International Statistical Genetics Workshop

IBG

Boulder, Colorado

Conor Dolan
Biological psychology, VU,
Amsterdam

Dirk Pelt
Biological psychology, VU,
Amsterdam

Multivariate twin models

- Section 1: from **uni**variate to **bi**variate
- Section 2: from **bi**variate to **multi**variate
- Section 3 (this video): independent pathway model (IPM) a.k.a. biometric model
  - Recap multivariate model
  - The common factor model
  - Applied to twins
- Section 4: the independent pathway model (continued)
- Section 5: the common pathway model

Refresher on the notation

Notation	
Phenotype (variable)	Ph (Ph_1, Ph_2) or sensible ("IQ")
Latent variable - additive genetic	A
Latent variable - shared environmental	C
Latent variable - dominance	D
Latent variable - unshared environmental	E
variance of phenotype	var(Ph) or σ_{Ph}^2
variance of A (D, C,E)	var(A) or σ_A^2 (σ_D^2 σ_C^2 σ_E^2)
standard deviation of A (D, C,E)	std(A) or σ_A (σ_D σ_C σ_E)
Covariance matrix of p phenotypes (pxp symmetric matrix) - Phenotypic	Σ_{Ph} containing p variances and ((p-1)*p)/2 covariances
Covariance matrix of p phenotypes (pxp symmetric matrix) - A (D, C, E)	Σ_A (Σ_D, Σ_C, Σ_E)
Covariance and correlation between X and Y (off-diagonal element in Σ_{Ph})	cov(X,Y) or $\sigma_{X,Y}$ or $s_{X,Y}$ cor(X,Y) or $\rho_{X,Y}$ or $r_{X,Y}$

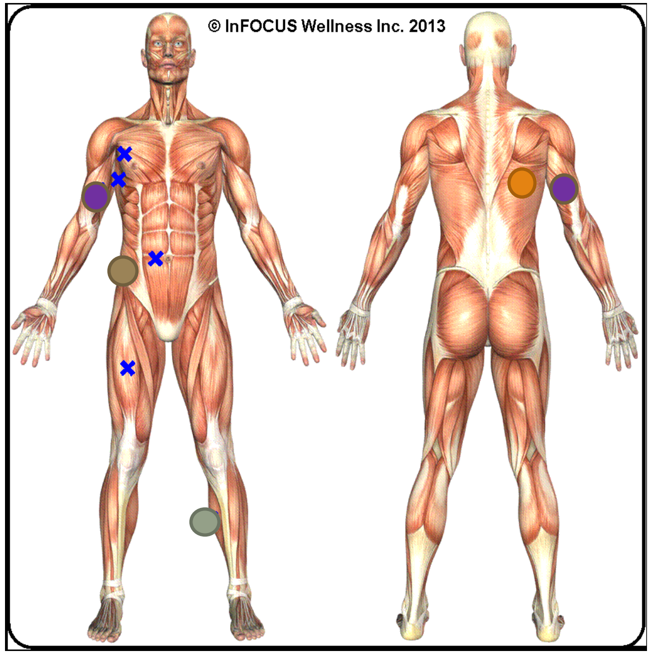
The generalization from p=1 (univariate) to p=2 (bivariate) to p>2 (multivariate).

Σ_{Ph} MZ	p phenotypes	p phenotypes
	$\Sigma_A + \Sigma_D + \Sigma_E$	$\Sigma_A + \Sigma_D$
	$\Sigma_A + \Sigma_D$	$\Sigma_A + \Sigma_D + \Sigma_E$

Σ_{Ph} DZ	p phenotypes	p phenotypes
	$\Sigma_A + \Sigma_D + \Sigma_E$	$½\Sigma_A + ¼\Sigma_D$
	$½\Sigma_A + ¼\Sigma_D$	$\Sigma_A + \Sigma_D + \Sigma_E$

p = 2 → $\Sigma_A + \Sigma_D + \Sigma_E$ = 2x2 → 4x4 matrix for twins

p = 4 → $\Sigma_A + \Sigma_D + \Sigma_E$ = 4x4 → 8x8 matrix for twins



Biceps and Triceps (purple ●), ~~Calf (green ●),~~

Subscapular (blue ●), Suprailiacal (orange ●).

