Copy files

- Go to Faculty\marleen\Boulder2012\Multivariate
- Copy all files to your own directory

- Go to Faculty\kees\Boulder2012\Multivariate
- Copy all files to your own directory

Introduction to Multivariate Genetic Analysis (1)

Marleen de Moor, Kees-Jan Kan & Nick Martin

Outline

- 11.00-12.30
  - Lecture Bivariate Cholesky Decomposition
  - Practical Bivariate analysis of IQ and attention problems
- 12.30-13.30 LUNCH
- 13.30-15.00
  - Lecture Multivariate Cholesky Decomposition
  - PCA versus Cholesky
  - Practical Tri- and Four-variate analysis of IQ, educational attainment and attention problems

Outline

- 11.00-12.30
  - Lecture Bivariate Cholesky Decomposition
  - Practical Bivariate analysis of IQ and attention problems
- 12.30-13.30 LUNCH
- 13.30-15.00
  - Lecture Multivariate Cholesky Decomposition
  - PCA versus Cholesky
  - Practical Tri- and Four-variate analysis of IQ, educational attainment and attention problems

Aim / Rationale multivariate models

Aim:

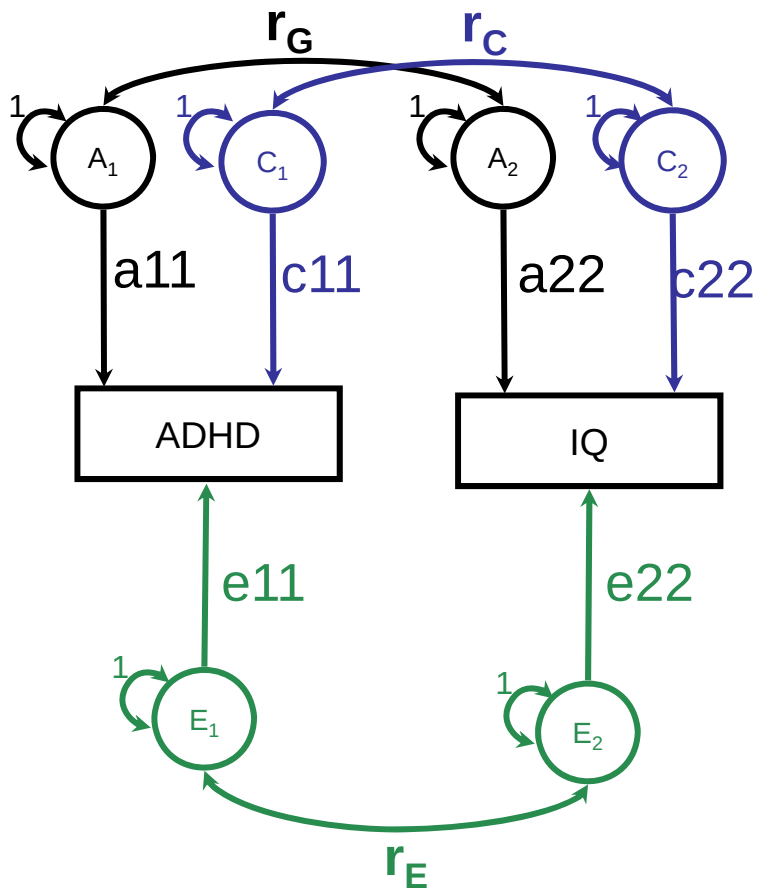
To examine the source of factors that make traits correlate or co-vary

Rationale:

- Traits may be correlated due to shared genetic factors (A or D) or shared environmental factors (C or E)
- Can use information on multiple traits from twin pairs to partition covariation into genetic and environmental components

Example

- Interested in relationship between ADHD and IQ
- How can we explain the association
  - Additive genetic factors (r_G)
  - Common environment (r_C)
  - Unique environment (r_E)



American Journal of Medical Genetics Part B (Neuropsychiatric Genetics) 124B:41–47 (2004)

Co-Occurrence of ADHD and Low IQ Has Genetic Origins

J. Kuntsi,[1] T.C. Eley,[1] A. Taylor,[1] C. Hughes,[2] P. Asherson,[1] A. Caspi,[1] and T.E. Moffitt[1]*

[1]*Social, Genetic and Developmental Psychiatry Centre, Institute of Psychiatry, King's College London, London, United Kingdom*

[2]*Centre for Family Research, University of Cambridge, Cambridge, United Kingdom*

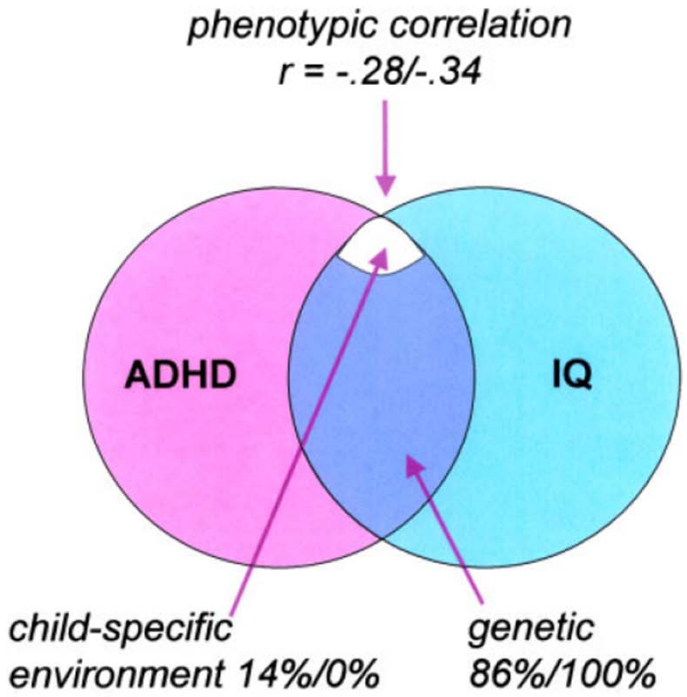


Fig. 1. Genetic and environmental contributions to the negative phenotypic correlation between IQ and both ADHD symptom scores and ADHD diagnosis. [Colour figure can be viewed in the online issue, which is available at www.interscience.wiley.com.]

Acta neurol. belg., 2006, 106, 191-207

Original articles

A longitudinal twin study on IQ, executive functioning, and attention problems during childhood and early adolescence

Tinca J. C. Polderman[1,2], M. Florencia Gosso[1,3], Danielle Posthuma[1], Toos C.E.M. van Beijsterveldt[1], Peter Heutink[1,3,4], Frank C. Verhulst[2] and Dorret I. Boomsma[1,4]

[1]Department of Biological Psychology, Vrije Universiteit Amsterdam ; [2]Department of Child and Adolescent Psychiatry, Erasmus University Rotterdam ; [3]Center for Neurogenomics and Cognitive Research - CNCR, Vrije Universiteit Amsterdam ; [4]Section of Medical Genomics, Department of Clinical Genetics and Anthropogenetics, VU Medical Center, Amsterdam, The Netherlands

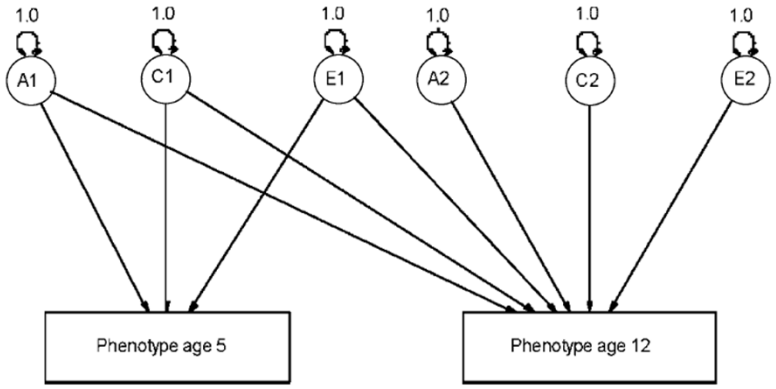


Fig. 3. — The bivariate (longitudinal) model represented for one individual

Bivariate Cholesky

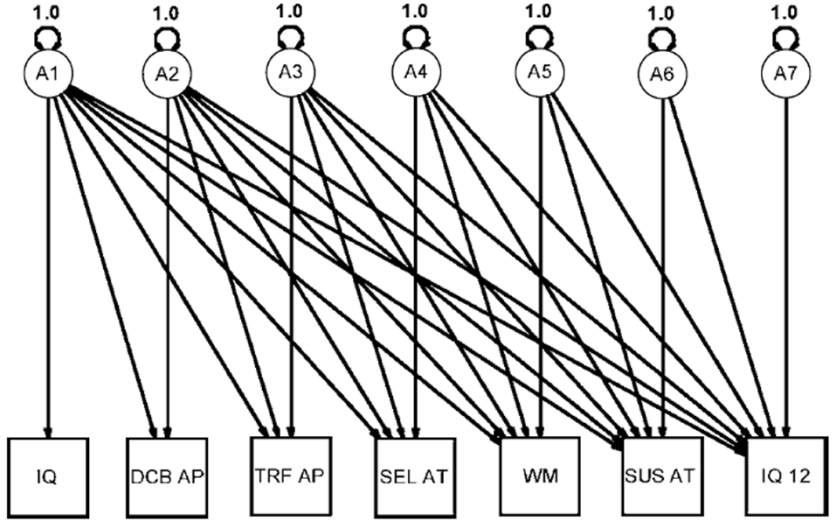


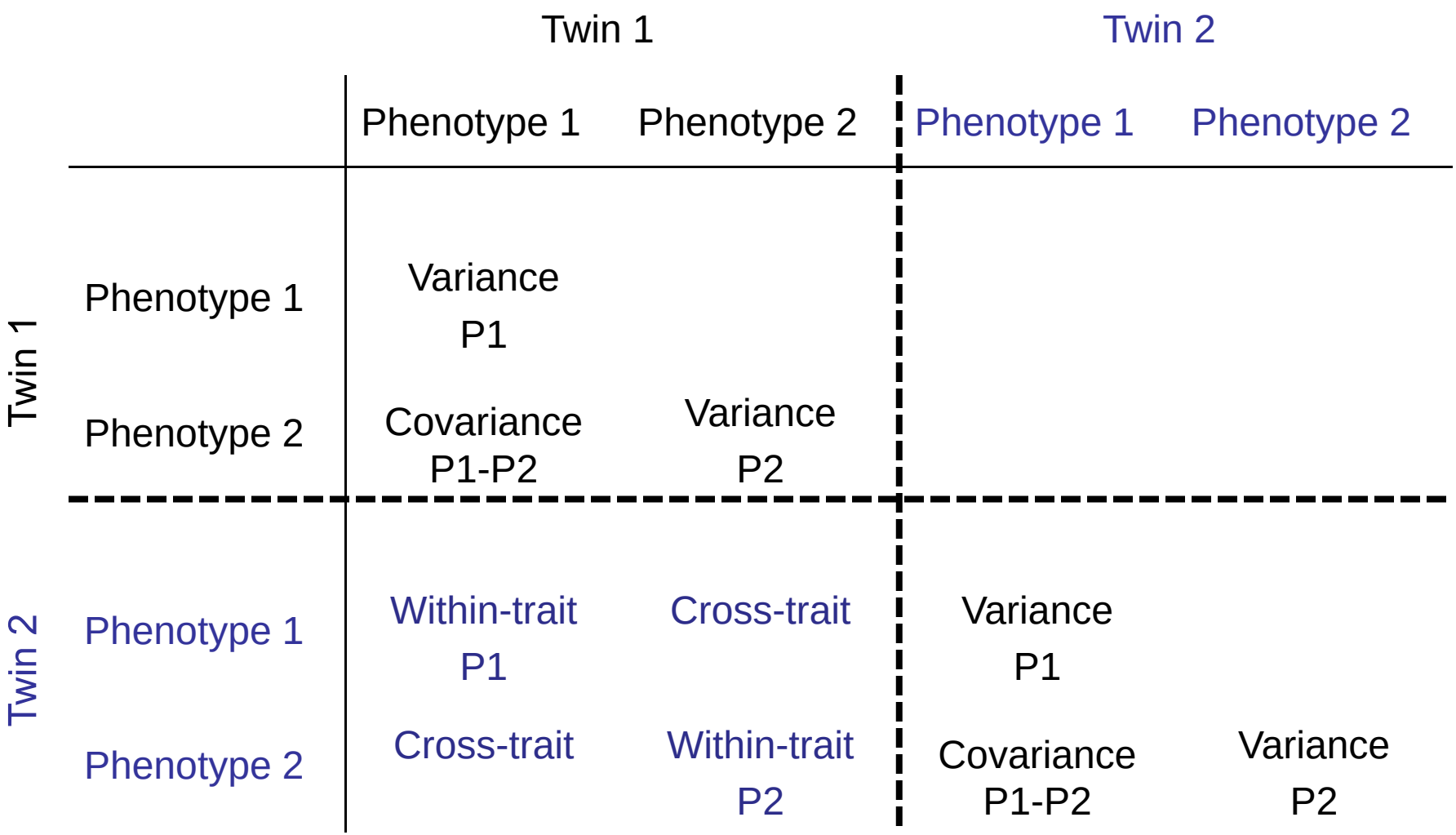
Fig. 4. — The multivariate (Cholesky) model with 7 variables represented for one individual

Multivariate Cholesky

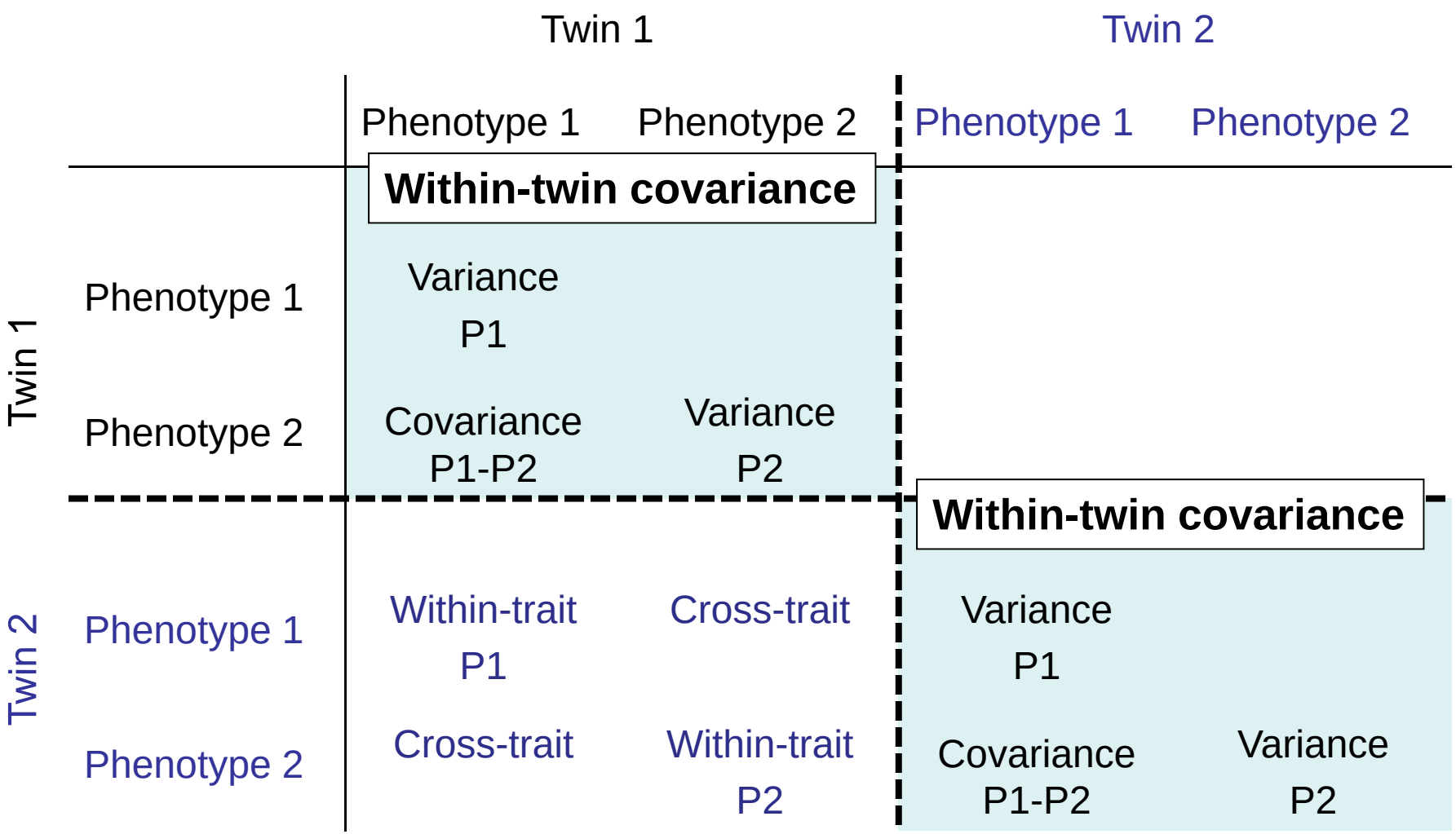
Sources of information

- Two traits measured in twin pairs
- Interested in:
  - Cross-trait covariance *within* individuals = phenotypic covariance
  - Cross-trait covariance *between* twins = cross-trait cross-twin covariance
  - MZ:DZ ratio of cross-trait covariance *between* twins

