Design

# Child of Twins Design



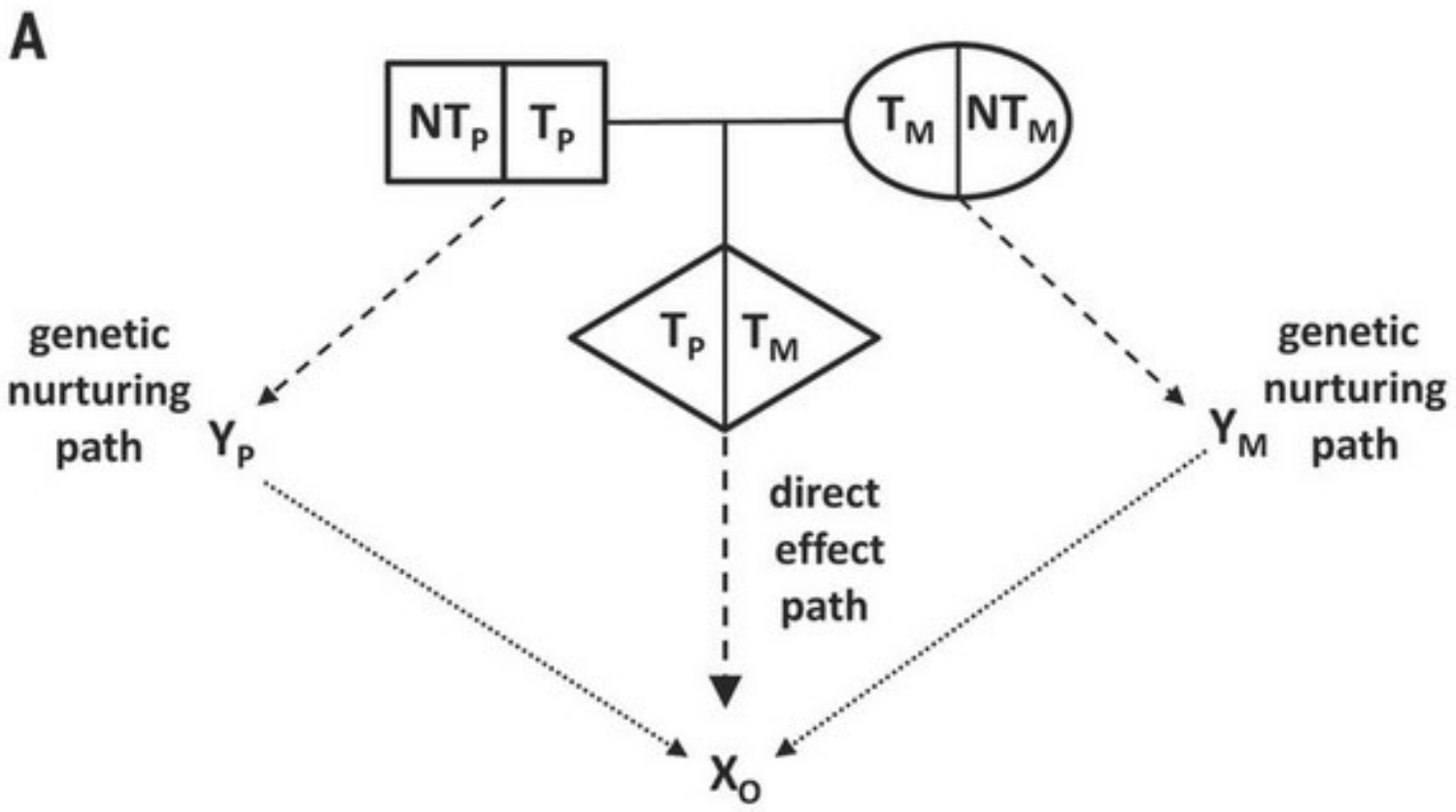

For example, see
Silberg, Maes, & Eaves (2010)
*Journal of Child Psychology and Psychiatry*

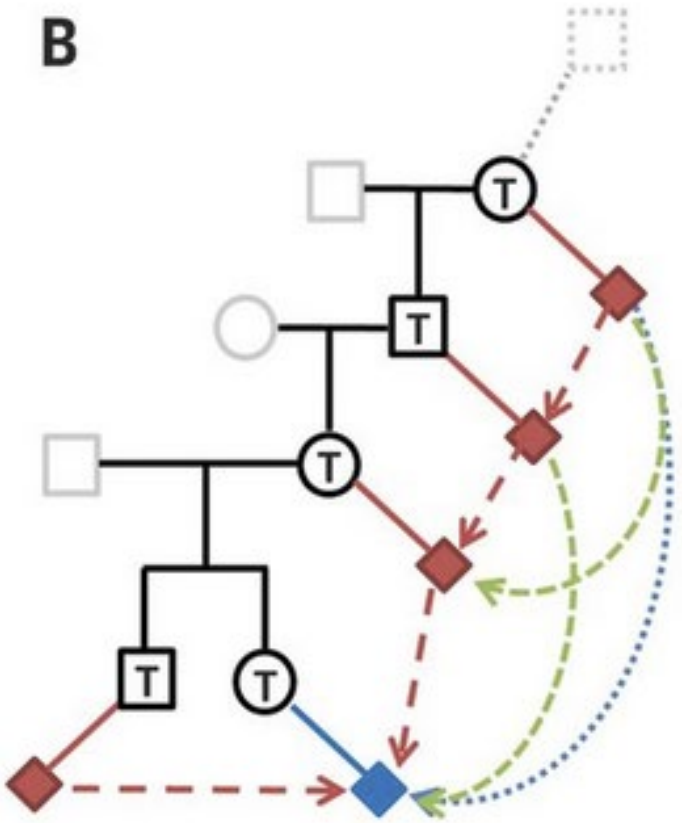
# How do we identify passive rGE?