Variance decomposition of Biceps, Triceps, Subscapular, and Suprailiacal (p=4)

$\Sigma_{Ph} =$

$\Sigma_A + \Sigma_D + \Sigma_E$

$4.93 + 4.97 + 2.72 = 12.62 = \sigma_{biceps}^2$

$.39 + .39 + .22 = 1$

$$\begin{matrix} \boxed{12.62} & 10.58 & 14.17 & 16.55 \\ \boxed{10.58} & 11.12 & 13.34 & 15.42 \\ 14.17 & 13.34 & 22.62 & 23.86 \\ 16.55 & 15.42 & 23.86 & 29.41 \end{matrix} =$$

$2.98 + 5.71 + 1.89 = 10.58 = \sigma_{biceps,triceps}$

$.28 + .54 + .18 = 1$

$$\begin{matrix} \boxed{4.93} & 2.98 & 4.20 & 5.43 \\ \boxed{2.98} & 2.30 & 1.88 & 2.09 \\ 4.20 & 1.88 & 6.59 & 6.60 \\ 5.43 & 2.09 & 6.60 & 8.97 \end{matrix} + \begin{matrix} \boxed{4.97} & 5.71 & 7.37 & 8.18 \\ \boxed{5.71} & 6.70 & 9.06 & 10.78 \\ 7.37 & 9.06 & 11.30 & 13.09 \\ 8.18 & 10.78 & 13.09 & 15.18 \end{matrix} + \begin{matrix} \boxed{2.72} & 1.89 & 2.61 & 2.94 \\ \boxed{1.89} & 2.12 & 2.41 & 2.55 \\ 2.61 & 2.41 & 4.73 & 4.17 \\ 2.94 & 2.55 & 4.17 & 5.25 \end{matrix}$$

Genetic and environmental correlations

R_A	bic	tri	s	si
bic	1.000	0.885	0.737	0.817
tri	0.885	1.000	0.483	0.460
s	0.737	0.483	1.000	0.858
si	0.817	0.460	0.858	1.000

R_E	bic	tri	s	si
bic	1.000	0.789	0.726	0.777
tri	0.789	1.000	0.761	0.765
s	0.726	0.761	1.000	0.836
si	0.777	0.765	0.836	1.000

R_D	bic	tri	s	si
bic	1.000	0.989	0.983	0.941
tri	0.989	1.000	1.041	1.069
s	0.983	1.041	1.000	0.999
si	0.941	1.069	0.999	1.000

!

Variance decomposition of Biceps, Triceps, Subscapular, and Suprailiacal (p=4)

Σ_{Ph} =

$\Sigma_A + \Sigma_D + \Sigma_E$

$$\begin{matrix} 12.62 & 10.58 & 14.17 & 16.55 \\ 10.58 & 11.12 & 13.34 & 15.42 \\ 14.17 & 13.34 & 22.62 & 23.86 \\ 16.55 & 15.42 & 23.86 & 29.41 \end{matrix} =$$

$$\begin{matrix} 4.93 & 2.98 & 4.20 & 5.43 \\ 2.98 & 2.30 & 1.88 & 2.09 \\ 4.20 & 1.88 & 6.59 & 6.60 \\ 5.43 & 2.09 & 6.60 & 8.97 \end{matrix} + \begin{matrix} 4.97 & 5.71 & 7.37 & 8.18 \\ 5.71 & 6.70 & 9.06 & 10.78 \\ 7.37 & 9.06 & 11.30 & 13.09 \\ 8.18 & 10.78 & 13.09 & 15.18 \end{matrix} + \begin{matrix} 2.72 & 1.89 & 2.61 & 2.94 \\ 1.89 & 2.12 & 2.41 & 2.55 \\ 2.61 & 2.41 & 4.73 & 4.17 \\ 2.94 & 2.55 & 4.17 & 5.25 \end{matrix}$$

We have estimated the 4x4 unconstrained covariance matrices Σ_A, Σ_D, and Σ_E

But we can estimate the matrices Σ_A, Σ_D, and Σ_E subject to constraints, subject to a model

The common factor model

APPLIED TO PHENOTYPIC DATA

Modelling covariance matrices - phenotypic common factor model

6 indicators of the Big 5 personality trait Openness to Experience (OtE) measured in 300 students. Correlation matrix:

```
      o1    o2    o3    o4    o5    o6
o1 1.000 0.258 0.325 0.130 0.119 0.173
o2 0.258 1.000 0.503 0.246 0.416 0.264
o3 0.325 0.503 1.000 0.202 0.236 0.290
o4 0.130 0.246 0.202 1.000 0.281 0.243
o5 0.119 0.416 0.236 0.281 1.000 0.132
o6 0.173 0.264 0.290 0.243 0.132 1.000
```