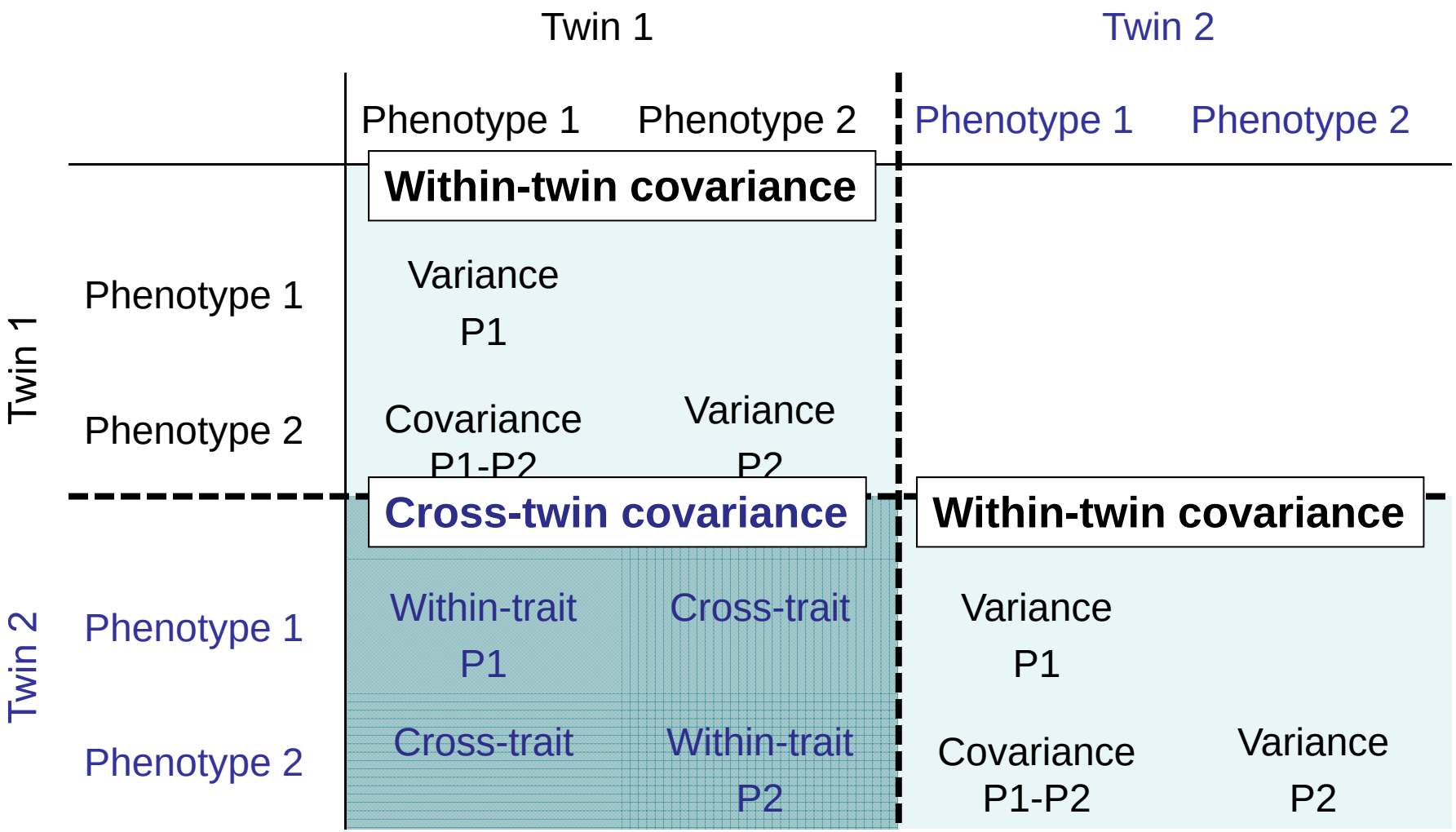
Observed Covariance Matrix: 4x4



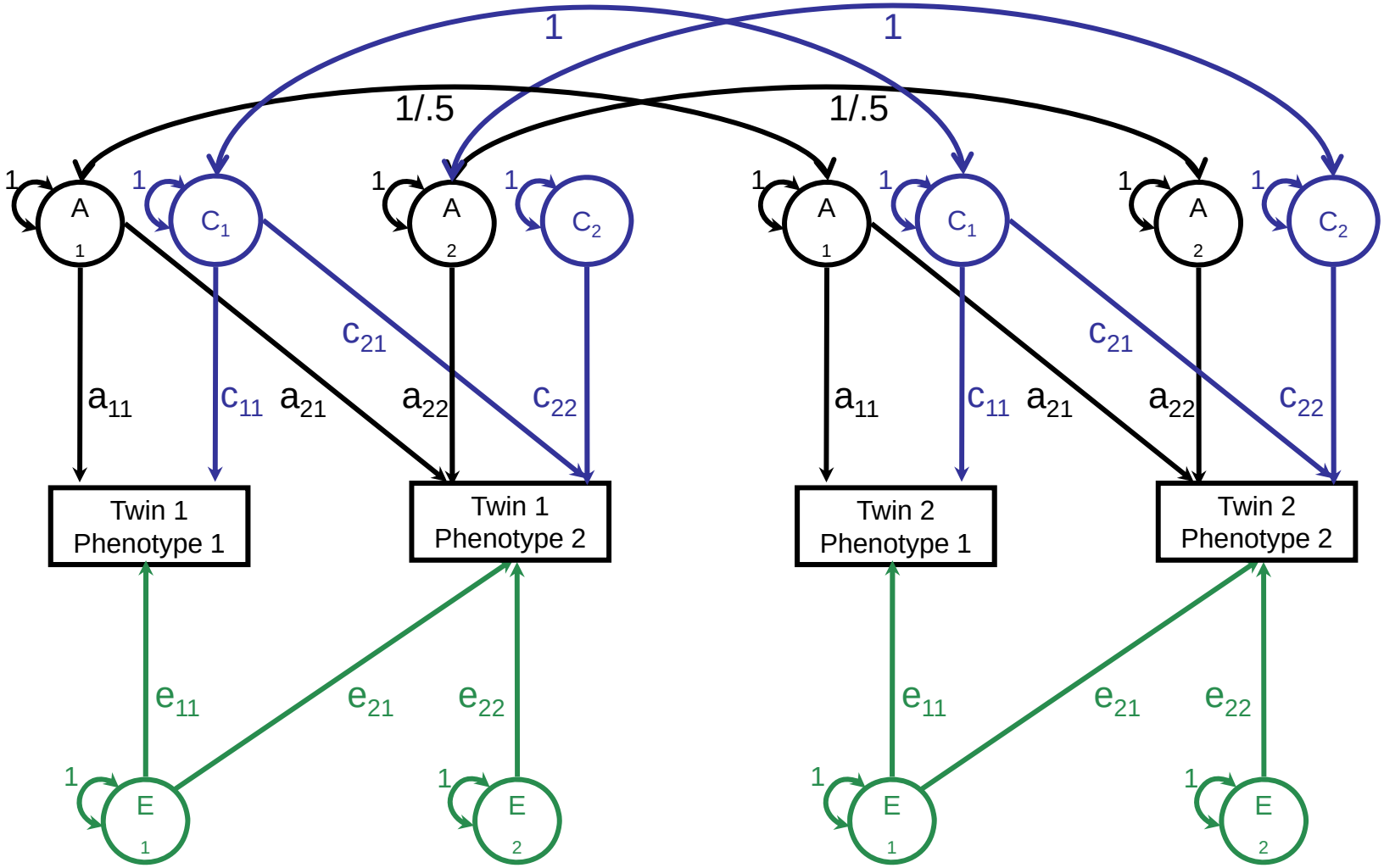
Observed Covariance Matrix: 4x4



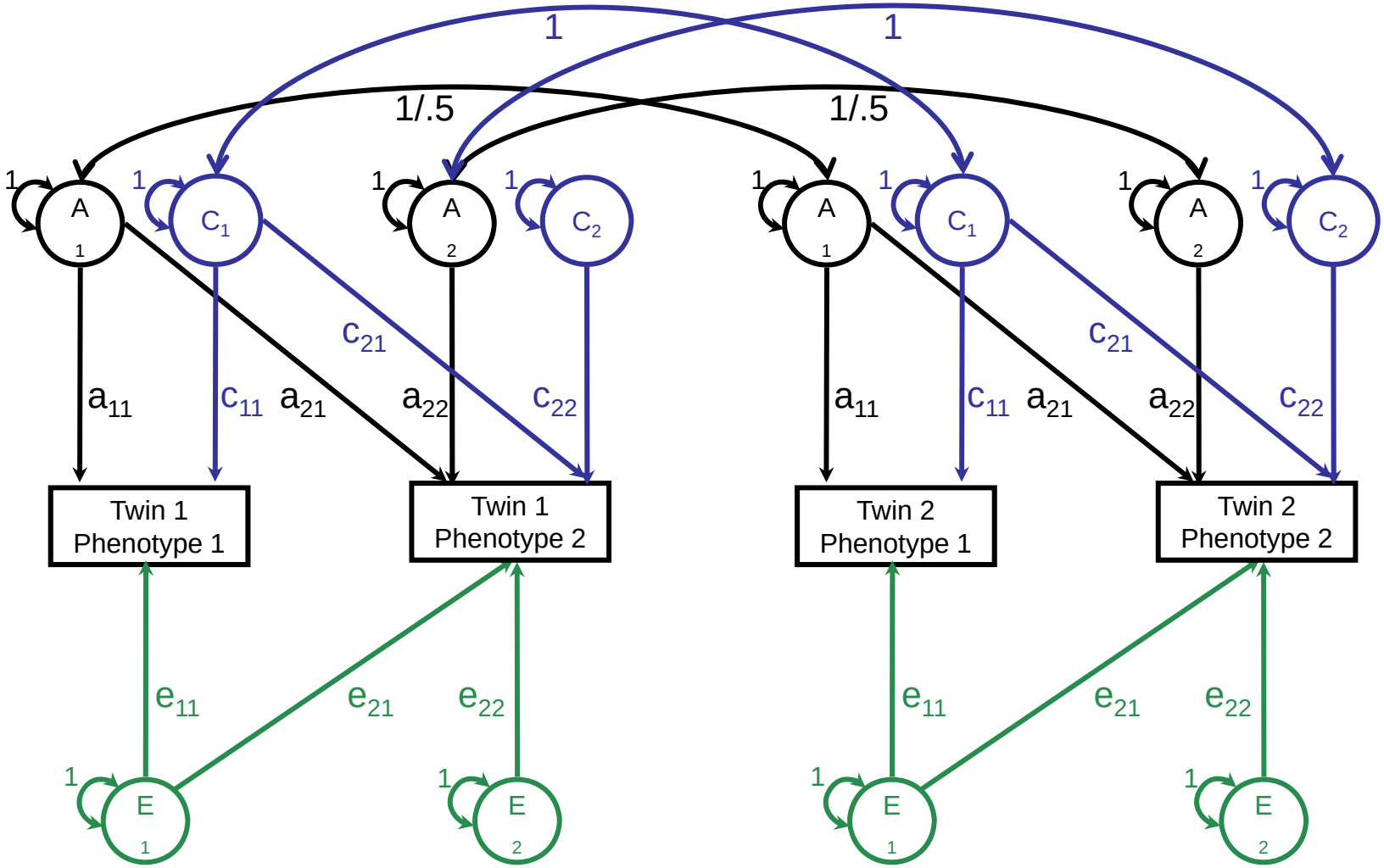
Observed Covariance Matrix: 4x4



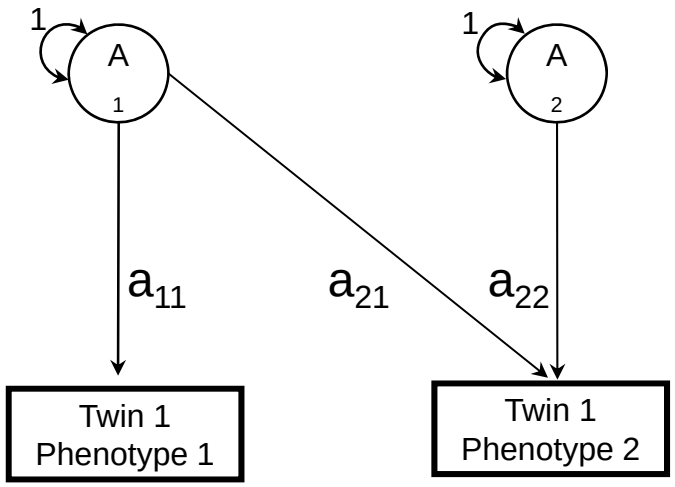
Cholesky decomposition



Now let's do the path tracing!