Adoption Design

Children of Twins Design

Genetic Nurture




Kong et al. (2018) *Science*

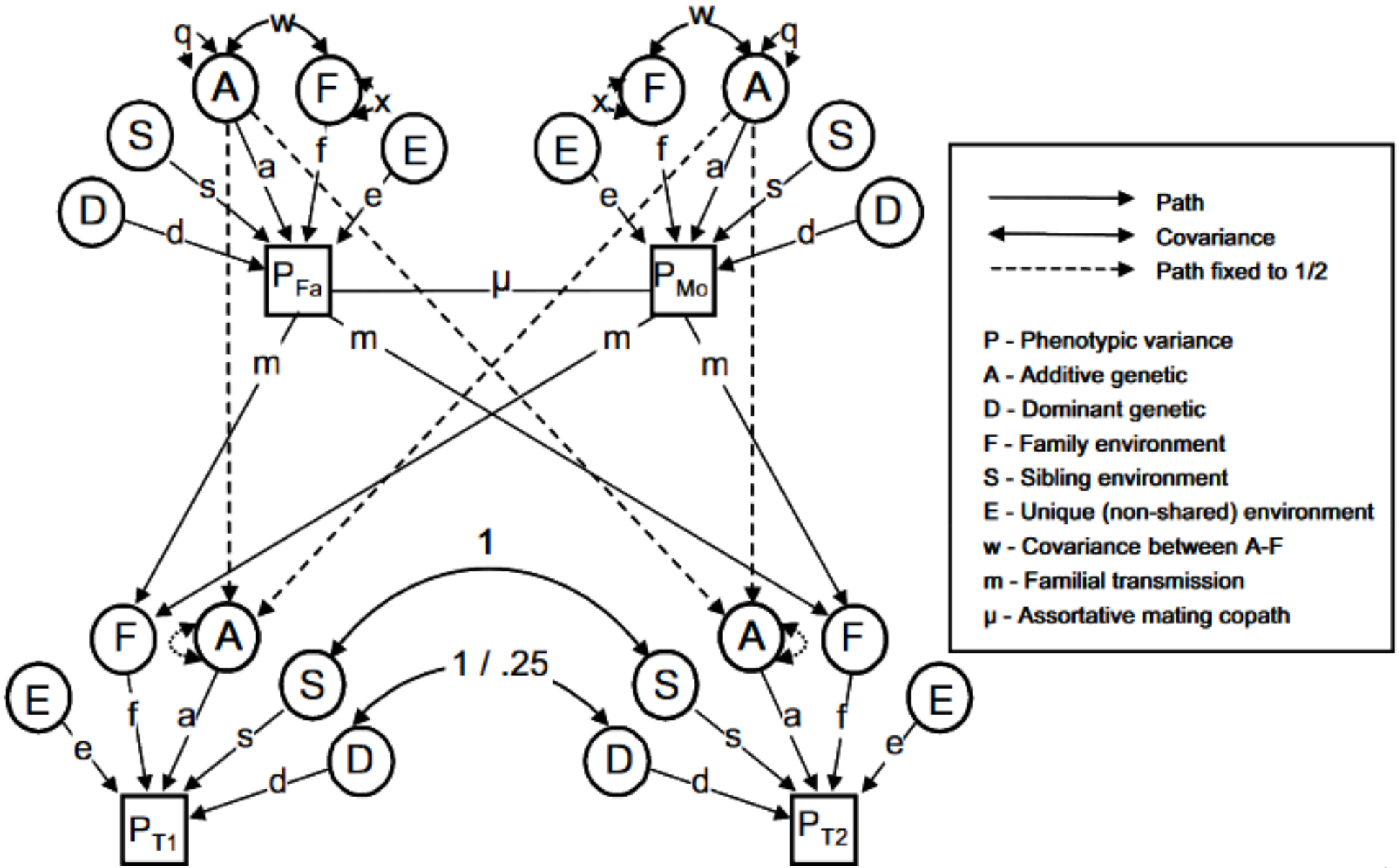
# How do we identify passive rGE?