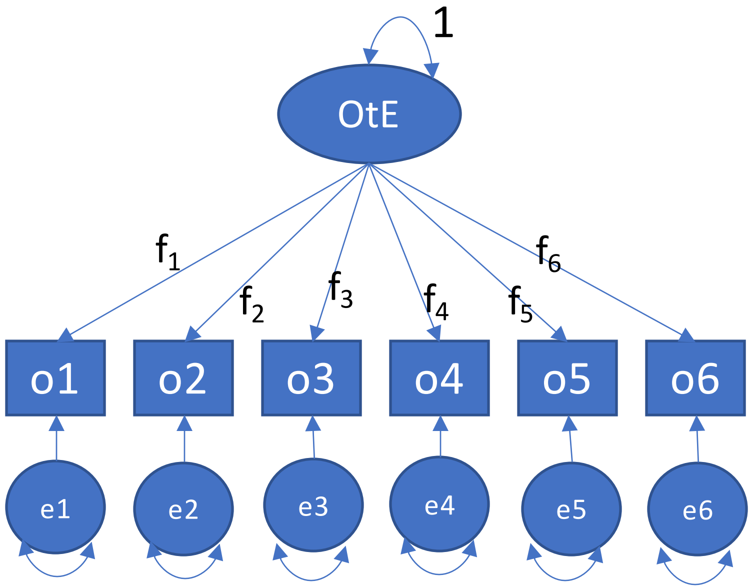
Psychometric theory of the common factor model

The common factor model is a covariance (correlation) structure model

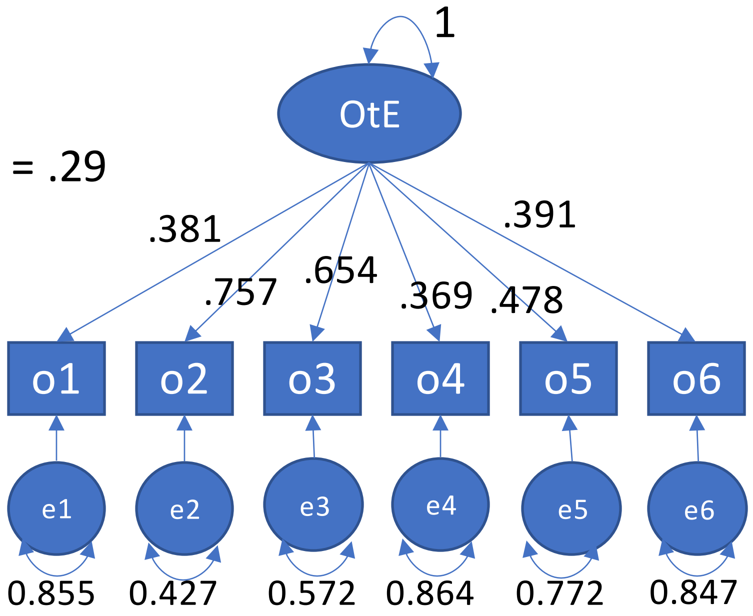
This accounts for the linear relations (covariances, correlations) between variables by positing a single common variable as its source

Single common factor model: regression with >3 dependents and 1 latent predictor

The model accounts for the observed correlation or covariance matrix.
Just like a twin model accounts for the observed covariance matrix.



$r_{o1,o2} = .381 * .757 = .29$



Consider o1 and OtE - related by linear regression:

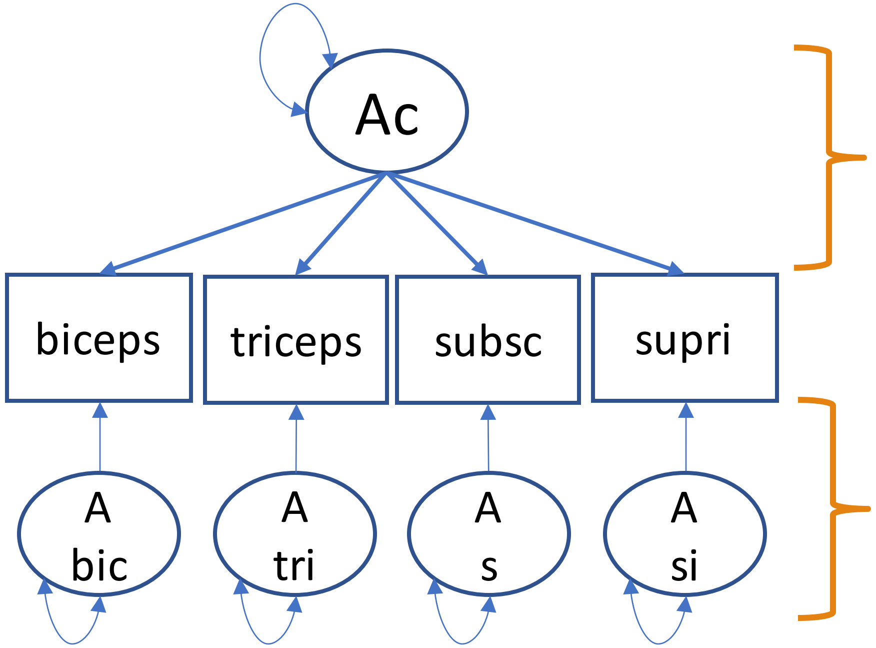
$o1 = f_1 * OtE + e1$ ..... regression model
$var(o1) = f_1^2 * var(OtE) + var(e1) = 1$ ..... decomposition of variance
$var(o1) = .381^2 * 1 + .855$ ..... variance as modelled
$f_1^2 * var(OtE) / var(o1) = R^2$ ..... proportion of explained variance