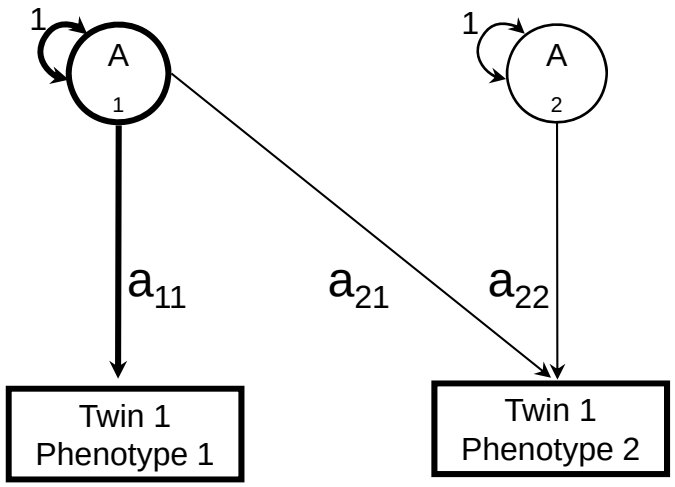


Within-Twin Covariances (A)



Twin 1		Twin 1
	Phenotype 1	Phenotype 2
Phenotype 1		
Phenotype 2		

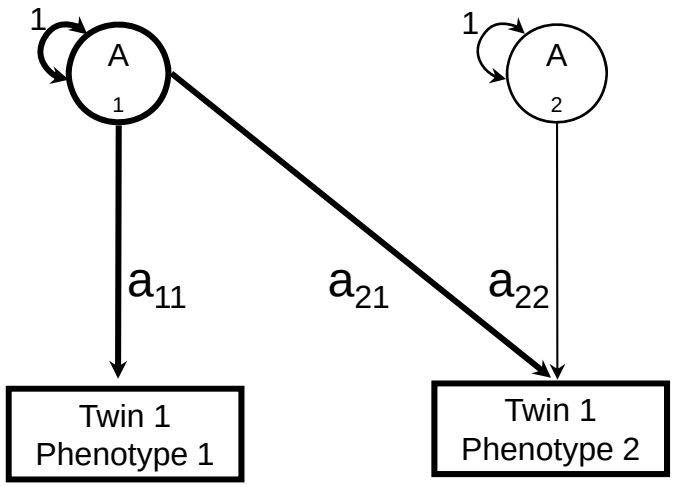
Within-Twin Covariances (A)



Twin 1

Twin 1	Phenotype 1	Phenotype 2
Phenotype 1	a_{11}^2	
Phenotype 2		

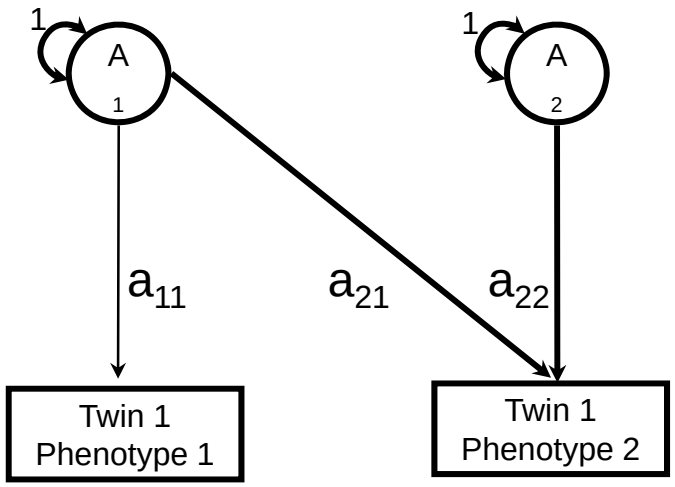
Within-Twin Covariances (A)



Twin 1

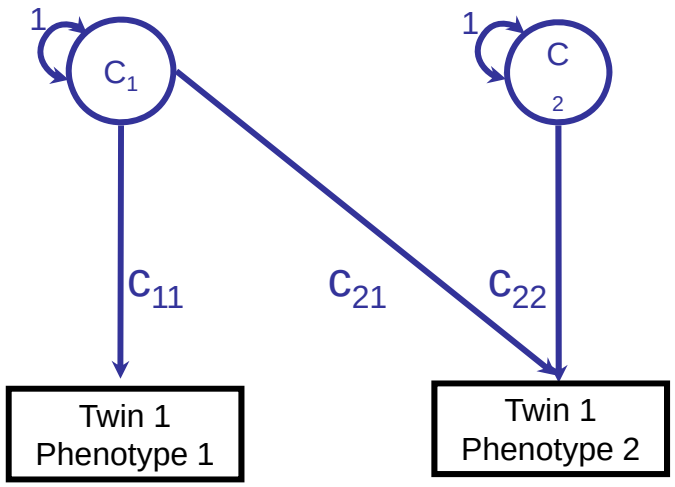
Twin 1	Phenotype 1	Phenotype 2
Phenotype 1	a_{11}^2	
Phenotype 2	$a_{11}a_{21}$	

Within-Twin Covariances (A)



	Twin 1 Phenotype 1	Twin 1 Phenotype 2
Twin 1 Phenotype 1	a_{11}^2	
Twin 1 Phenotype 2	$a_{11}a_{21}$	$a_{22}^2+a_{21}^2$

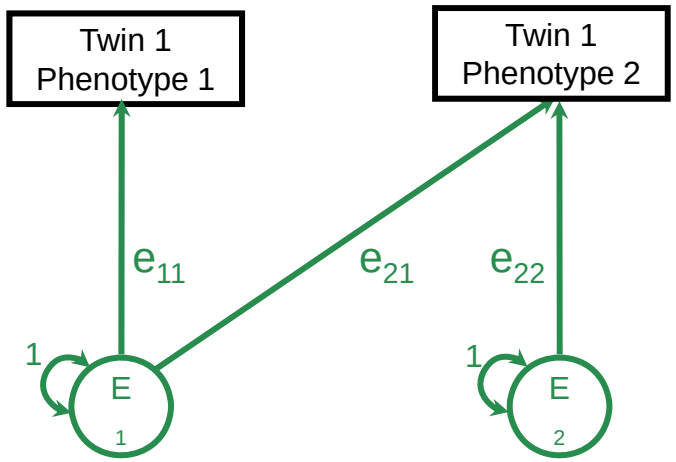
Within-Twin Covariances (C)



Twin 1

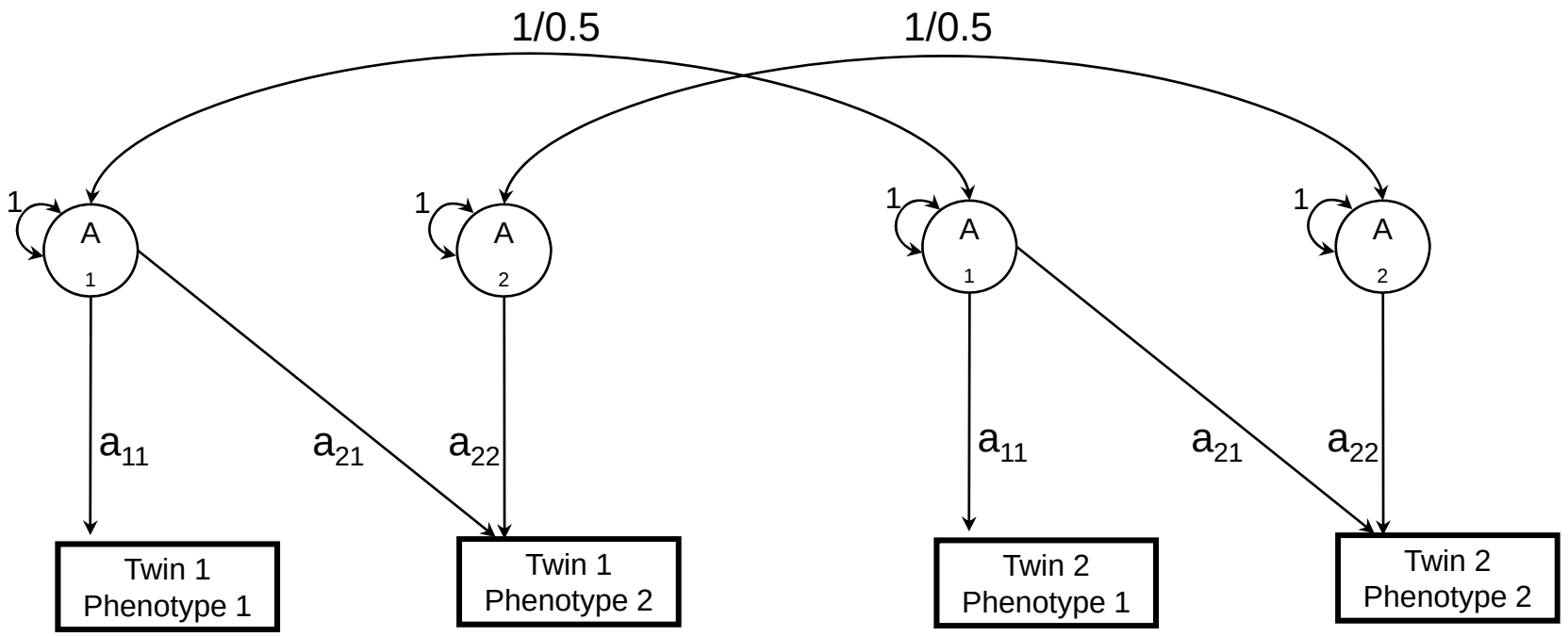
Twin 1	Phenotype 1	Phenotype 2
Phenotype 1	$a_{11}^2+c_{11}^2$	
Phenotype 2	$a_{11}a_{21}+c_{11}c_{21}$	$a_{22}^2+a_{21}^2+c_{22}^2+c_{21}^2$

Within-Twin Covariances (E)



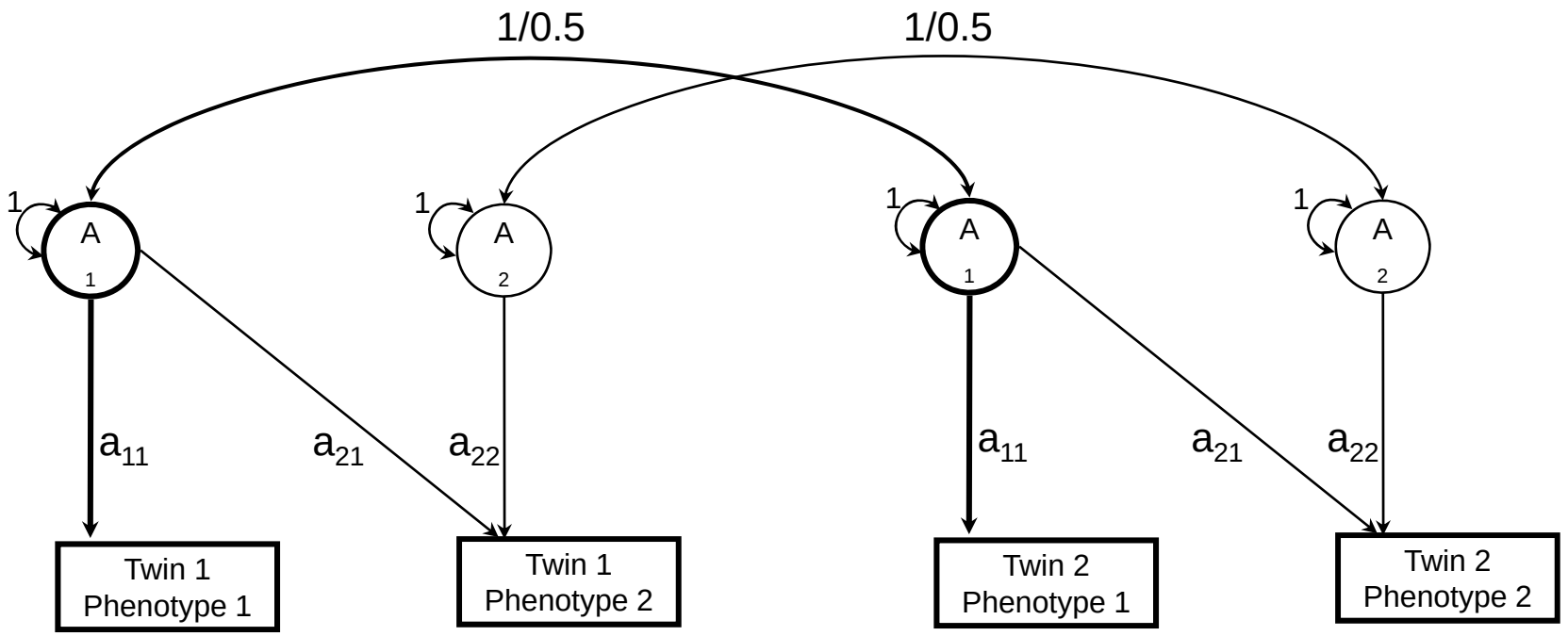
Twin 1	Twin 1	
	Phenotype 1	Phenotype 2
Phenotype 1	$a_{11}^2+c_{11}^2+e_{11}^2$	
Phenotype 2	$a_{11}a_{21}+c_{11}c_{21}+e_{11}e_{21}$	$a_{22}^2+a_{21}^2+c_{22}^2+c_{21}^2+e_{22}^2+e_{21}^2$

Cross-Twin Covariances (A)



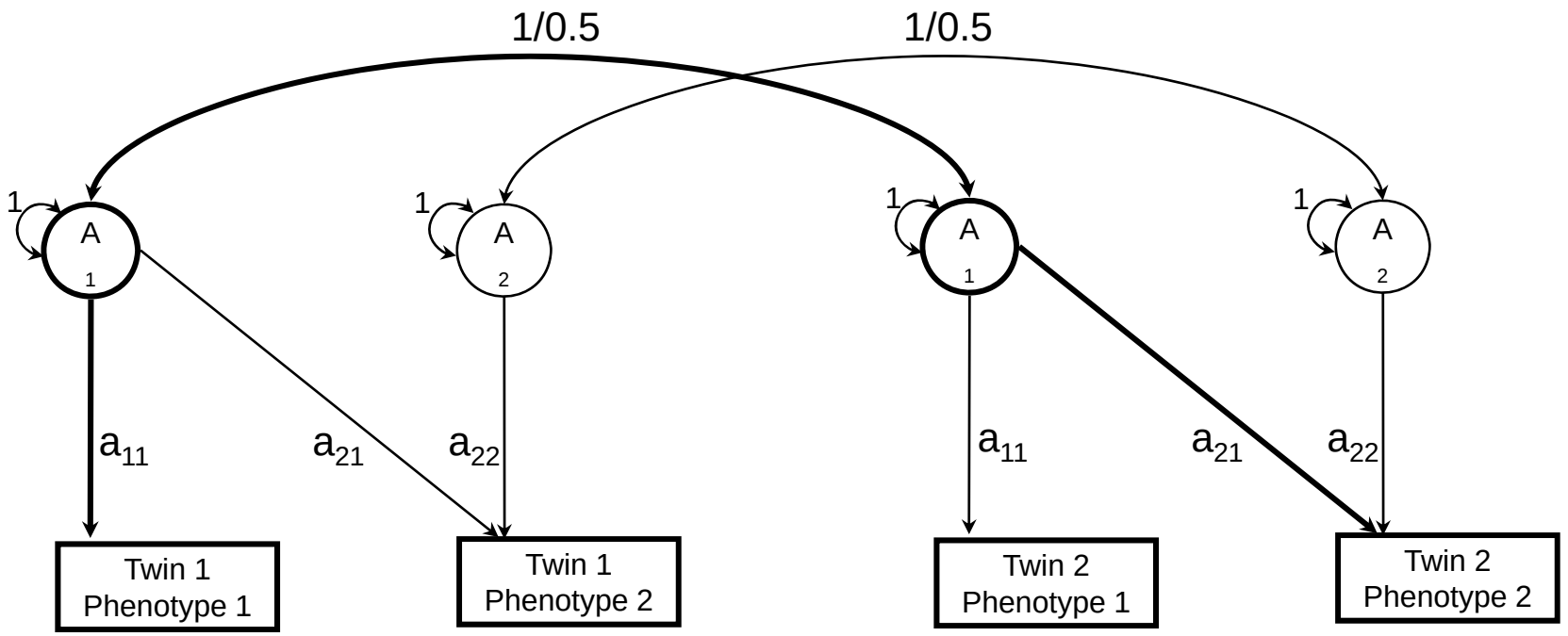
Twin 2	Twin 1 Phenotype 1	Twin 1 Phenotype 2
Phenotype 1		
Phenotype 2		

Cross-Twin Covariances (A)