Adoption Design

Children of Twins Design

Genetic Nurture

Extended Twin Models / Cascade Model (next session)




Keller et al. (2009)
*Twin Research & Human Genetics*

# How do we identify passive rGE?

Adoption Design

Children of Twins Design

Genetic Nurture

Extended Twin Models / Cascade Model (next session)

Twin + PGI Approach (Dolan et al. 2021; Bruins, et al. 2023)

Behavior Genetics (2021) 51:237–249
https://doi.org/10.1007/s10519-020-10035-7

ORIGINAL RESEARCH

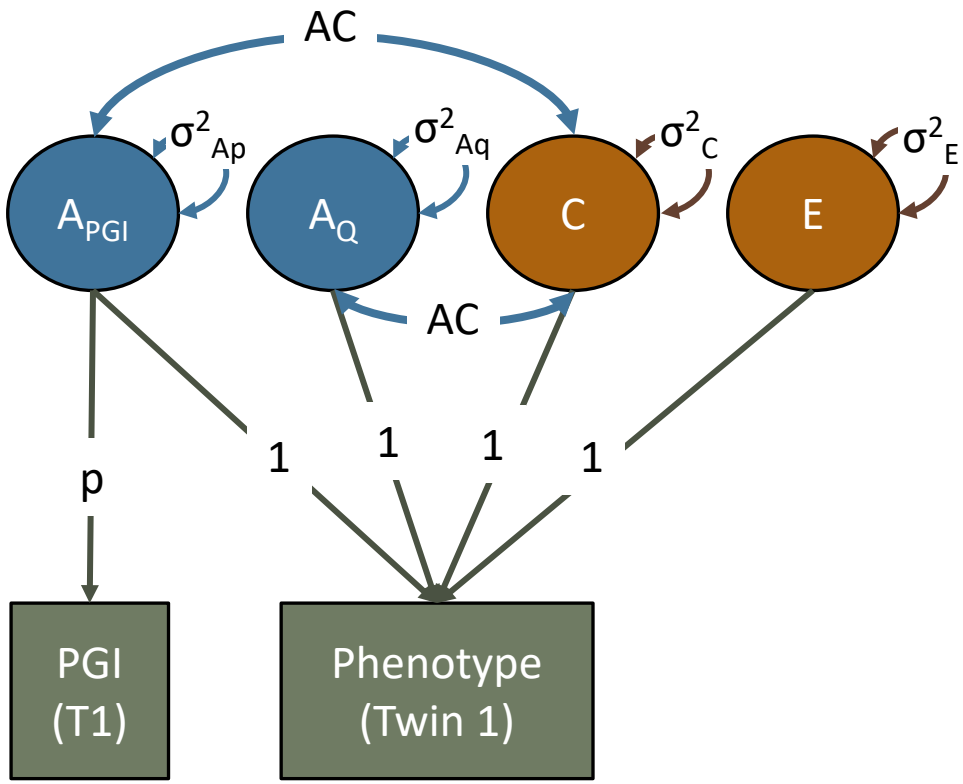
# Incorporating Polygenic Risk Scores in the ACE Twin Model to Estimate A–C Covariance

Conor V. Dolan[1,4] · Roel C. A. Huijskens[1] · Camelia C. Minică[3,5] · Michael C. Neale[1,2] · Dorret I. Boomsma[1]

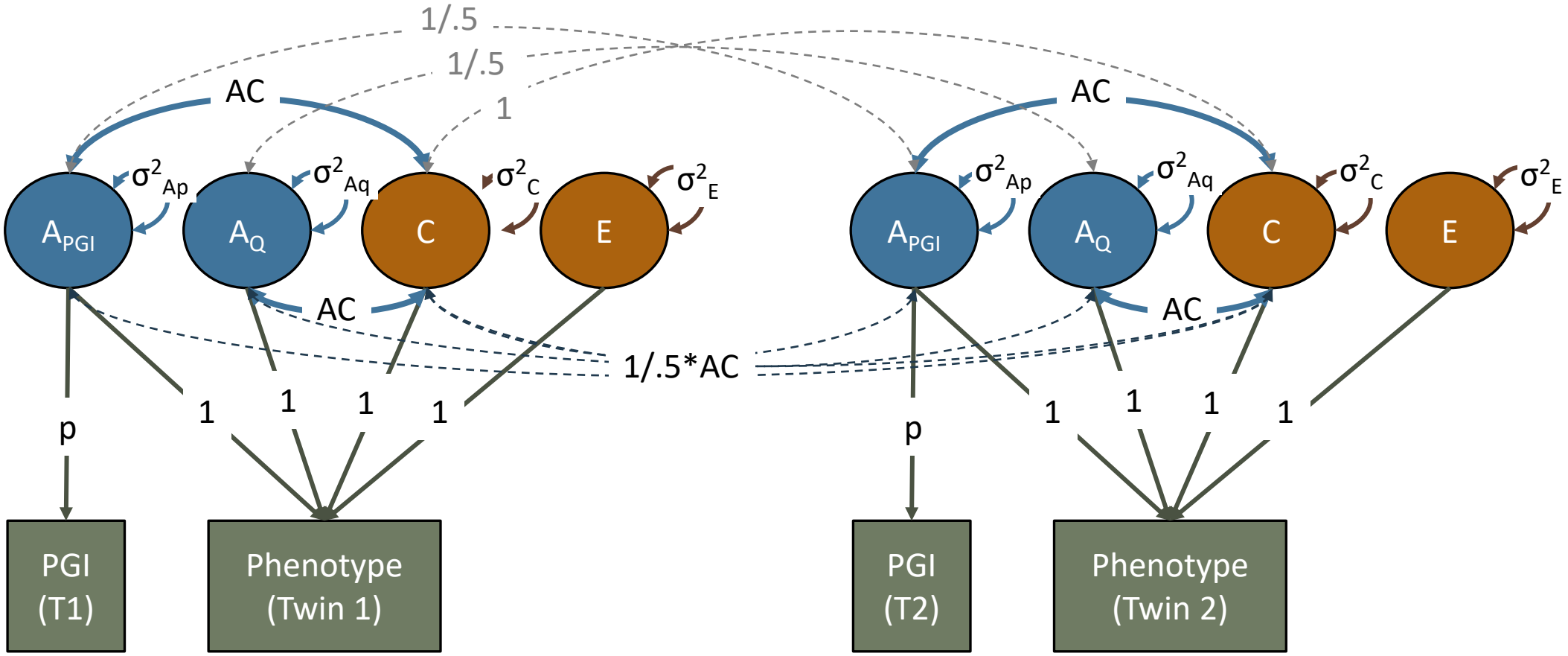
Received: 25 June 2020 / Accepted: 5 December 2020 / Published online: 1 February 2021