The common factor model

APPLIED TO $\Sigma_A + \Sigma_D + \Sigma_E$

From phenotypic to genetic covariance/correlation matrices

Single common factor model - a phenotype (psychometric) model (Σ_{Ph})
Apply the same model: Single common factor model - a possible model for Σ_A



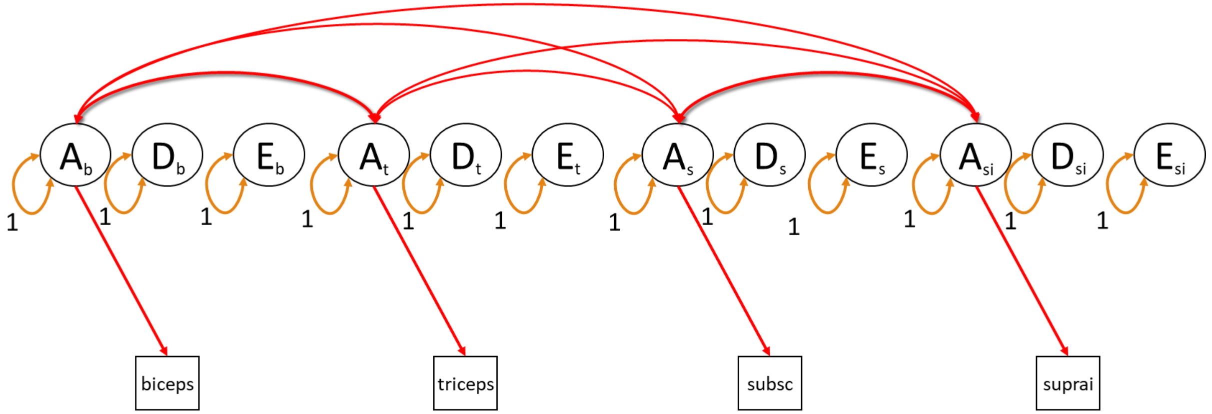
Note. Subscript c in Ac stands for "common"

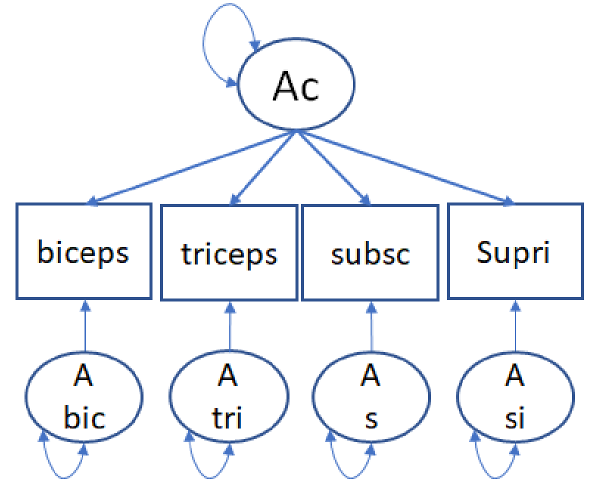
Genetic theory of the common factor model

there are genes with common additive genetic effects on all 4 skinfold phenotypes

there are genes with phenotype specific additive genetic effects - unique influences on the 4 phenotypes are a source of variance, but not covariance

