# Introduction to Multivariate Genetic Analysis

#### **Danielle Posthuma & Meike Bartels**

# **Aim and Rationale**

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 shared environmental factors (C or E)
- Can use information on multiple traits from twin pairs to partition covariation into genetic and environmental components

# **Example 1**

Why do traits correlate/covary? How can we explain the association? Additive genetic factors (r<sub>G</sub>) Shared environment (r<sub>C</sub>) Non-shared environment (r<sub>E</sub>)

Kuntsi et al. (2004) *Am J Med Genet B*, 124:41



# Example 2

- Associations between phenotypes over time
  - Does anxiety in childhood lead to depression in adolescence?
- How can we explain the association?
  - Additive genetic factors (a<sub>21</sub>)
  - Shared environment (c<sub>21</sub>)
  - Non-shared environment (e<sub>21</sub>)
  - How much is not explained by prior anxiety?
- Rice et al. (2004) BMC Psychiatry 4:43



# **Sources of Information**

- As an example: two traits measured in twin pairs
- Interested in:
  - Cross-trait covariance within individuals
  - Cross-trait covariance between twins
  - MZ:DZ ratio of cross-trait covariance between twins

# **Observed Covariance Matrix**

		Tw	in 1	Twin 2			
		Phenotype 1	Phenotype 2	Phenotype 1	Phenotype 2		
Twin 1	Phenotype 1	Variance P1					
	Phenotype 2	Covariance P1-P2	Variance P2				
Twin 2	Phenotype 1	Within-trait P1	Cross-trait	Variance P1			
	Phenotype 2	Cross-trait	Within-trait P2	Covariance P1-P2	Variance P2		

# **Observed Covariance Matrix**

		Twin 1		Twin 2			
		Phenotype 1	Phenotype 2	Phenotype 1	Phenotype 2		
		Within-twin	covariance				
Twin 1	Phenotype 1	Variance P1					
	Phenotype 2	Covariance P1-P2	Variance P2				
				Within-twin	covariance		
win 2	Phenotype 1	Within-trait P1	Cross-trait	Variance P1			
-	Phenotype 2	Cross-trait	Within-trait P2	Covariance P1-P2	Variance P2		

# **Observed Covariance Matrix**

		Twin 1		Twin 2			
		Phenotype 1	Phenotype 2	Phenotype 1	Phenotype 2		
		Within-twin	covariance				
Twin 1	Phenotype 1	Variance P1					
	Phenotype 2	Covariance	Variance P2				
		Cross-twin	covariance	Within-twin covariance			
win 2	Phenotype 1	Within-trait P1	Cross-trait	Variance P1			
-	Phenotype 2	Cross-trait	Within-trait P2	Covariance P1-P2	Variance P2		

# **SEM: Cholesky Decomposition**



# **SEM: Cholesky Decomposition**



# **SEM: Cholesky Decomposition**



# Why Fit This Model?

- Covariance matrices must be positive definite
- If a matrix is positive definite, it can be decomposed into the product of a triangular matrix and its transpose:

A = a\*a<sup>T</sup>

- Many other multivariate models possible
  - Depends on data and hypotheses of interest

# **Cholesky Decomposition**

# **Path Tracing**























		Twin 1		Twin 2			
		Phenotype 1	Phenotype 2	Phenotype 1	Phenotype 2		
		Within-twin	covariance				
vin 1	Phenotype 1	a <sub>11</sub> <sup>2</sup> +c <sub>11</sub> <sup>2</sup> +e <sub>11</sub> <sup>2</sup>					
ŕ	Phenotype 2	a <sub>11</sub> a <sub>21</sub> +c <sub>11</sub> c <sub>21</sub> + e <sub>11</sub> e <sub>21</sub>	$a_{22}^{2}+a_{21}^{2}+c_{22}^{2}+c_{21}^{2}+c_{21}^{2}+c_{21}^{2}+c_{21}^{2}$				
		Cross-twin	covariance	Within-twin covariance			
win 2	Phenotype 1	1/.5a <sub>11</sub> <sup>2</sup> +c <sub>11</sub> <sup>2</sup>		a <sub>11</sub> <sup>2</sup> +c <sub>11</sub> <sup>2</sup> +e <sub>11</sub> <sup>2</sup>			
-	Phenotype 2	1/.5a <sub>11</sub> a <sub>21</sub> + c <sub>11</sub> c <sub>21</sub>	$1/.5a_{22}^{2}+1/.5$ $a_{21}^{2}+c_{22}^{2}+c_{21}^{2}$	a <sub>11</sub> a <sub>21</sub> +c <sub>11</sub> c <sub>21</sub> + e <sub>11</sub> e <sub>21</sub>	$a_{22}^{2}+a_{21}^{2}+c_{22}^{2}+c_{21}^{2}+c_{21}^{2}+c_{22}^{2}+c_{21}^{2}$		