## Abstract

The assumption in the twin model that genotypic and environmental variables are uncorrelated is primarily made to ensure parameter identification, not because researchers necessarily think that these variables are uncorrelated. Although the biasing effects of such correlations are well understood, a method to estimate these parameters in the twin model would be useful. Here we explore the possibility of relaxing this assumption by adding polygenic scores to the (univariate) twin model. We demonstrate that this extension renders the additive genetic (A)—common environmental (C) covariance ($\sigma_{AC}$) identified. We study the statistical power to reject $\sigma_{AC}=0$ in the ACE model and present the results of simulations.




Dolan et al. (2021) *Behavior Genetics*




Dolan et al. (2021) *Behavior Genetics*

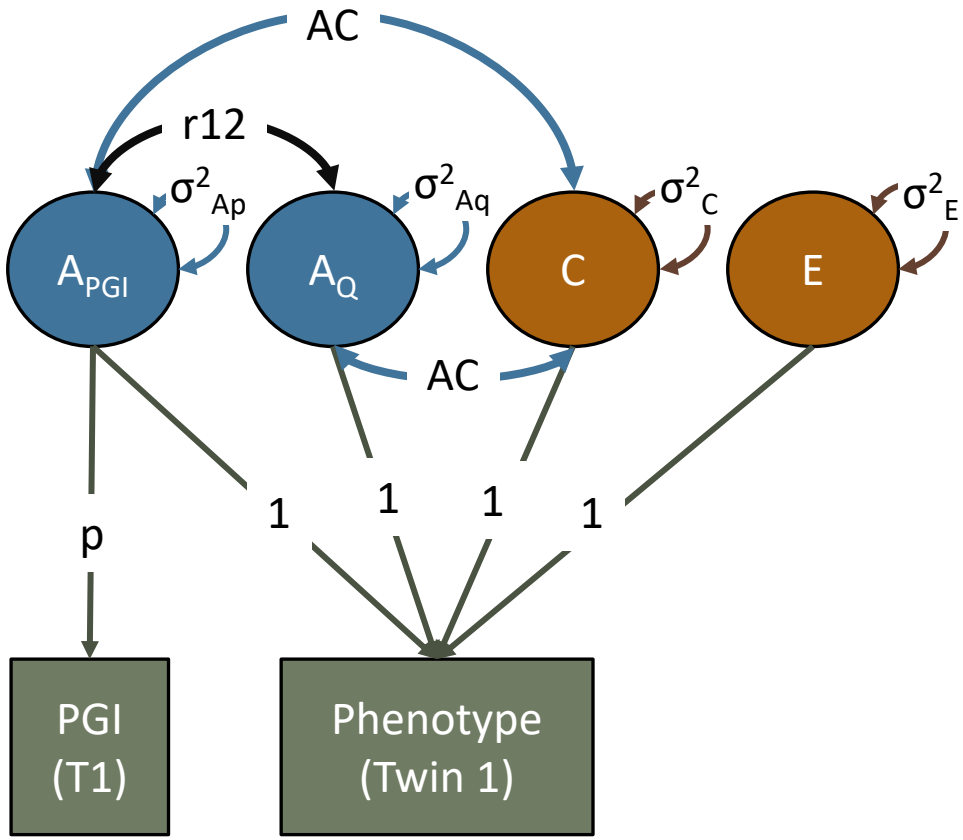
# Assumptions and Discussion

$A_{PGI}$ is representative subset of $A_Q$
- PGIs are effectively capturing same information as the remaining heritability, but imperfectly
- But is this really true? $A_{PGI}$ could never have de novo mutation or rare variants

$A_{PGI}$ is correlated with C to the same extend that $A_Q$ is
- What happens if this is not the case?
- By assumption we could only fit one of these two options, but by toggling between them we could try both