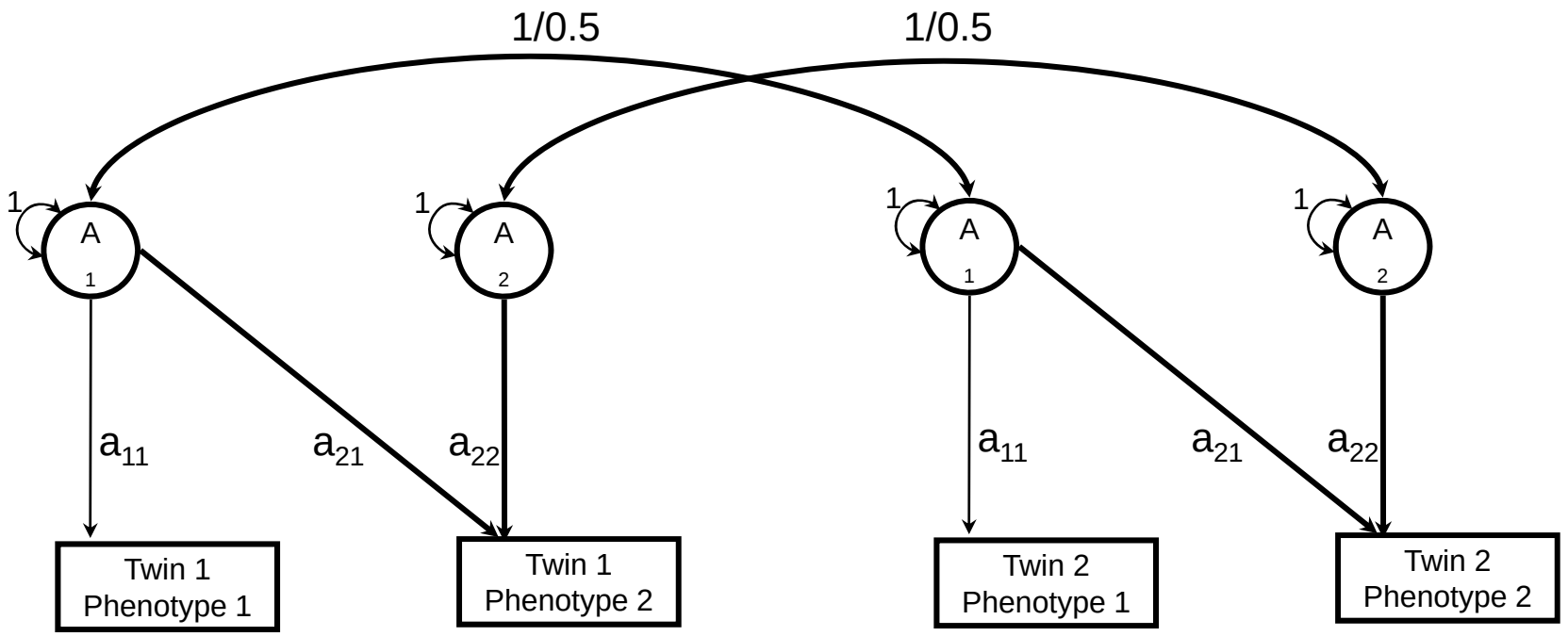
	Twin 1	
Twin 2	Phenotype 1	Phenotype 2
Phenotype 1	$1/0.5a_{11}^2$	
Phenotype 2		

Cross-Twin Covariances (A)



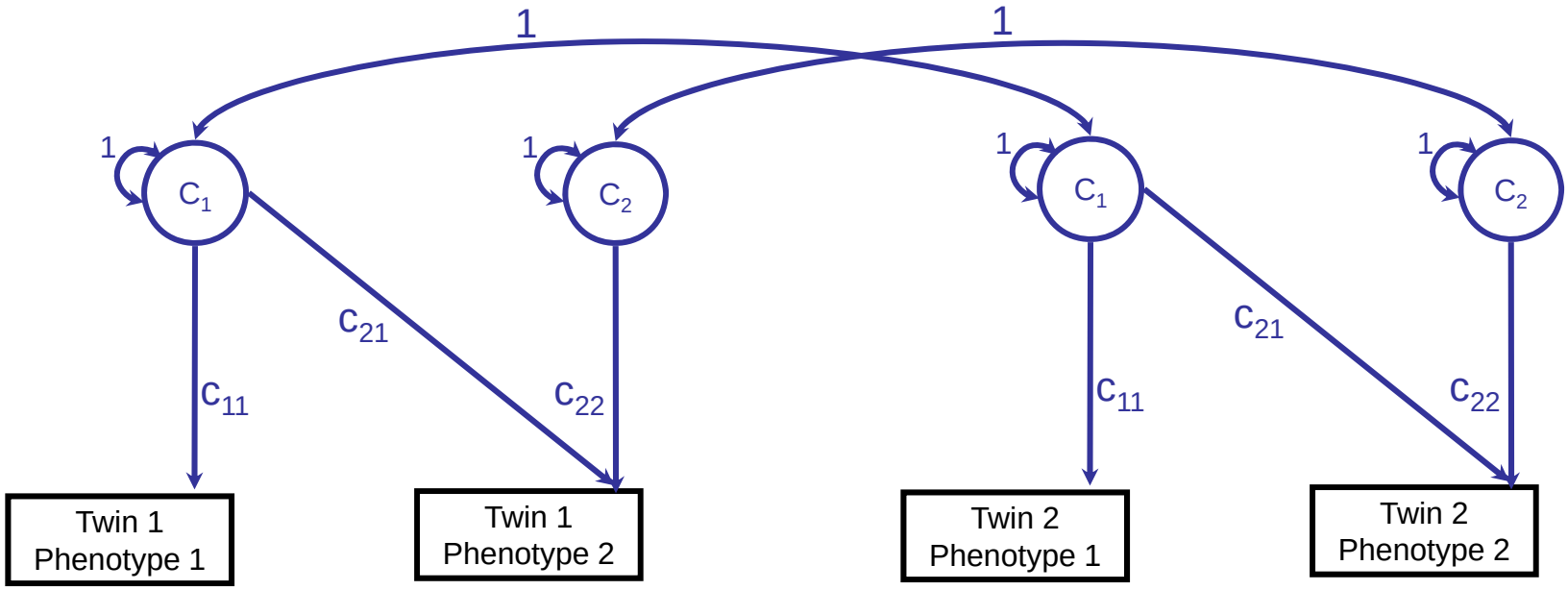
		Twin 1	
		Phenotype 1	Phenotype 2
Twin 2	Phenotype 1	$1/0.5a_{11}^2$	
	Phenotype 2	$1/0.5a_{11}a_{21}$	

Cross-Twin Covariances (A)



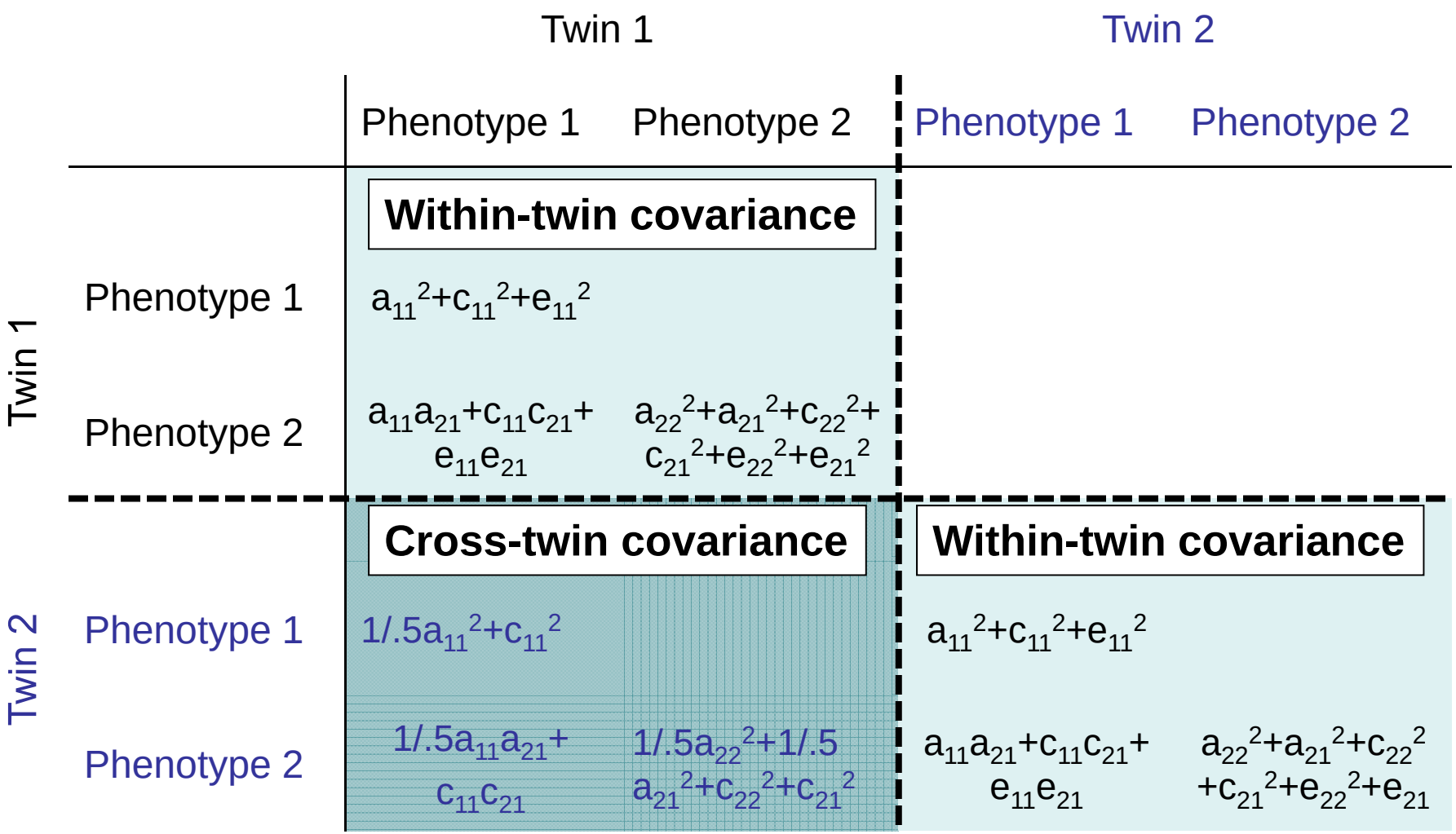
	Twin 1	
Twin 2	Phenotype 1	Phenotype 2
Phenotype 1	$1/0.5a_{11}^2$	
Phenotype 2	$1/0.5a_{11}a_{21}$	$1/0.5a_{22}^2+1/0.5a_{21}^2$

Cross-Twin Covariances (C)

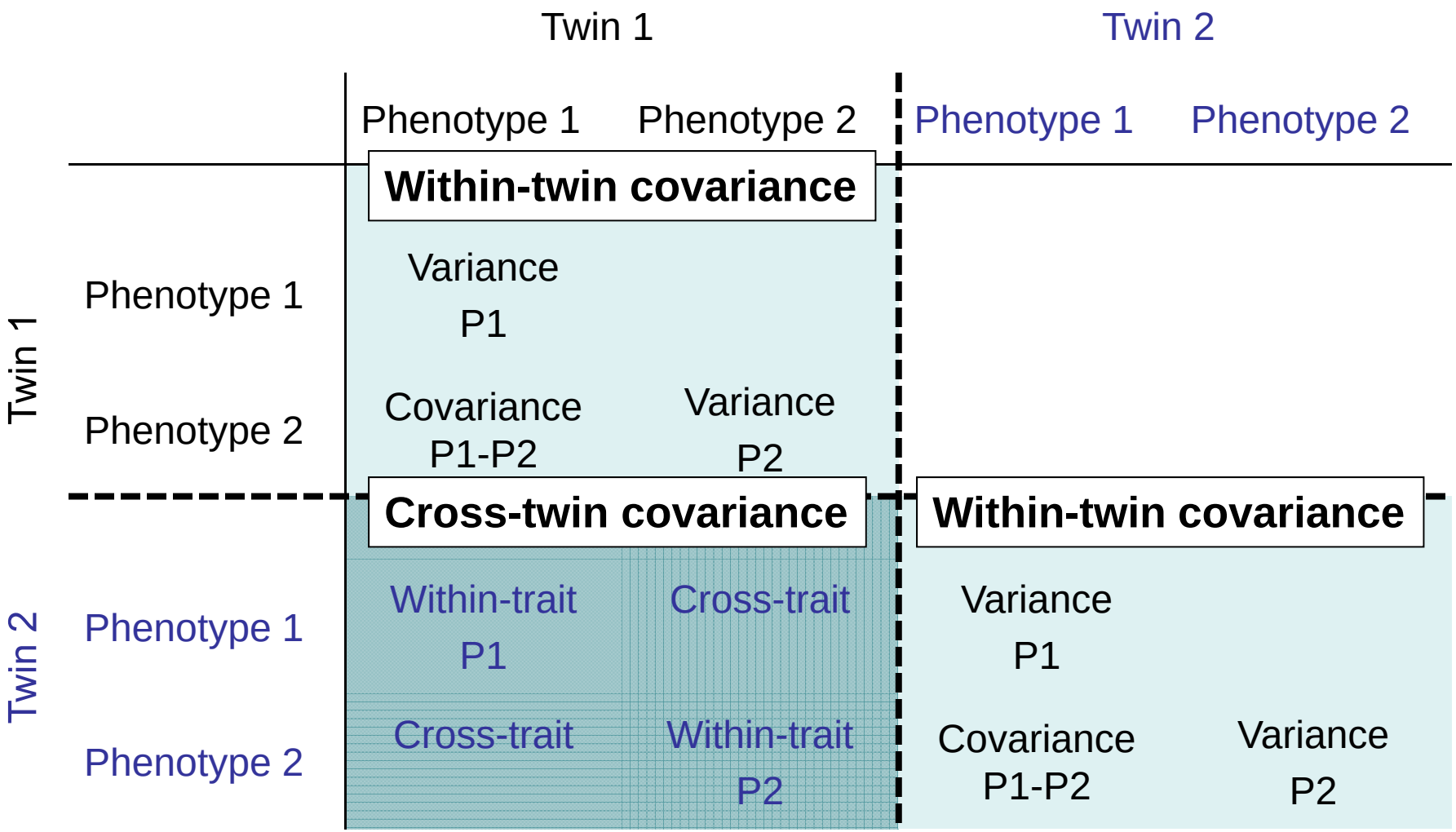


Twin 2 \ Twin 1	Phenotype 1	Phenotype 2
Phenotype 1	$1/0.5a_{11}^2+c_{11}^2$	
Phenotype 2	$1/0.5a_{11}a_{21}+c_{11}c_{21}$	$1/0.5a_{22}^2+1/0.5a_{21}^2+c_{22}^2+c_{21}^2$