		Twin 1		Twin 2			
		Phenotype 1	Phenotype 2	Phenotype 1 Phe		Phenotype 2	henotype 2
		Within-twin	covariance				
Twin 1	Phenotype 1	Variance P1			Cross-twin covariance	cross-trait differs by	
	Phenotype 2	Covariance P1-P2	Variance P2		Zygosity		
		Cross-twin covariance		V	Vithin-twin	covariance	
win 2	Phenotype 1	Within-trait P1	Cross-trait		Variance P1		
ŕ	Phenotype 2	Cross-trait	Within-trait P2	C	Covariance P1-P2	Variance P2	











# Summary

- Within-individual cross-trait covariance implies common aetiological influences
- Cross-twin cross-trait covariance implies common aetiological influences are familial
- Whether familial influences genetic or environmental shown by MZ:DZ ratio of cross-twin cross-trait covariances

# **Cholesky Decomposition Bivariate Genetic analyses**

# Specification in OpenMx



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



Path Tracing:  $\Sigma_{A} = \begin{bmatrix} a_{11}^{2} & a_{11}a_{21} \\ a_{21}a_{11} & a_{21}^{2} + a_{22}^{2} \end{bmatrix}$ a Lower 2 x 2:  $\begin{bmatrix} a_{1} & a_{2} \\ P_{1} & a_{11} & 0 \\ P_{2} & a_{21} & a_{22} \end{bmatrix}$ 



$$\begin{split} \Sigma_{A} &= a^{*} a^{T} \\ \Sigma_{A} &= a^{\%} {}^{*} {}^{\%} t(a) \end{split} = \begin{bmatrix} a_{11} & 0 \\ a_{21} & a_{22} \end{bmatrix} {}^{*} \begin{bmatrix} a_{11} & a_{21} \\ 0 & a_{22} \end{bmatrix} \\ &= \begin{bmatrix} a_{11}^{2} + 0 \times 0 & a_{11}a_{21} + 0 \times a_{22} \\ a_{21}a_{11} + 0 \times a_{22} & a_{21}^{2} + a_{22}^{2} \end{bmatrix} \end{split}$$



## **Total Within-Twin Covar.**

$$\Sigma_{A} = a \% *\% t(a) = \begin{bmatrix} a_{11}^{2} & a_{11}a_{21} \\ a_{21}a_{11} & a_{21}^{2} + a_{22}^{2} \end{bmatrix} \Sigma_{C} = c \% *\% t(c) = \begin{bmatrix} c_{11}^{2} & c_{11}c_{21} \\ c_{21}c_{11} & c_{21}^{2} + c_{22}^{2} \end{bmatrix}$$
$$\Sigma_{E} = e \% *\% t(e) = \begin{bmatrix} e_{11}^{2} & e_{11}e_{21} \\ e_{21}e_{11} & e_{21}^{2} + e_{22}^{2} \end{bmatrix}$$

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

$$\Sigma_{V} = \Sigma_{A} + \Sigma_{C} + \Sigma_{E}$$

$$\Sigma_{V} = \begin{bmatrix} a_{11}^{2} + c_{11}^{2} + e_{11}^{2} & a_{11}a_{21} + c_{11}c_{21} + e_{11}e_{21} \\ a_{21}a_{11} + c_{21}c_{11} + e_{11}e_{21} & a_{21}^{2} + a_{22}^{2} + c_{21}^{2} + c_{22}^{2} + e_{21}^{2} + e_{22}^{2} \end{bmatrix}$$

# **OpenMx Matrices & Algebra**

```
OpenMx
multACEModel <- mxModel("multACE", ``</pre>
                                      mxModel("ACE",
# Matrices a, c, and e to store a, c, and e path coefficients
mxMatrix( type="Lower", nrow=nv, ncol=nv, free=TRUE, values=.6, name="a"),
mxMatrix( type="Lower", nrow=nv, ncol=nv, free=TRUE, values=.6, name="c"),
mxMatrix( type="Lower", nrow=nv, ncol=nv, free=TRUE, values=.6, name="e"),
# Matrices A, C, and E compute variance components mxAlgebra(
expression=a %*% t(a), name="A" ),
mxAlgebra( expression=c %*% t(c), name="C" ),
mxAlgebra( expression=e %*% t(e), name="E" ),
# Algebra to compute total variances and standard deviations (diagonal only)
mxAlgebra( expression=A+C+E, name="V" ),
mxMatrix( type="Iden", nrow=nv, ncol=nv, name="I"), mxAlgebra(
expression=solve(sqrt(I*V)), name="isd"),
```





Within-traits P11-P12 =  $0.5a_{11}^2$ P21-P22 =  $0.5a_{22}^2+0.5a_{21}^2$ 



```
Path Tracing:
```