$A_{PGI}$ and $A_Q$ are both nonzero

$A_{PGI}$ is not correlated with $A_Q$


AC
r12
σ²Ap
σ²Aq
σ²C
σ²E
APGI
AQ
C
E
AC
p
1
1
1
1
PGI
(T1)
Phenotype
(Twin 1)

MULTIVARIATE BEHAVIORAL RESEARCH
https://doi.org/10.1080/00273171.2023.2228763



# Environment-by-PGS Interaction in the Classical Twin Design: An Application to Childhood Anxiety and Negative Affect

Susanne Bruins[a,b], Jouke-Jan Hottenga[a], Michael C. Neale[a,c], René Pool[a,b], Dorret I. Boomsma[a,b,d], and Conor V. Dolan[a]

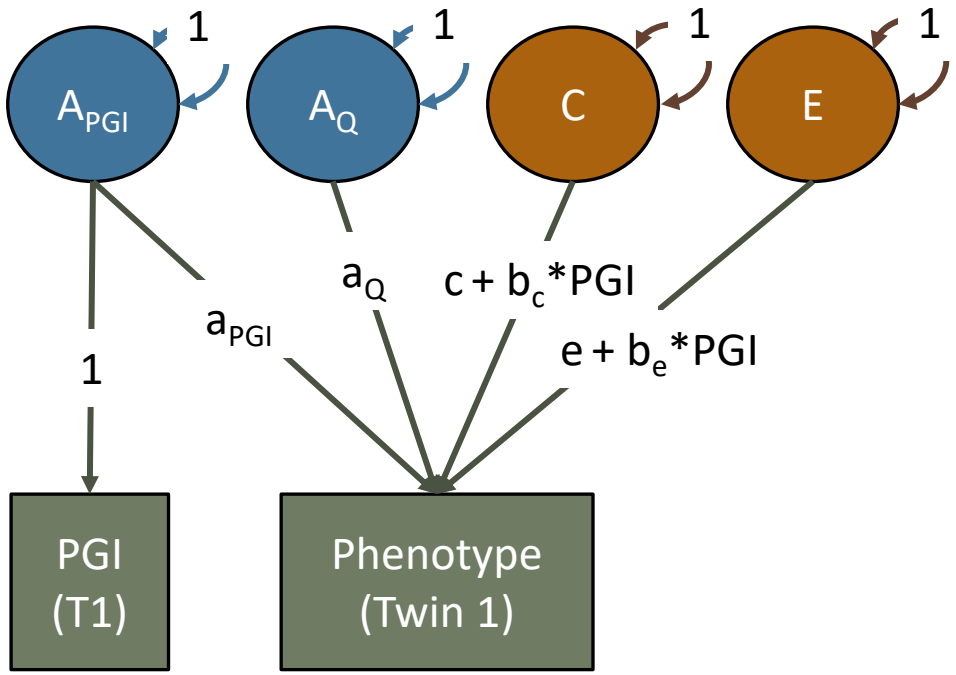
[a]Department of Biological Psychology, Vrije Universiteit; [b]Amsterdam Public Health research institute; [c]Virginia Institute for Psychiatric and Behavioral Genetics, Virginia Commonwealth University; [d]Amsterdam Reproduction and Development research institute

**ABSTRACT**

One type of genotype-environment interaction occurs when genetic effects on a phenotype are moderated by an environment; or when environmental effects on a phenotype are moderated by genes. Here we outline these types of genotype-environment interaction models, and propose a test of genotype-environment interaction based on the classical twin design, which includes observed genetic variables (polygenic scores: PGSs) that account for part of the genetic variance of the phenotype. We introduce environment-by-PGS interaction and the results of a simulation study to address statistical power and parameter recovery. Next, we apply the model to empirical data on anxiety and negative affect in children. The power to detect environment-by-PGS interaction depends on the heritability of the phenotype, and the strength of the PGS. The simulation results indicate that under realistic conditions of sample size, heritability and strength of the interaction, the environment-by-PGS model is a viable approach to detect genotype-environment interaction. In 7-year-old children, we defined two PGS based on the largest genetic association studies for 2 traits that are genetically correlated to childhood anxiety and negative affect, namely major depression (MDD) and intelligence (IQ). We find that common environmental influences on negative affect are amplified for children with a lower IQ-PGS.

**KEYWORDS**

Genotype-environment (g × e) interaction; moderation; mono- and dizygotic twins; polygenic scores (pgs); genetic covariance structure modeling




Bruins et al. (2023) *Multivariate Behavioral Research*

# Polygenic Indices (PGIs/PGSs)

Quantification of genetic liability for a trait

- In the context of a disease, it may capture risk due to genes (i.e., polygenic risk score; PRS)

Analytical Concerns:

- Focus analyses on a single ancestry at a time?
- Control for ancestry-based principal components
- Monozygotic twins are 100% identical and therefore have the same PGI

# PGI + Twin Practical

UNIX:

```
mkdir Twin_PGI/
cp -r /home/faculty/daneg/Twin_PGI/* Twin_PGI/
cd Twin_PGI/practical
```