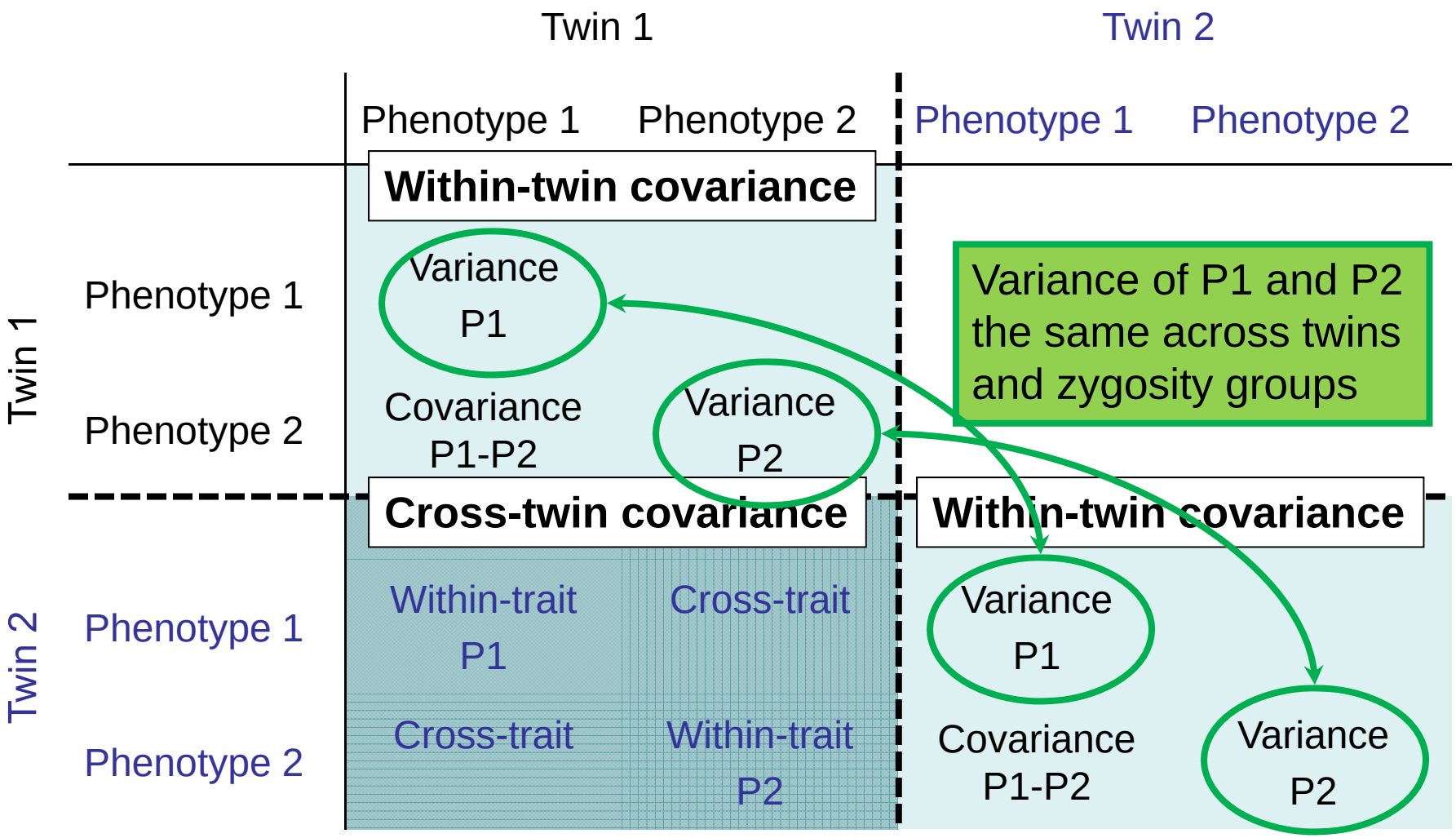
Predicted Model



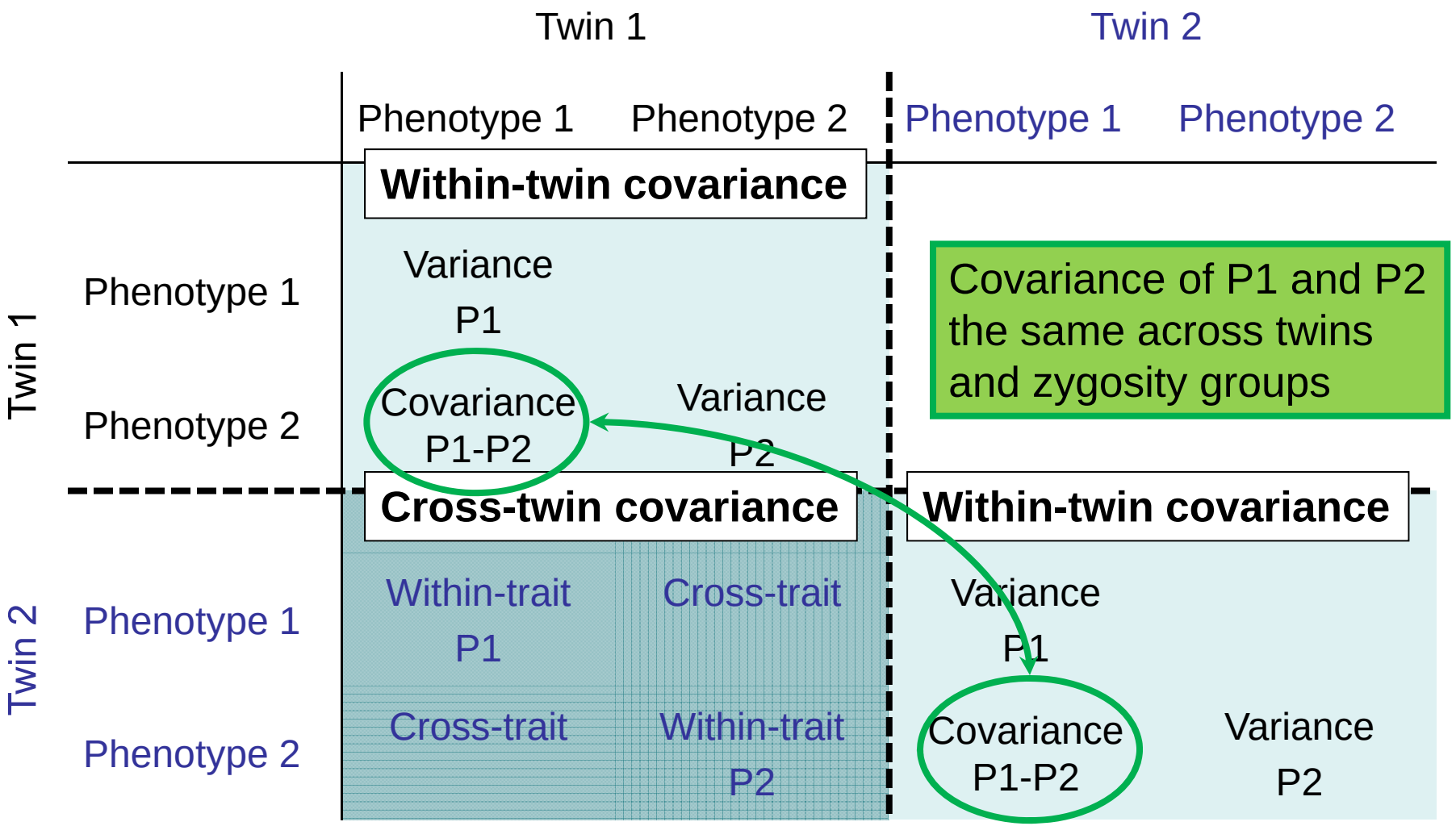
Predicted Model



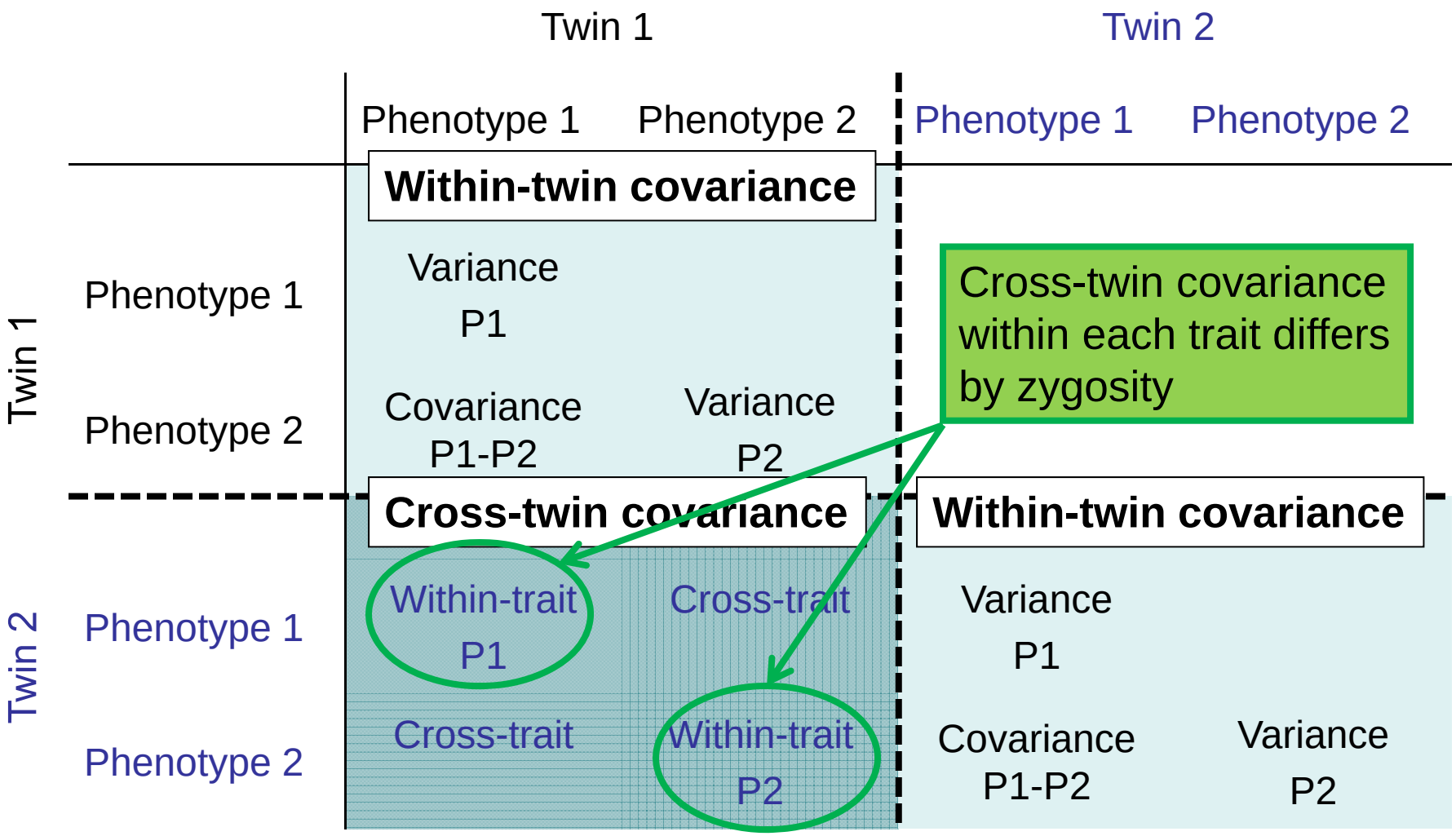
Predicted Model



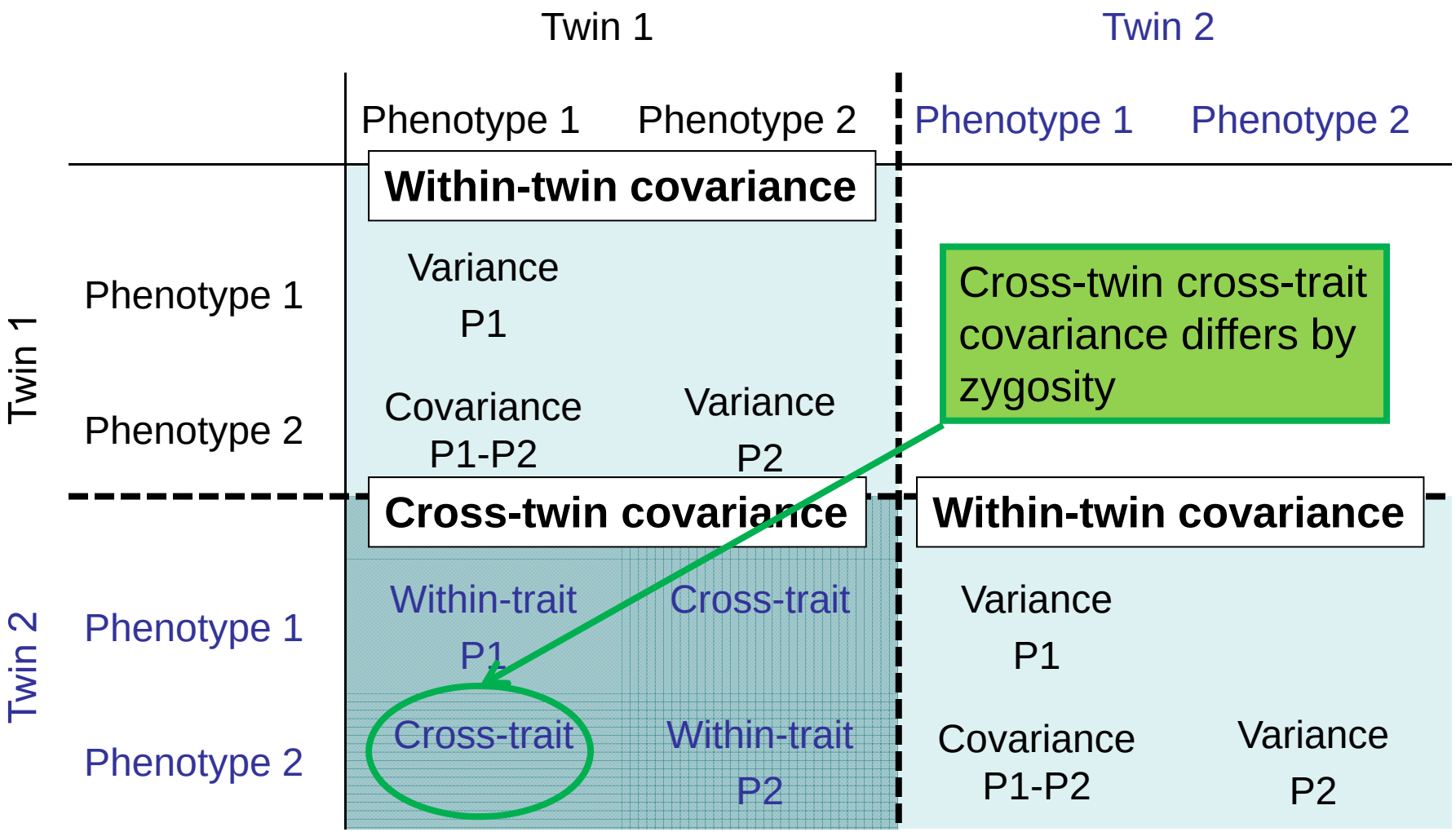
Predicted Model



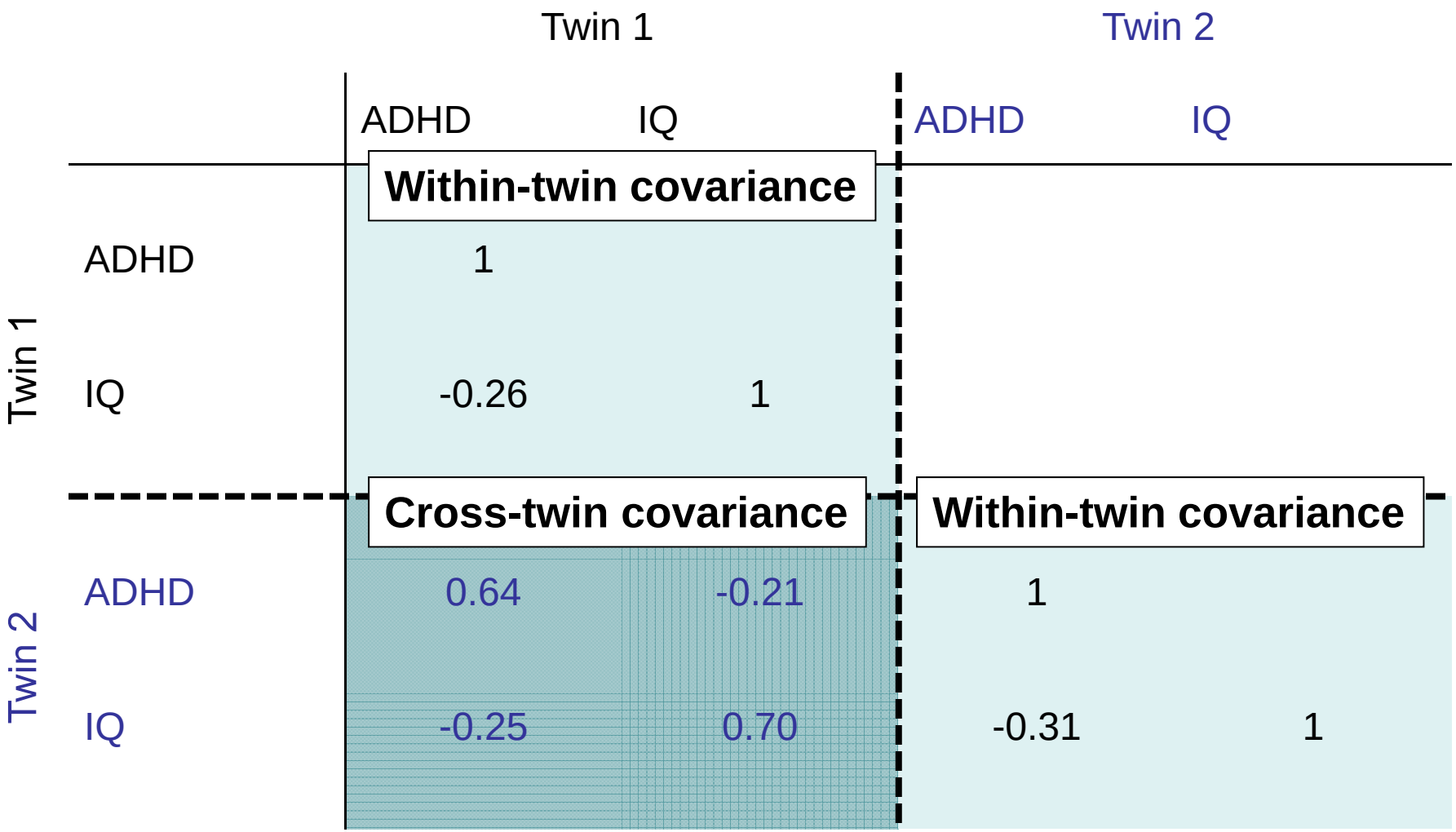
Predicted Model



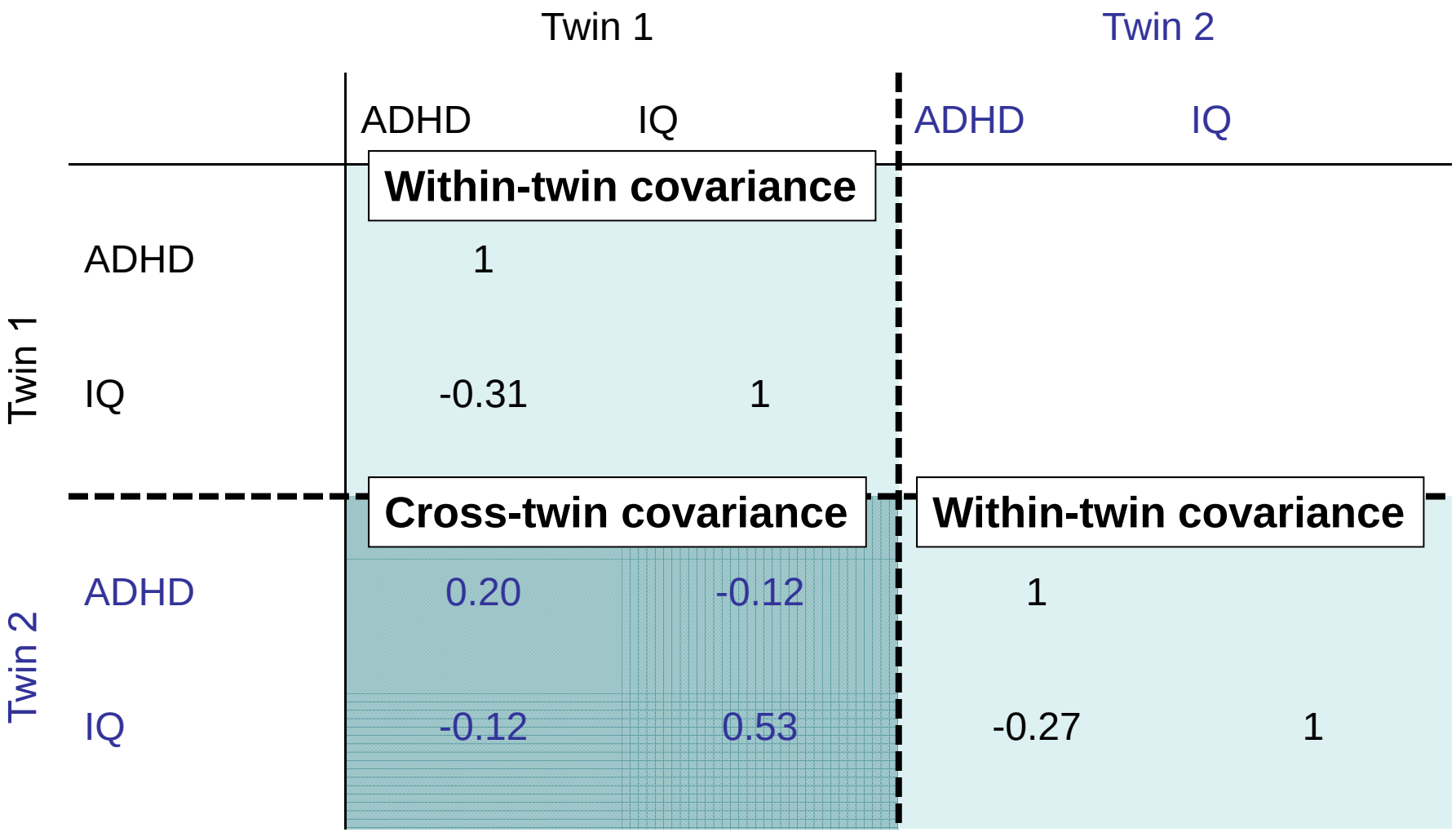
Predicted Model



Example covariance matrix MZ



Example covariance matrix DZ



Kuntsi et al. study

TABLE IA. Within-Pair Pearson Correlations: ADHD Symptom Scores and IQ

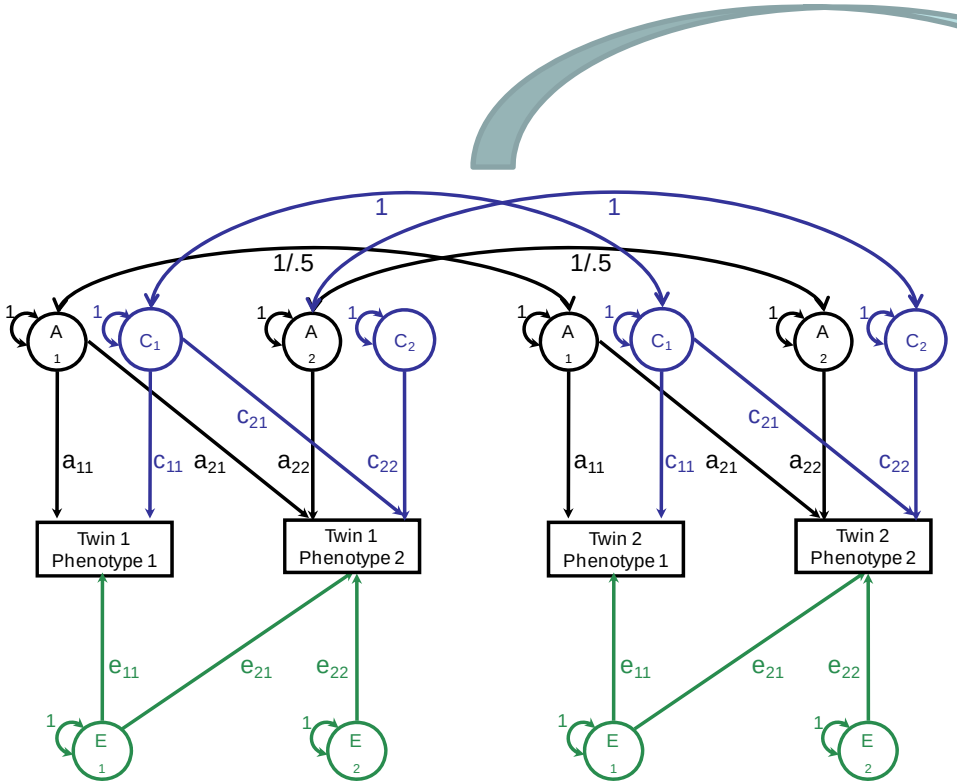
	Twin 1 ADHD symptoms	Twin 1 IQ	Twin 2 ADHD symptoms	Twin 2 IQ
MZ twins				
Twin 1 ADHD symptoms	1.00			
Twin 1 IQ	−0.26	1.00		
Twin 2 ADHD symptoms	0.64	−0.21	1.00	
Twin 2 IQ	−0.25	0.70	−0.31	1.00
Mean (SD)	15.46 (11.52)[a]	96.81 (13.64)	14.99 (11.12)[a]	97.15 (14.32)
DZ twins				
Twin 1 ADHD symptoms	1.00			
Twin 1 IQ	−0.31	1.00		
Twin 2 ADHD symptoms	0.20	−0.12	1.00	
Twin 2 IQ	−0.12	0.53	−0.27	1.00
Mean (SD)	16.93 (11.86)[a]	98.88 (14.61)	14.30 (11.00)[a]	98.44 (15.07)

[a]Prior to transformation.

Summary

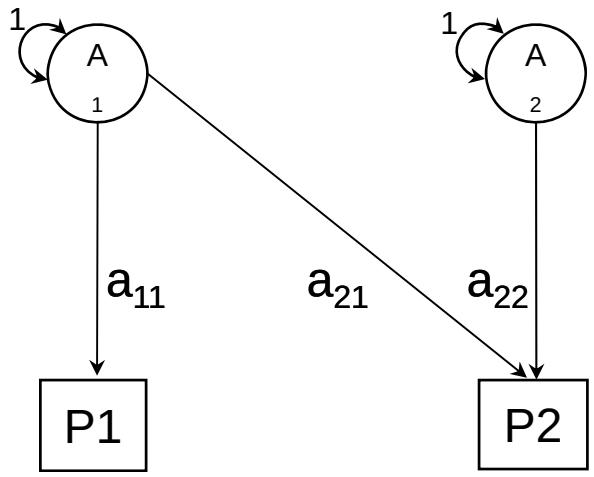
- Within-twin cross-trait covariance (phenotypic covariance) implies common aetiological influences
- Cross-twin cross-trait covariances >0 implies common aetiological influences are familial
- Whether familial influences are genetic or common environmental is shown by MZ:DZ ratio of cross-twin cross-trait covariances

Specification in OpenMx?



OpenMx script….

Within-Twin Covariance (A)



Path Tracing:

$$\Sigma_A = \begin{bmatrix} a_{11}^2 & a_{11}a_{21} \\ a_{21}a_{11} & a_{21}^2 + a_{22}^2 \end{bmatrix}$$