From… → To …



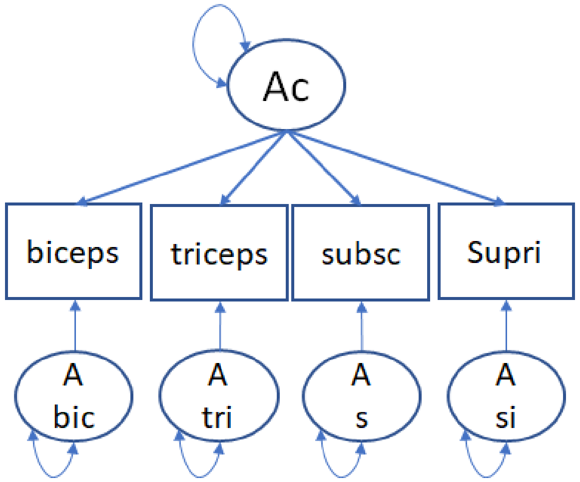


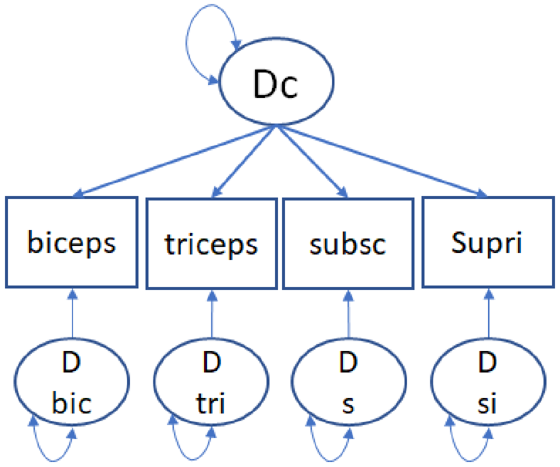
$$\Sigma_A \quad \begin{array}{lcccc} & \text{bic} & \text{tri} & \text{s} & \text{si} \\ \text{bic} & 4.93 & 2.98 & 4.20 & 5.43 \\ \text{tri} & 2.98 & 2.30 & 1.88 & 2.09 \\ \text{s} & 4.20 & 1.88 & 6.59 & 6.60 \\ \text{si} & 5.43 & 2.09 & 6.60 & 8.97 \end{array}$$

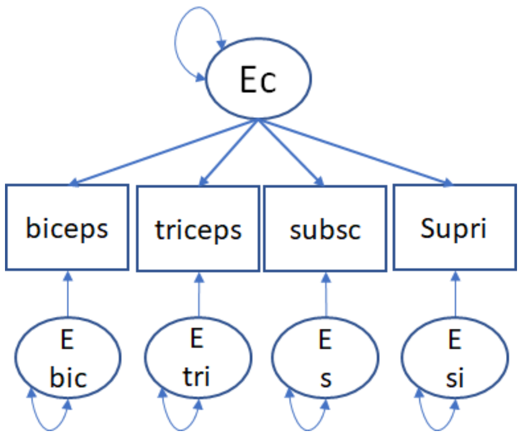
Replace the unconstrained matrices Σ_A, Σ_D, and Σ_E by the constrained counterparts … constrained according to the common factor model

Σ_A, unconstrained contains 4*5/2 = 10 parameters

Σ_A, constrained contains 4+4 = 8 parameters







Σ_A

	bic	tri	s	si
bic	4.93	2.98	4.20	5.43
tri	2.98	2.30	1.88	2.09
s	4.20	1.88	6.59	6.60
si	5.43	2.09	6.60	8.97

Σ_D

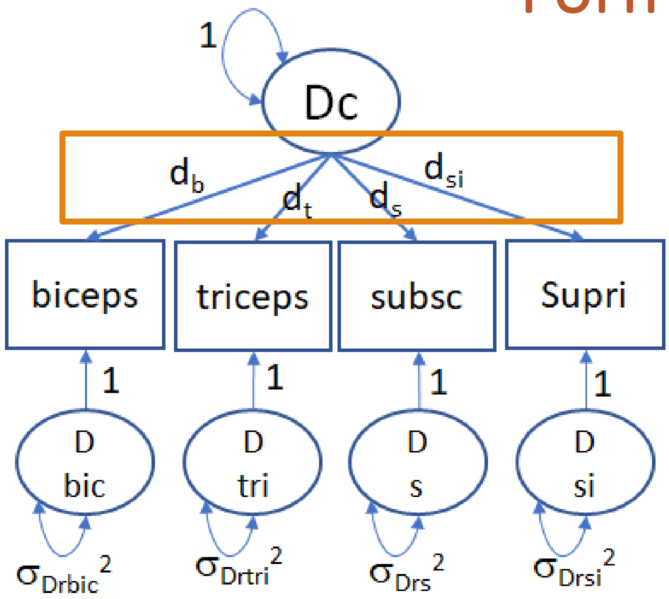
	bic	tri	s	si
bic	4.97	5.71	7.37	8.18
tri	5.71	6.70	9.06	10.78
s	7.37	9.06	11.30	13.09
si	8.18	10.78	13.09	15.18

Σ_E

	bic	tri	s	si
bic	2.72	1.89	2.61	2.94
tri	1.89	2.12	2.41	2.55
s	2.61	2.41	4.73	4.17
si	2.94	2.55	4.17	5.25

Replace the unconstrained matrices Σ_A, Σ_D, and Σ_E by the constrained counterparts
... constrained according to the common factor model for Σ_A, Σ_D and Σ_E