Rstudio:

```
system("mkdir $HOME/Twin_PGI/")
system("cp -R /home/faculty/daneg/2024/Twin_PGI/* $HOME/Twin_PGI/")
setwd("~/Twin_PGI/")
```

https://qimr.az1.qualtrics.com/jfe/form/SV_dpy2LRErlNs6Jmu

# Practical considerations

| Parameter | Full Model | No AC Covariance |
|---|---|---|
| $A_{PGI}$ | 16% | 23% |
| $A_Q$ | 24% | 25% |
| C | 24% | 35% |
| E | 16% | 17% |
| AC Covariance | 19% | - |

Information Criteria:

| df Penalty | Parameters Penalty | Sample-Size Adjusted

AIC: -990.4557 13009.54 13009.62

BIC: -40151.9657 13054.35 13028.94

CFI: NA

TLI: 1 (also known as NNFI)

RMSEA: 0 [95% CI (NA, NA)]

Prob(RMSEA <= 0.05): NA

To get additional fit indices, see help(mxRefModels)

```
# Fit model normally
ACModelFit <- mxTryHard(ACModel, extraTries = 200)
summary(ACModelFit)

## Fit reference models to get CFI etc.
factorSat <- mxRefModels(ACModelFit, run=T)
#summary(factorSat$Independence)
#summary(factorSat$Saturated)

summary(ACModelFit, refModels=factorSat)
```

chi-square: $\chi^2$ ( df=15 ) = -0.0003213478, p = 1
Information Criteria:

| | df Penalty | Parameters Penalty | Sample-Size Adjusted |
|---|---|---|---|
| AIC: | -990.4557 | 13009.54 | 13009.62 |
| BIC: | -40151.9657 | 13054.35 | 13028.94 |

CFI: 1.005239
TLI: 1.003144 (also known as NNFI)
RMSEA: 0 *(Non-centrality parameter is negative) [95% CI (0, 0)]
Prob(RMSEA <= 0.05): 1

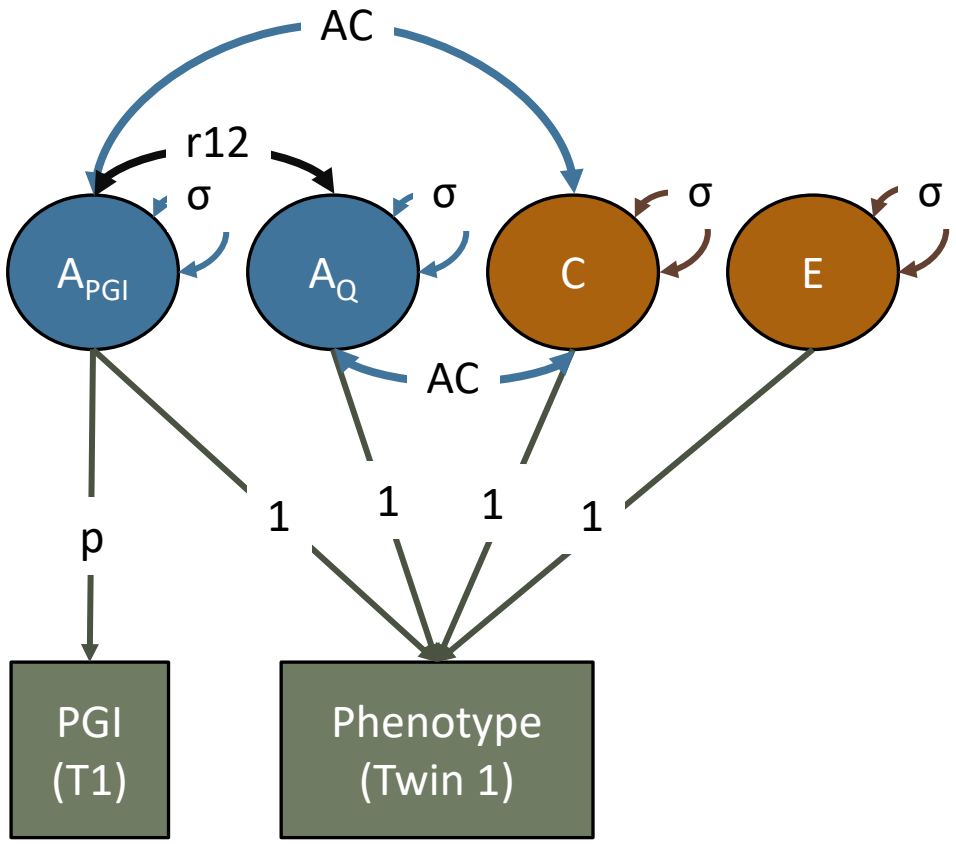

AC
r12
σ
σ
σ
σ
$A_{PGI}$
$A_Q$
C
E
AC
p
1
1
1
1
PGI
(T1)
Phenotype
(Twin 1)