Lower 2 x 2 matrix:

	a1	a2
P1		0
P2	a_{21}	a_{22}

$$\Sigma_A = a * a^T$$

$$\Sigma_A = a\%*\%t(a)$$

$$= \begin{bmatrix} & 0 \\ a_{21} & a_{22} \end{bmatrix} * \begin{bmatrix} & a_{21} \\ 0 & a_{22} \end{bmatrix}$$

$$= \begin{bmatrix} a_{11}^2 + 0 \times 0 & a_{11}a_{21} + 0 \times a_{22} \\ a_{21}a_{11} + 0 \times a_{22} & a_{21}^2 + a_{22}^2 \end{bmatrix}$$

Within-Twin Covariance (A)

$$\Sigma_A = a * a^T$$
$$\Sigma_A = a\%*\%t(a)$$

$$= \begin{bmatrix} & 0 \\ a_{21} & a_{22} \end{bmatrix} * \begin{bmatrix} & a_{21} \\ 0 & a_{22} \end{bmatrix}$$

$$= \begin{bmatrix} a_{11}^2 + 0 \times 0 & a_{11}a_{21} + 0 \times a_{22} \\ a_{21}a_{11} + 0 \times a_{22} & a_{21}^2 + a_{22}^2 \end{bmatrix}$$

OpenMx

```
Vars <- c("FSIQ","AttProb")
nv  <- length(Vars)
aLabs <- c("a11“, "a21", "a22")

pathA <- mxMatrix(name = "a", type = "Lower", nrow = nv, ncol = nv, labels = aLabs)
covA <- mxAlgebra(name = "A", expression = a %*% t(a))
```

Within-Twin Covariance (A+C+E)

Σ_A = a %*% t(a)

Σ_C = c %*% t(c) = $\begin{bmatrix} c_{11}^2 & c_{11}c_{21} \\ c_{21}c_{11} & c_{21}^2 + c_{22}^2 \end{bmatrix}$

Σ_E = e %*% t(e) = $\begin{bmatrix} e_{11}^2 & e_{11}e_{21} \\ e_{21}e_{11} & e_{21}^2 + e_{22}^2 \end{bmatrix}$

Using matrix addition, the total within-twin covariance for the phenotypes is defined as:

$$\Sigma_V = \square + \Sigma_C + \Sigma_E$$

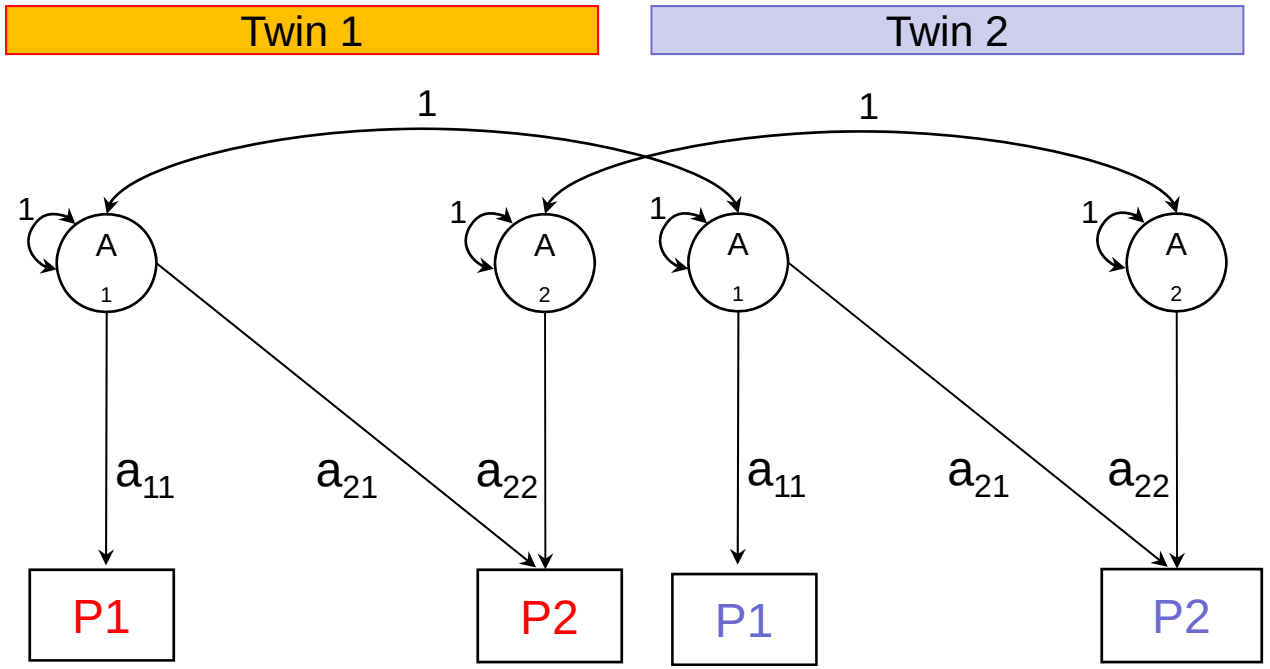
$$\Sigma_V = \begin{bmatrix} \square + c_{11}^2 + e_{11}^2 & \square + c_{11}c_{21} + e_{11}e_{21} \\ \square + c_{21}c_{11} + e_{11}e_{21} & \square + c_{21}^2 + c_{22}^2 + e_{21}^2 + e_{22}^2 \end{bmatrix}$$