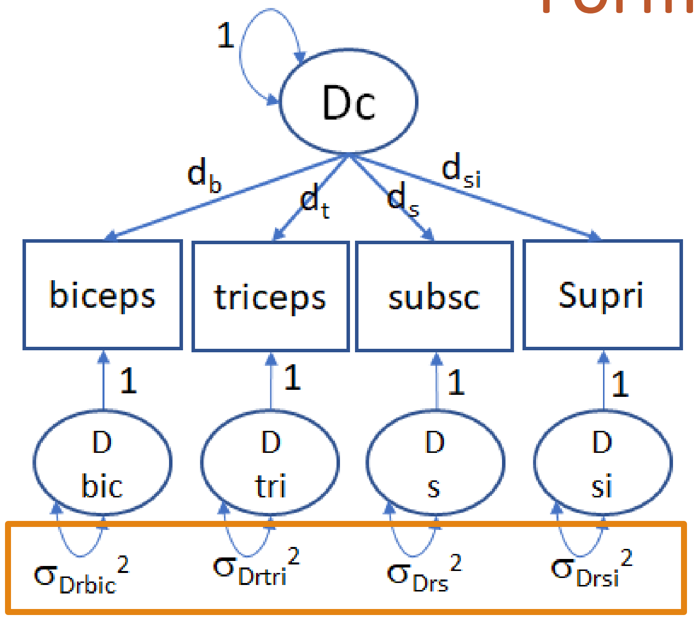
Formal representation of the model



$$L_D = \begin{pmatrix} d_b \\ d_t \\ d_s \\ d_{si} \end{pmatrix} \quad * \quad L_D^t = \begin{pmatrix} d_b & d_t & d_s & d_{si} \end{pmatrix}$$

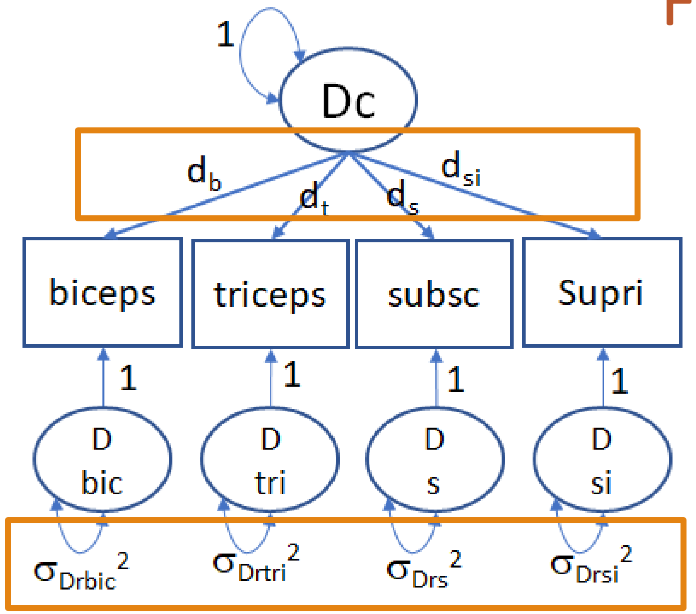
$$\Sigma_D = L_D\, L_D^t + T_D = \begin{pmatrix} d_b^2 & d_b d_t & d_b d_s & d_b d_{si} \\ d_b d_t & d_t^2 & d_t d_s & d_t d_{si} \\ d_b d_s & d_t d_s & d_s^2 & d_s d_{si} \\ d_b d_{si} & d_t d_{si} & d_s d_{si} & d_{si}^2 \end{pmatrix}$$

Formal representation of the model



$$\Sigma_D = L_D L_D^t + T_D = \begin{pmatrix} \sigma_{Drbic}^2 & 0 & 0 & 0 \\ 0 & \sigma_{Drtri}^2 & 0 & 0 \\ 0 & 0 & \sigma_{Drs}^2 & 0 \\ 0 & 0 & 0 & \sigma_{Drsi}^2 \end{pmatrix}$$

Formal representation of the model



$$L_D = \begin{pmatrix} d_b \\ d_t \\ d_s \\ d_{si} \end{pmatrix} * L_D^t = \begin{pmatrix} d_b & d_t & d_s & d_{si} \end{pmatrix} + T_D = \begin{pmatrix} \sigma_{Drbic}^2 & 0 & 0 & 0 \\ 0 & \sigma_{Drtri}^2 & 0 & 0 \\ 0 & 0 & \sigma_{Drs}^2 & 0 \\ 0 & 0 & 0 & \sigma_{Drsi}^2 \end{pmatrix}$$