Within-traits P11-P12 =  $0.5a_{11}^2$ P21-P22 =  $0.5a_{22}^2$ + $0.5a_{21}^2$ 

Cross-traits P11-P22 =  $0.5a_{11}a_{21}$ P21-P12 =  $0.5a_{21}a_{11}$ 

#### Additive Genetic Cross-Twin Covariance (DZ)



#### Additive Genetic Cross-Twin Covariance (MZ)



 $1 \otimes \Sigma_{A} = 1\% x\% (a\% *\% t(a))$ 

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

#### Common Environment Cross-Twin Covariance



$$1 \otimes \Sigma_{c} = 1\% x\% (c\% *\% t(c)) = \begin{bmatrix} c_{11}^{2} & c_{11}c_{21} \\ c_{21}c_{11} & (c_{21}^{2} + c_{22}^{2}) \end{bmatrix}$$

# **Covariance Model for Twin Pairs**

OpenMx					
# Algebra for expected variance/covariance matrix in MZ mxAlgebra(					
expression= rbind ( cbind(A+C+E , A+C),					
cbind(A+C , A+C+E)),					
name="expCovMZ" ),					
# Algebra for expected variance/covariance matrix in DZ, note use of 0.5, converted to 1*1 matrix					
<pre>mxAlgebra( expression= rbind ( cbind(A+C+E , 0.5%x%A+C),</pre>					
cbind(0.5%x%A+C , A+C+E)),					
name="expCovDZ" ) ),					

# Obtaining Standardised Estimates

# **Correlated Factors Solution**



- Each variable decomposed into genetic/environmental components
- Correlations across variables estimated
- Results from Cholesky can be converted to this model

#### **Genetic correlation**



# Standardized drawing or correlated factors solution



# **Standardized solution**

 A correlation coefficient is a standardized covariance that lies between -1 and 1 so that it is easier to interpret

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

#### **Covariance to Correlation**

$$r_{12} = \frac{\sigma_{12}^2}{\sqrt{\sigma_{11}^2 \times \sigma_{22}^2}} = \frac{1}{\sqrt{\sigma_{11}^2}} * \sigma_{12}^2 * \frac{1}{\sqrt{\sigma_{22}^2}}$$

#### Using matrix algebra notation:

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

#### **Genetic Correlations**

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

# Specification in OpenMx

$$\begin{bmatrix} 1 & r_G \\ r_G & 1 \end{bmatrix} = \begin{bmatrix} \frac{1}{\sqrt{\sigma_{A_{11}}^2}} & 0 \\ 0 & \frac{1}{\sqrt{\sigma_{A_{22}}^2}} \end{bmatrix} * \begin{bmatrix} \sigma_{A_{11}}^2 & \sigma_{A_{12}}^2 \\ \sigma_{A_{21}}^2 & \sigma_{A_{22}}^2 \end{bmatrix} * \begin{bmatrix} \frac{1}{\sqrt{\sigma_{A_{11}}^2}} & 0 \\ 0 & \frac{1}{\sqrt{\sigma_{A_{22}}^2}} \end{bmatrix}$$
$$= \sqrt{(I \cdot A)}^{-1} * A * \sqrt{(I \cdot A)}^{-1}$$
Where I is an identity matrix:
$$\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$
$$\text{and I.A} = \begin{bmatrix} \sigma_{A_{11}}^2 & 0 \\ 0 & \sigma_{A_{22}}^2 \end{bmatrix}$$



#### Genetic correlation & contribution to observed correlation



If the rg = 1, the two sets of genes overlap completely

If however a11 and a22 are near to zero, genes do not contribute to the observed correlation

Twin 1

The contribution to the observed correlation is a function of both heritabilities and the rg

# **Interpreting Results**

- High genetic correlation = large overlap in genetic effects on the two phenotypes
- Does it mean that the phenotypic correlation between the traits is largely due to genetic effects?
  - No: the substantive importance of a particular  $r_G$  depends the value of the correlation **and** the value of the  $\sqrt{\sigma_A^2}$  paths i.e. importance is also determined by the heritability of each phenotype

# Example



Proportion of  $r_P$  due to additive genetic factors:

 $\left(\sqrt{a_x^2} * r_g * \sqrt{a_y^2}\right) / r_p$ 



## **Standardised Results**



# **Interpretation of Correlations**

Consider two traits with a phenotypic correlation of 0.40 :

- $h_{P1}^2 = 0.7$  and  $h_{P2}^2 = 0.6$  with  $r_G = .3$
- Correlation due to additive genetic effects = ?
- Proportion of phenotypic correlation attributable to additive genetic effects = ?
- $h_{P1}^2 = 0.2$  and  $h_{P2}^2 = 0.3$  with  $r_G = 0.8$
- Correlation due to additive genetic effects = ?
- Proportion of phenotypic correlation attributable to additive genetic effects = ?

Correlation due to A:  $\sqrt{h_{P1}^2 * r_G * \sqrt{h_{P2}^2}}$ Divide by  $r_P$  to find proportion of phenotypic correlation.

# Summary

• Genetic correlation  $(r_g)$  is the correlation between two latent genetic factors

The proportion of the genetic factors to the observed correlation is a function of the r<sub>g</sub> and the heritabilities of the two traits

#### More Variables...



#### More Variables...



#### **Expanded Matrices**



## **OpenMx Parameter Matrices**



# Lunch

After lunch: practical bivariate and trivariate genetic analysis