**Table 1** Statistical power to reject the null hypothesis that A–C covariance is zero (alpha = 0.05)

| | *pr*PRS | *pr*Ph | $\sigma^2_A$ | $\sigma^2_C$ | $\sigma^2_E$ | $\sigma_{AC}$ | $r_{AC}$ | $\sigma^2_{Ph}$ | $pr2\sigma_{AC}$ | Power | N (0.80) |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 0.2 | 0.054 | 0.3 (0.27) | 0.2 (0.18) | 0.5 (0.45) | 0.048 | 0.2 | 1.09 | 0.088 | 0.216 | 11,420 |
| 2 | 0.2 | 0.052 | 0.3 (0.26) | 0.2 (0.17) | 0.5 (0.43) | 0.073 | 0.3 | 1.14 | 0.128 | 0.415 | 5158 |
| 3 | 0.2 | 0.053 | 0.3 (0.27) | 0.3 (0.26) | 0.4 (0.36) | 0.060 | 0.2 | 1.12 | 0.107 | 0.323 | 6971 |
| 4 | 0.2 | 0.050 | 0.3 (0.25) | 0.3 (0.25) | 0.4 (0.34) | 0.090 | 0.3 | 1.18 | 0.153 | 0.606 | 3161 |
| 5 | 0.4 | 0.109 | 0.3 (0.27) | 0.2 (0.18) | 0.5 (0.45) | 0.048 | 0.2 | 1.09 | 0.088 | 0.402 | 5352 |
| 6 | 0.4 | 0.104 | 0.3 (0.26) | 0.2 (0.17) | 0.5 (0.43) | 0.073 | 0.3 | 1.14 | 0.128 | 0.724 | 2408 |
| 7 | 0.4 | 0.107 | 0.3 (0.27) | 0.3 (0.26) | 0.4 (0.36) | 0.060 | 0.2 | 1.12 | 0.107 | 0.595 | 3241 |
| 8 | 0.4 | 0.101 | 0.3 (0.25) | 0.3 (0.25) | 0.4 (0.34) | 0.090 | 0.3 | 1.18 | 0.153 | 0.906 | 1462 |
| 9 | 0.2 | 0.088 | 0.5 (0.44) | 0.2 (0.18) | 0.3 (0.27) | 0.063 | 0.2 | 1.12 | 0.112 | 0.238 | 10,107 |
| 10 | 0.2 | 0.084 | 0.5 (0.42) | 0.2 (0.17) | 0.3 (0.25) | 0.094 | 0.3 | 1.18 | 0.159 | 0.455 | 4594 |
| 11 | 0.2 | 0.086 | 0.5 (0.43) | 0.3 (0.26) | 0.2 (0.17) | 0.077 | 0.2 | 1.15 | 0.134 | 0.362 | 6084 |
| 12 | 0.2 | 0.081 | 0.5 (0.40) | 0.3 (0.24) | 0.2 (0.16) | 0.116 | 0.3 | 1.23 | 0.189 | 0.661 | 2784 |
| 13 | 0.4 | 0.177 | 0.5 (0.44) | 0.2 (0.18) | 0.3 (0.26) | 0.063 | 0.2 | 1.12 | 0.112 | 0.465 | 4481 |
| 14 | 0.4 | 0.168 | 0.5 (0.42) | 0.2 (0.17) | 0.3 (0.25) | 0.094 | 0.3 | 1.18 | 0.158 | 0.794 | 2028 |
| 15 | 0.4 | 0.173 | 0.5 (0.43) | 0.3 (0.26) | 0.2 (0.17) | 0.077 | 0.2 | 1.15 | 0.133 | 0.682 | 2650 |
| 16 | 0.4 | 0.162 | 0.5 (0.41) | 0.3 (0.24) | 0.2 (0.16) | 0.116 | 0.3 | 1.23 | 0.188 | 0.950 | 1206 |

Given $\sigma^2_A = \sigma^2_{Ap} + \sigma^2_{Aq}$, *pr*PRS equals $\sigma^2_{Ap}/\sigma^2_A$, i.e., the proportion of additive genetic variance attributable to the PRS, and *pr*Ph is the proportion of phenotypic variance attributable to the PRS, $\sigma^2_{Ap}/\sigma^2_{Ph}$; $r_{AC}$ and $\sigma_{AC}$ are the correlation and covariance of A and C, $\sigma^2_{Ph}$ is the phenotypic variance; $pr2*\sigma_{AC}$ is the proportion of phenotypic variance due to $2*\sigma_{AC}$

The standardized A, C, E variance components are given in parentheses. For instance, in setting 16, the raw variance is 0.5 + 0.3 + 0.2 + 0.116*2 = 1.23, and the standardized variance is 0.41 + 0.24 + 0.16 + 0.188 = ~

```
 mxAlgebra(SA1 + SA2 + 2*r12*sqrt(SA1*SA2), name="SA"),
 mxAlgebra(r12*sqrt(SA1)*sqrt(SA2), name="C12"),
 mxAlgebra( (SA1+r12*sqrt(SA1*SA2))*solve(SA), name="g1"),
 mxAlgebra( (SA2+r12*sqrt(SA1*SA2))*solve(SA), name="g2"),
# AC covariances
 mxAlgebra(g1*(CAC), name="CA1"),
 mxAlgebra(g2*(CAC), name="CA2"),
….

 mxAlgebra(expression=rbind(
#          E   A1   A2   C       E   A1    A2    C
 cbind(SE,   0,   0,   0,        0,   0,   0,      0),       # E
 cbind( 0, SA1, C12, CA1,   0, SA1, C12, CA1),    # A1
 cbind( 0, C12, SA2, CA2,   0, C12, SA2, CA2),    # A2
 cbind( 0, CA1, CA2,  SC,    0, CA1, CA2, SC),     # C

 cbind( 0,   0,   0,   0,         SE,   0,    0,    0),       # E
 cbind( 0, SA1, C12, CA1,    0, SA1,  C12,  CA1),  # A1
 cbind( 0, C12, SA2, CA2,    0, C12,  SA2,  CA2),  # A2
 cbind( 0, CA1, CA2,  SC,     0, CA1,  CA2,   SC)),  # C
name='latSmz'),
```