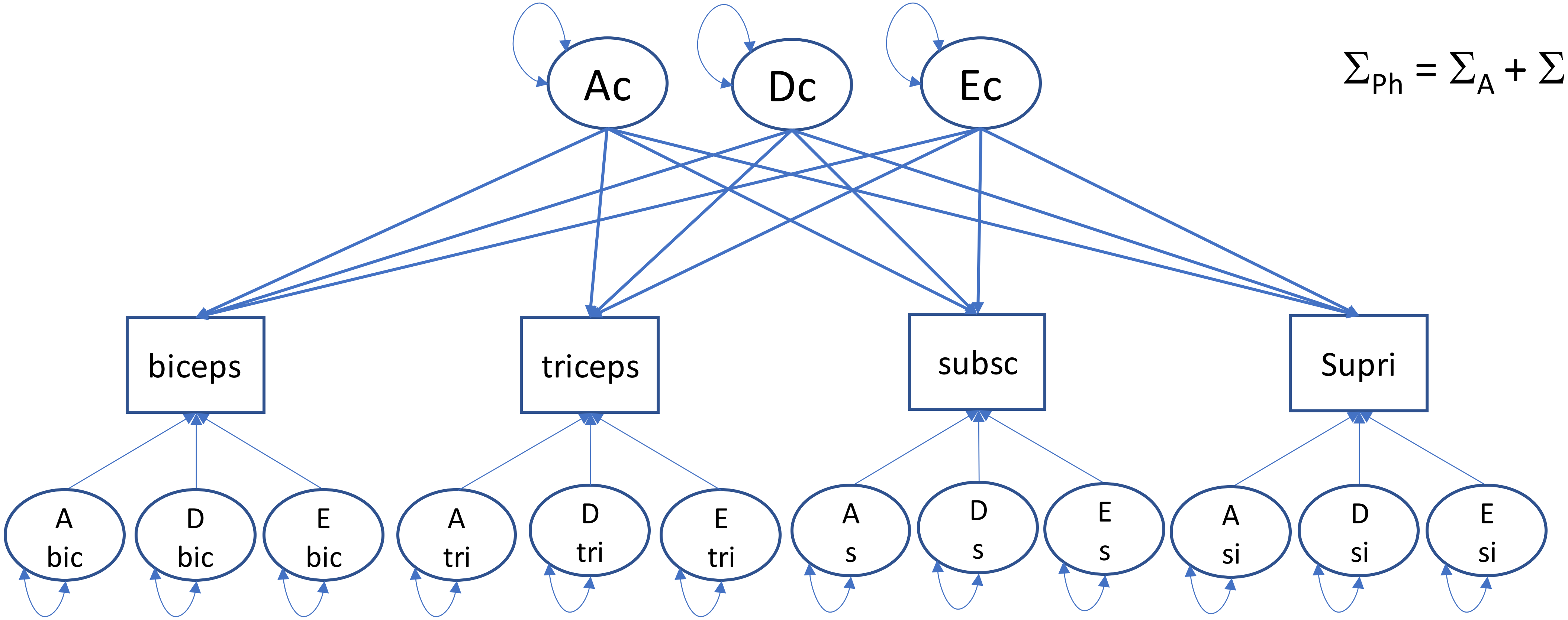
$$\Sigma_D = L_D \, L_D^t + T_D = \begin{pmatrix} d_b^2+\sigma_{Drbic}^2 & d_b d_t & d_b d_s & d_b d_{si} \\ d_b d_t & d_t^2+\sigma_{Drtri}^2 & d_t d_s & d_t d_{si} \\ d_b d_s & d_t d_s & d_s^2+\sigma_{Drs}^2 & d_s d_{si} \\ d_b d_{si} & d_t d_{si} & d_s d_{si} & d_{si}^2+\sigma_{Drsi}^2 \end{pmatrix}$$

4 factor loadings (d_b d_t d_s d_{si}) and 4 residual variances (σ_{Drbic}^2 σ_{Drtri}^2 σ_{Drs}^2 σ_{Drsi}^2) = 8 parameters

Welcome to the independent pathway model !
- common factor model applied to Σ_A, Σ_D, & Σ_E

$$\Sigma_A = L_A L_A^t + T_A$$
$$\Sigma_D = L_D L_D^t + T_D$$
$$\Sigma_E = L_E L_E^t + T_E$$

$$\Sigma_{Ph} = \Sigma_A + \Sigma_D + \Sigma_E$$

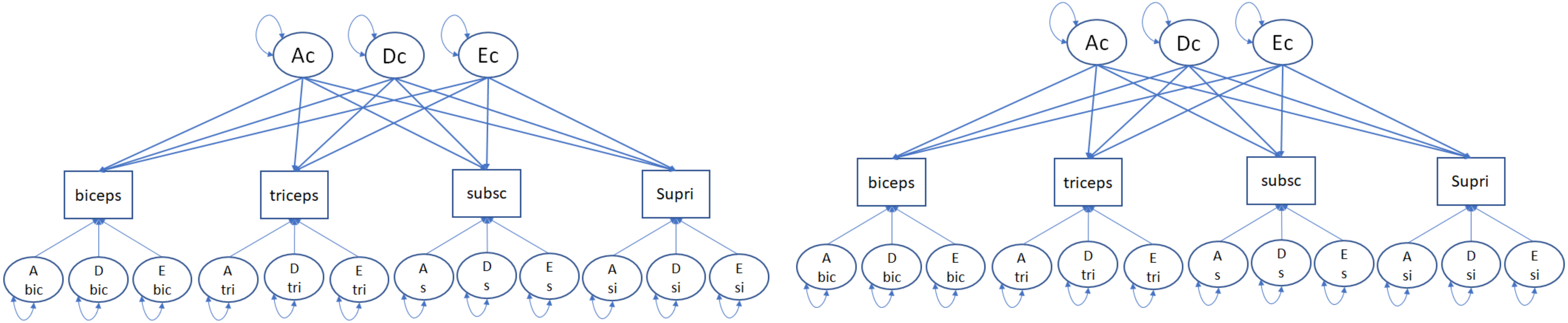


Now on to twins…

$\Sigma_A = L_A L_A^t + T_A$
$\Sigma_D = L_D L_D^t + T_D$
$\Sigma_E = L_E L_E^t + T_E$