OpenMx Matrices & Algebra

OpenMx

```
Vars <- c("FSIQ","AttProb")
nv <- length(Vars)

aLabs <- c("a11“, "a21", "a22")
cLabs <- c("c11", "c21", "c22")
eLabs <- c("e11", "e21", "e22")

# Matrices a, c, and e to store a, c, and e Path Coefficients
pathA <- mxMatrix(name = "a", type = "Lower", nrow = nv, ncol = nv, labels = aLabs)
pathC <- mxMatrix(name = "c", type = "Lower", nrow = nv, ncol = nv, labels = cLabs)
pathE <- mxMatrix(name = "e", type = "Lower", nrow = nv, ncol = nv, labels = eLabs)

# Matrices generated to hold A, C, and E computed Variance Components
covA <- mxAlgebra(name = "A", expression = a %*% t(a))
covC <- mxAlgebra(name = "C", expression = c %*% t(c))
covE <- mxAlgebra(name = "E", expression = e %*% t(e))

# Algebra to compute total variances and standard deviations (diagonal only)
covPh <- mxAlgebra(name = "V", expression = A+C+E)
matI <- mxMatrix(name= "I", type="Iden", nrow = nv, ncol = nv)
invSD <- mxAlgebra(name ="iSD", expression = solve(sqrt(I*V)))
```

MZ Cross-Twin Covariance (A)



Cross-twin within-trait:

P1-P1 = $1*a_{11}^2$

P2-P2 = $1*a_{22}^2+1*a_{21}^2$

Cross-twin cross-trait:

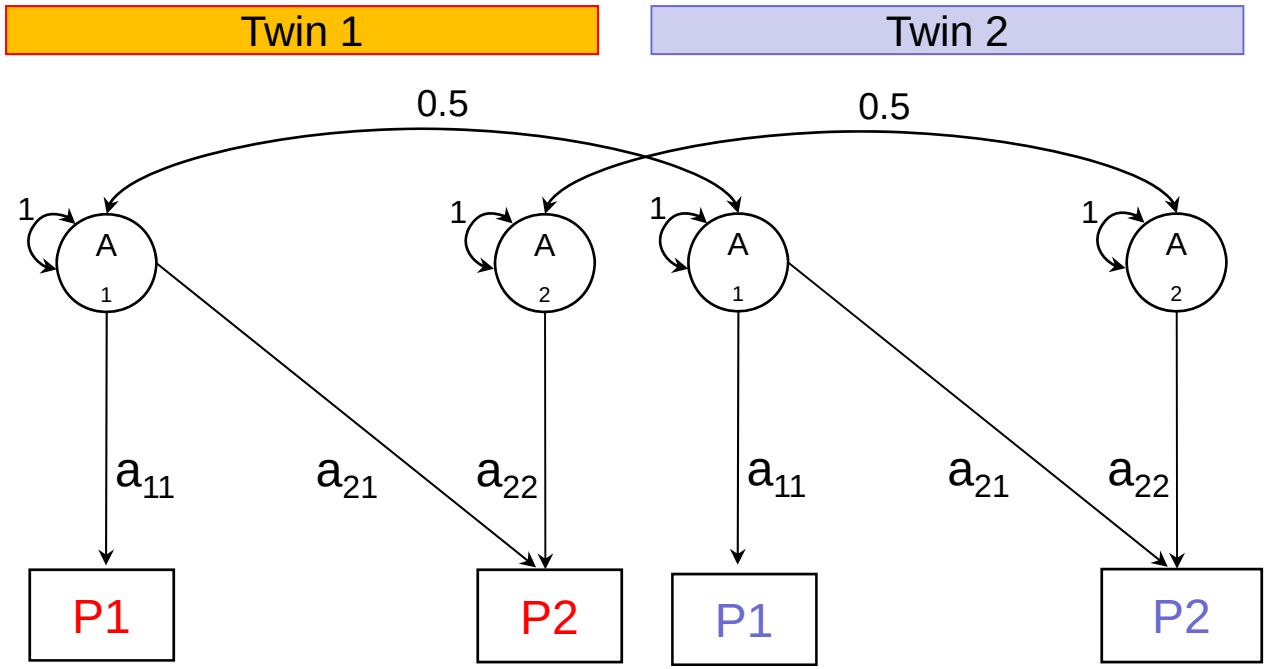
P1-P2 = $1*a_{11}a_{21}$

P2-P1 = $1*a_{21}a_{11}$

$$A = a\%*\%t(a)$$

$$= \begin{bmatrix} a_{11}^2 & a_{11}a_{21} \\ a_{21}a_{11} & (a_{21}^2 + a_{22}^2) \end{bmatrix}$$

DZ Cross-Twin Covariance (A)



Cross-twin within-trait:

P1-P1 = $0.5a_{11}^2$

P2-P2 = $0.5a_{22}^2+0.5a_{21}^2$

Cross-twin cross-trait:

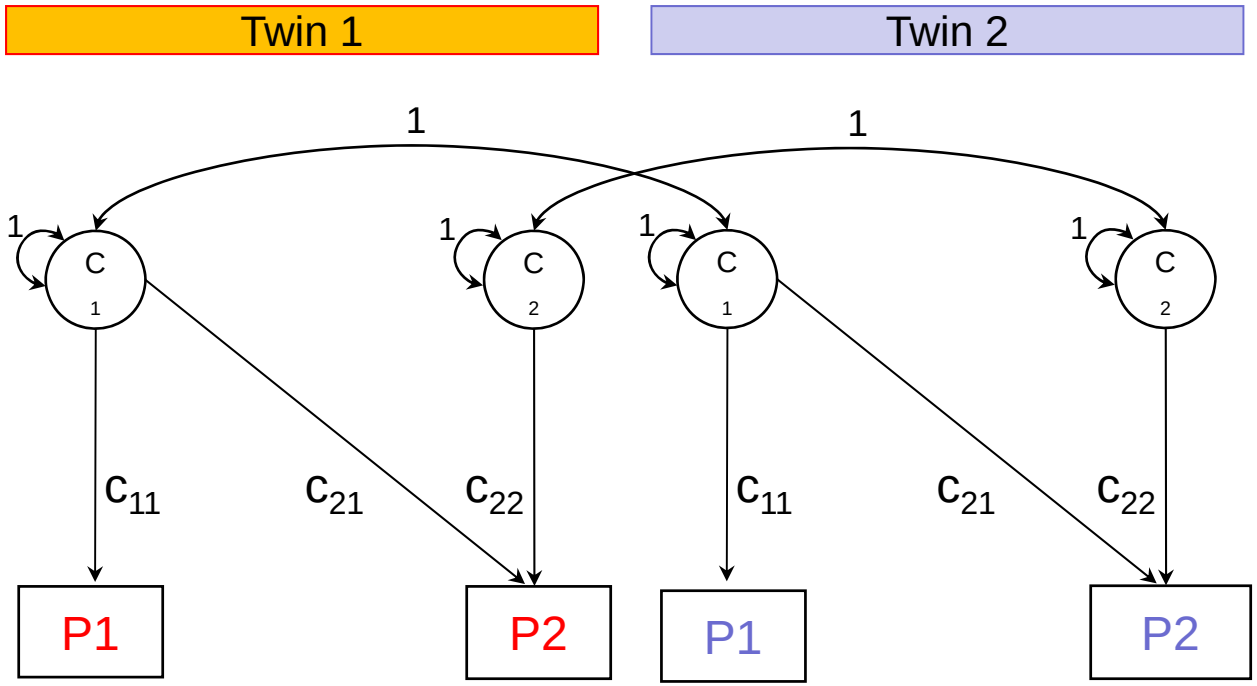
P1-P2 = $0.5a_{11}a_{21}$

P2-P1 = $0.5a_{21}a_{11}$

$$0.5\%x\%A = 0.5\%x\%(a\%*\%t(a))$$

$$= \begin{bmatrix} 0.5a_{11}^2 & 0.5a_{11}a_{21} \\ 0.5a_{21}a_{11} & 0.5(a_{21}^2 + a_{22}^2) \end{bmatrix}$$

MZ/DZ Cross-Twin Covariance (C)



Cross-twin within-trait:

P1-P1 = $1*c_{11}^2$

P2-P2 = $1*c_{22}^2+1*c_{21}^2$

Cross-twin cross-trait:

P1-P2 = $1*c_{11}c_{21}$

P2-P1 = $1*c_{21}c_{11}$

$$C = c\%*\%t(c)$$

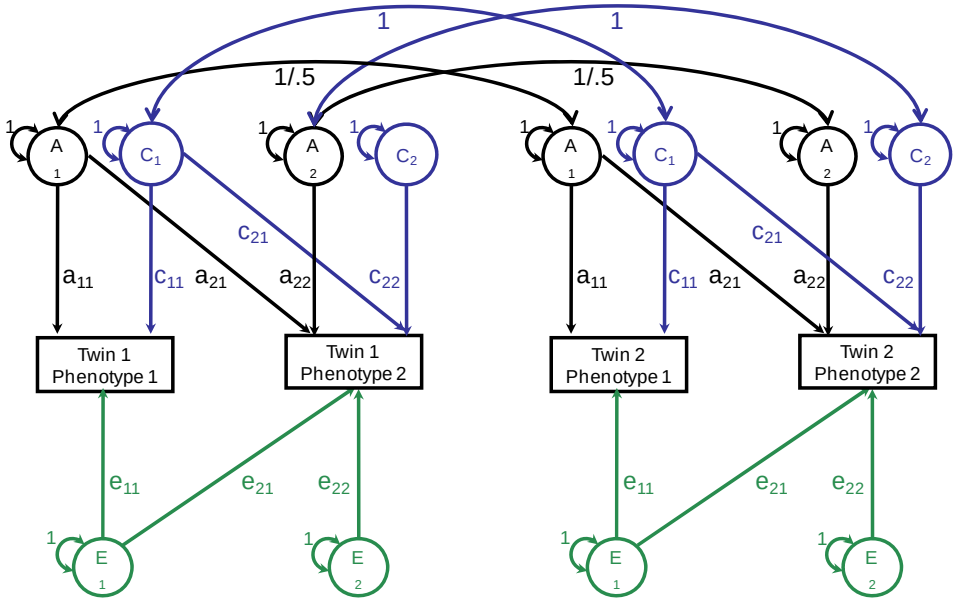
$$= \begin{bmatrix} c_{11}^2 & c_{11}c_{21} \\ c_{21}c_{11} & (c_{21}^2 + c_{22}^2) \end{bmatrix}$$

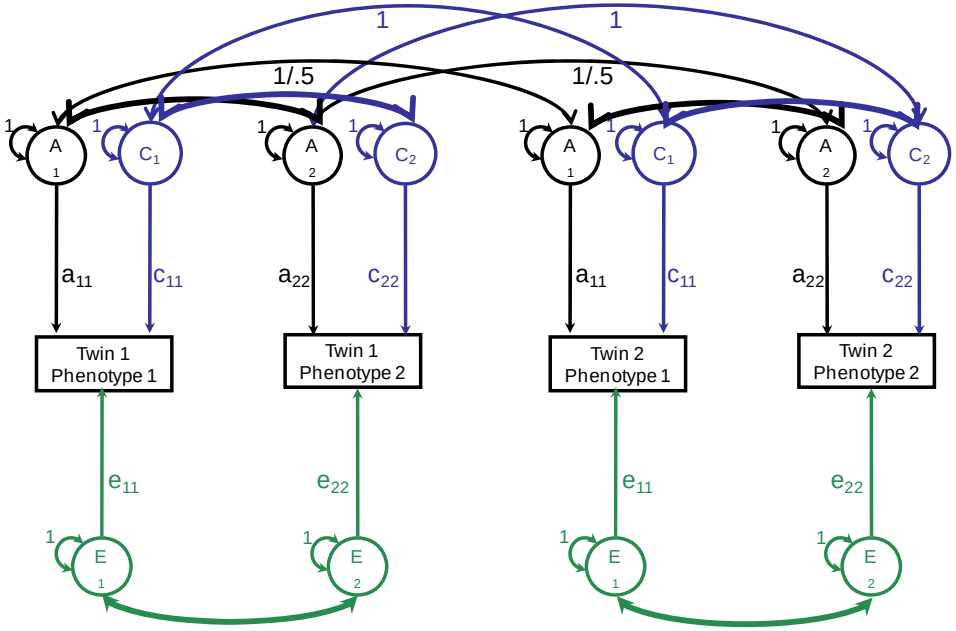
Covariance Model for Twin Pairs

OpenMx

```
# Algebra for expected variance/covariance matrix in MZ
expCovMZ <- mxAlgebra(name = "expCovMZ",
                      expression = rbind (cbind(A+C+E, A+C),
                                          cbind(A+C,   A+C+E) ) )

# Algebra for expected variance/covariance matrix in DZ
expCovDZ <- mxAlgebra(name = "expCovDZ",
                      expression = rbind (cbind(A+C+E,     0.5%x%A+C),
                                          cbind(0.5%x%A+C, A+C+E) ) )
```