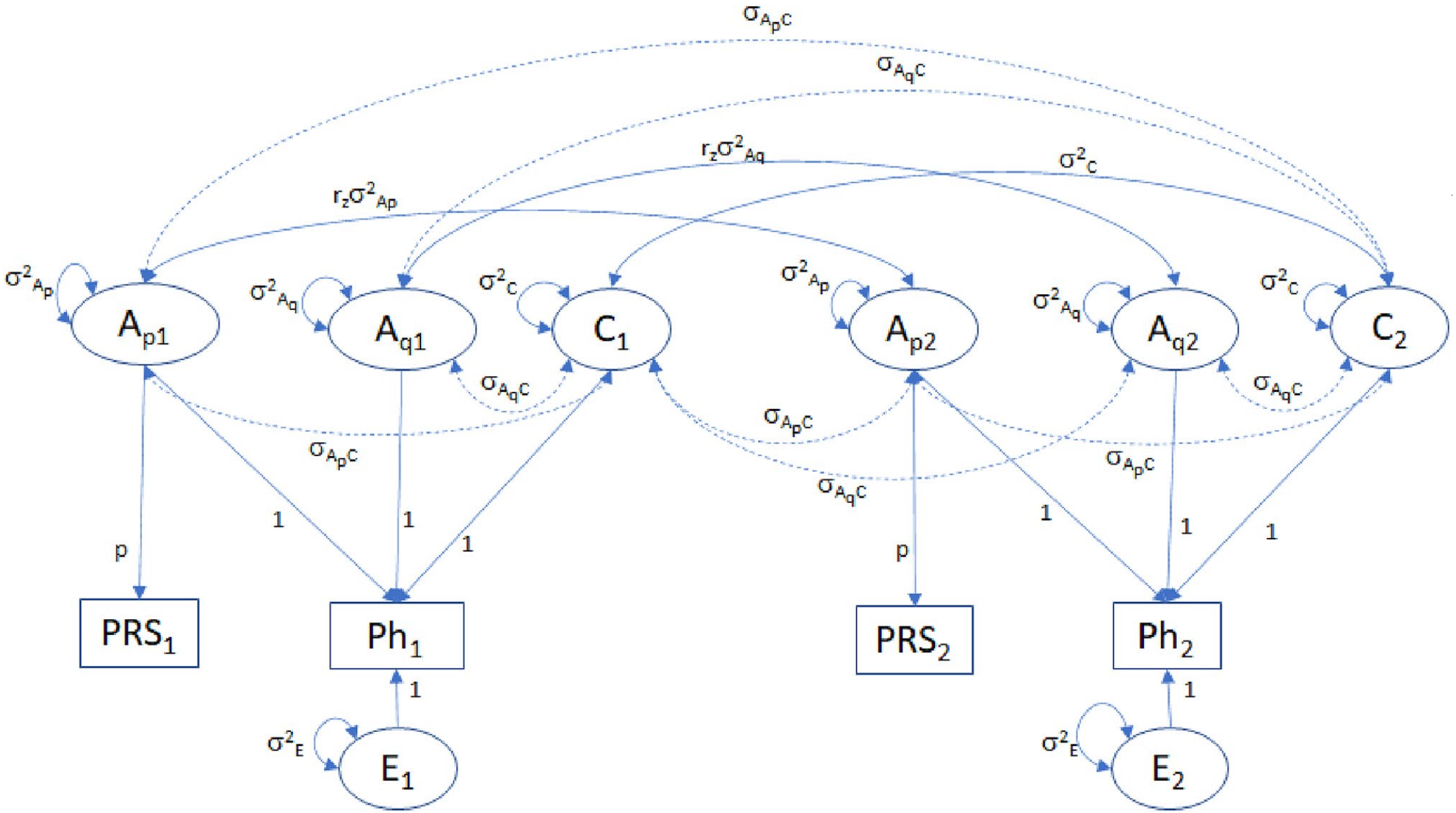
Extra development notes:

---

Could add the r12 assumption test

Regress PGS out of mean?

> If you look in this file you'll have all the alebra you want to do. We'll source that function, do an ACE function, ACE + covariate, ACE cov. Then spit out those models (possibly with the additional 2 models).


σ²Ap
σ²Aq
σ²C
rzσ²Ap
rzσ²Aq
σApC
σAqC
Ap1
Aq1
C1
Ap2
Aq2
C2
p
1
PRS1
Ph1
PRS2
Ph2
σ²E
E1
E2

Evokative rGE