MZ	p phenotypes	p phenotypes
	$\Sigma_A + \Sigma_D + \Sigma_E$	$\Sigma_A + \Sigma_D$
	$\Sigma_A + \Sigma_D$	$\Sigma_A + \Sigma_D + \Sigma_E$

DZ	p phenotypes	p phenotypes
	$\Sigma_A + \Sigma_D + \Sigma_E$	$.5*\Sigma_A + .25*\Sigma_D$
	$.5*\Sigma_A + .25*\Sigma_D$	$\Sigma_A + \Sigma_D + \Sigma_E$

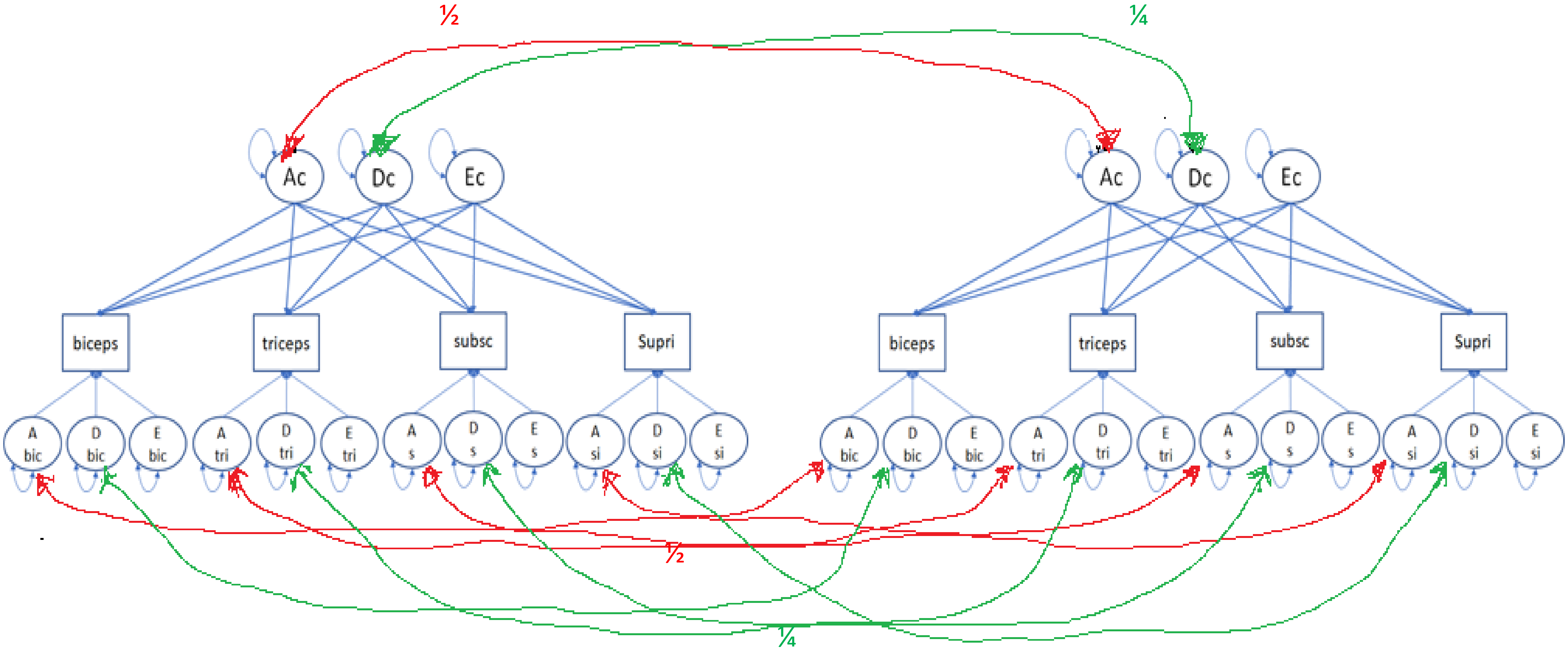


twin 1

twin 2

Connect the twin 1 and twin 2 A, D variables correctly

For DZs



½
¼
Ac
Dc
Ec
Ac
Dc
Ec
biceps
triceps
subsc
Supri
biceps
triceps
subsc
Supri
A bic
D bic
E bic
A tri
D tri
E tri
A s
D s
E s
A si
D si
E si
A bic
D bic
E bic
A tri
D tri
E tri
A s
D s
E s
A si
D si
E si
½
¼