Unstandardized vs standardized solution

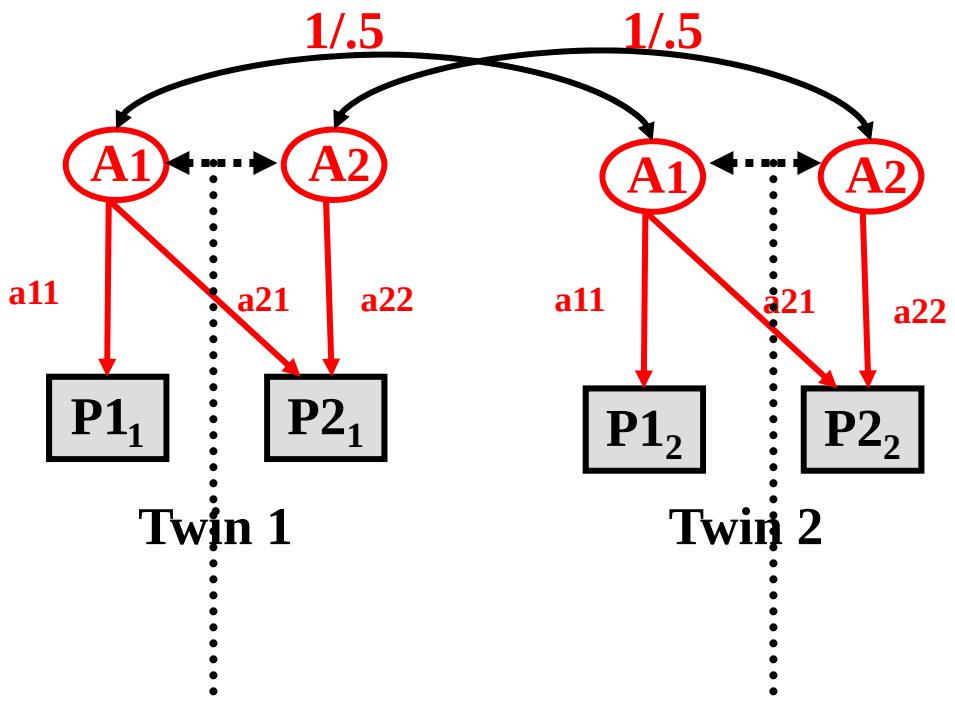




Genetic correlation

- It is calculated by dividing the genetic covariance by the square root of the product of the genetic variances of the two variables

Genetic correlation

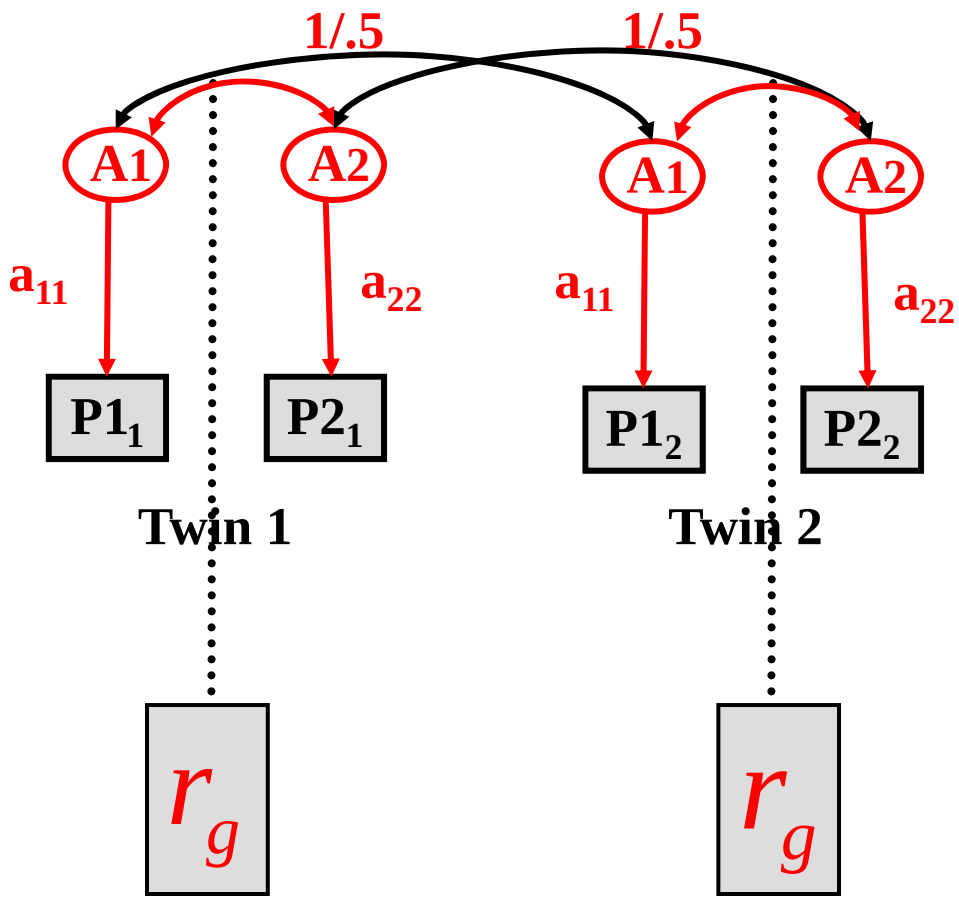


$$r_g = \frac{a_{21}a_{11}}{\sqrt{a_{11}^2 * (a_{21}^2 + a_{22}^2)}}$$

Genetic covariance

Sqrt of the product of the genetic variances

Standardized Solution = Correlated Factors Solution



Genetic correlation – matrix algebra

$$\Sigma_A = \begin{bmatrix} a_{11}^2 & a_{11}a_{21} \\ a_{21}a_{11} & a_{21}^2 + a_{22}^2 \end{bmatrix}$$

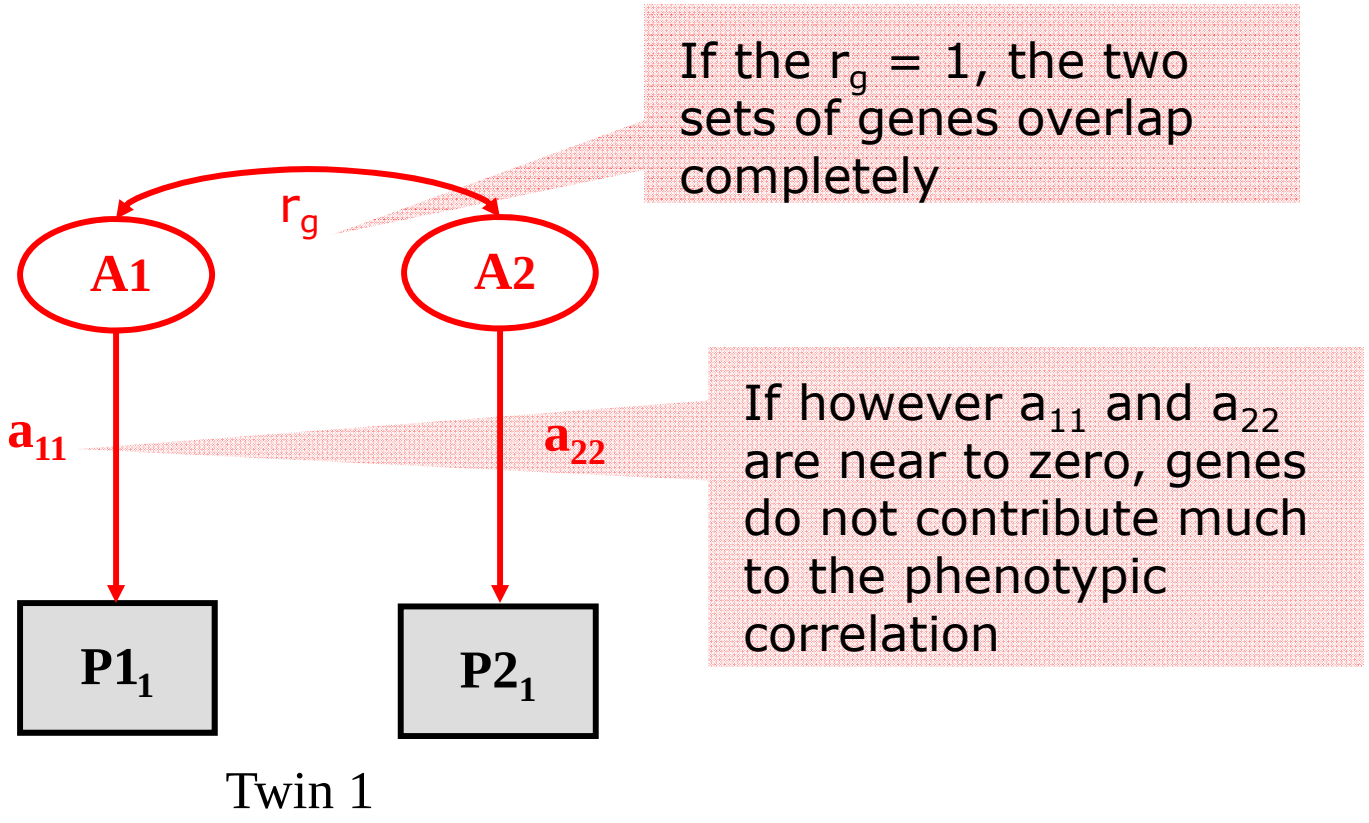
$$= \begin{bmatrix} \sigma_{A_{11}}^2 & \sigma_{A_{12}}^2 \\ \sigma_{A_{21}}^2 & \sigma_{A_{22}}^2 \end{bmatrix}$$

$$\begin{bmatrix} 1 & r_G \\ r_G & 1 \end{bmatrix} = \begin{bmatrix} \frac{1}{\sqrt{\sigma_{A_{11}}^2}} & 0 \\ 0 & \frac{1}{\sqrt{\sigma_{A_{22}}^2}} \end{bmatrix} * \begin{bmatrix} \sigma_{A_{11}}^2 & \sigma_{A_{12}}^2 \\ \sigma_{A_{21}}^2 & \sigma_{A_{22}}^2 \end{bmatrix} * \begin{bmatrix} \frac{1}{\sqrt{\sigma_{A_{11}}^2}} & 0 \\ 0 & \frac{1}{\sqrt{\sigma_{A_{22}}^2}} \end{bmatrix}$$

OpenMx

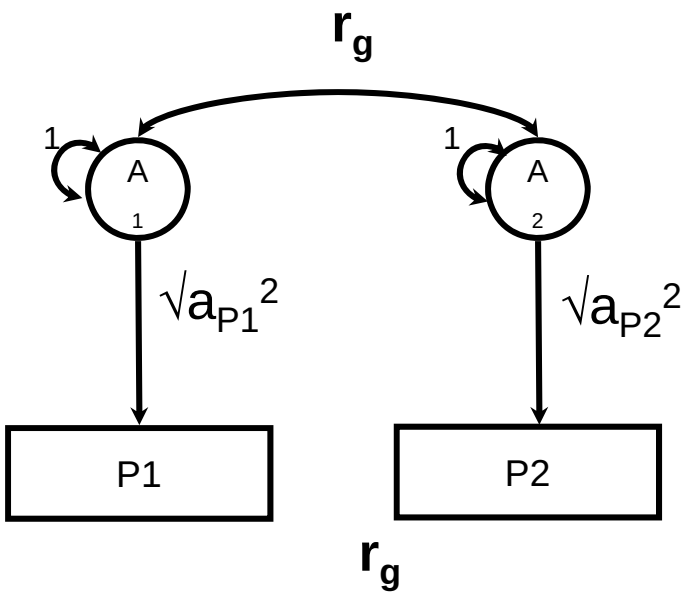
```
corA <- mxAlgebra(name ="rA", expression = solve(sqrt(I*A))%*%A%*%solve(sqrt(I*A)))
```

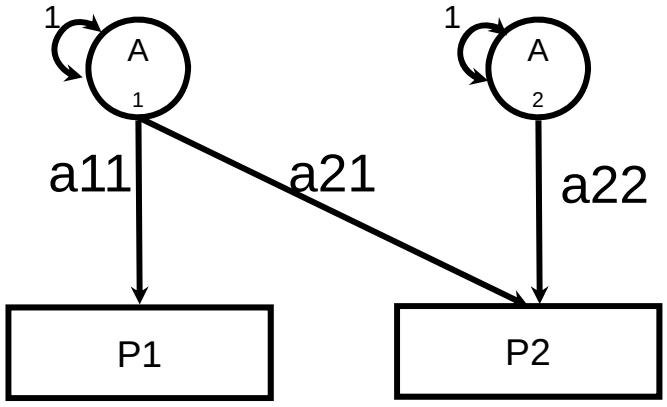
Contribution to phenotypic correlation



➢ The contribution to the phenotypic correlation is a function of both heritabilities and the r_g

Contribution to phenotypic correlation





Proportion of $r_{P1,P2}$ due to additive genetic factors:

$$\frac{\left(\sqrt{a_{P1}^2} * r_g * \sqrt{a_{P2}^2}\right)}{r_{P1,P2}}$$

$$\frac{a21a11}{a21a11 + c21c11 + e21e11}$$

Contribution to phenotypic correlation

OpenMx

```
ACEcovMatrices <- c("A","C","E","V","A/V","C/V","E/V")
ACEcovLabels <-
("covComp_A","covComp_C","covComp_E","Var","stCovComp_A","stCovComp_C","stCovComp_E")
formatOutputMatrices(CholACEFit,ACEcovMatrices,ACEcovLabels,Vars,4)
```

```
[1] "Matrix ACE.A/ACE.V"
    stCovComp_A1 stCovComp_A2
LP1 0.2491       0.4784
LP2 0.4784       0.2673
```

Proportion of the phenotypic correlation due to genetic effects

```
[1] "Matrix ACE.C/ACE.V"
    stCovComp_C1 stCovComp_C2
LP1 0.2441       0.2417
LP2 0.2417       0.0292
```

$$\begin{matrix} c^2_{P1} & C_{COV} \\ C_{COV} & c^2_{P2} \end{matrix}$$

Proportion of the phenotypic correlation due to shared environmental effects

```
[1] "Matrix ACE.E/ACE.V"
    stCovComp_E1 stCovComp_E2
LP1 0.5069       0.2799
LP2 0.2799       0.7035
```

$$\begin{matrix} e^2_{P1} & E_{COV} \\ E_{COV} & e^2_{P2} \end{matrix}$$

Proportion of the phenotypic correlation due to unshared environmental effects

Summary / Interpretation

- Genetic correlation (r_g) = the correlation between two latent genetic factors
  - High genetic correlation = large overlap in genetic effects on the two phenotypes

- Contribution of genes to phenotypic correlation = The proportion of the phenotypic correlation explained by the overlapping genetic factors
  - This is a function of the r_g and the heritabilities of the two traits

Outline

- 11.00-12.30
  - Lecture Bivariate Cholesky Decomposition
  - Practical Bivariate analysis of IQ and attention problems
- 12.30-13.30 LUNCH
- 13.30-15.00
  - Lecture Multivariate Cholesky Decomposition
  - Practical Tri- and Four-variate analysis of IQ, educational attainment and attention problems

Practical

- Replicate findings from Kuntsi et al.

- 126 MZ and 126 DZ twin pairs from Netherlands Twin Register
- Age 12

- FSIQ
- Attention Problems (AP) [mother-report]

Practical – exercise 1

- Script CholeskyBivariate.R
- Dataset Cholesky.dat

- Run script **up to saturated model**

Practical – exercise 1

- Fill in the table with correlations:

MZ	FSIQ1	AP1	FSIQ2	AP2
FSIQ1	1			
AP1		1		
FSIQ2			1	
AP2				1

DZ	FSIQ1	AP1	FSIQ2	AP2
FSIQ1	1			
AP1		1		
FSIQ2			1	
AP2				1

Practical – exercise 1 - Questions

- Are correlations similar to those reported by Kuntsi et al.?
- What is the phenotypic correlation between FSIQ and AP?
- What are the MZ and DZ cross-twin cross-trait correlations?
- What are your expectations for the common aetiological influences?
  - Are they familial?
  - If yes, are they genetic or shared environmental?

Practical – exercise 2

- Run **Bivariate ACE model** in the script
- Look whether you understand the output. If not, ask us!
- Adapt the first submodel such that you drop all C
- Compare fit of AE model with ACE model

Script: CholeskyBivariate.R

Practical – exercise 2

- Fill in the table with fit statistics:

	-2LL	df	chi2	Δdf	P-value
ACE model			-	-	-
AE model					

- Question:
  - Is C significant?

Practical – exercise 3

- Now try to fill in the estimates for all paths in the path model (grey boxes):