Conor Dolan
Biological psychology, VU,
Amsterdam

International Statistical Genetics Workshop
IBG
Boulder, Colorado

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: independent pathway model (IPM) a.k.a. biometric model
- Section 4 (this video): the independent pathway model (continued)
  - Recap independent pathway model
  - Estimating and comparing models
  - Interpreting and reporting results
- 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 $r_{X,Y}$

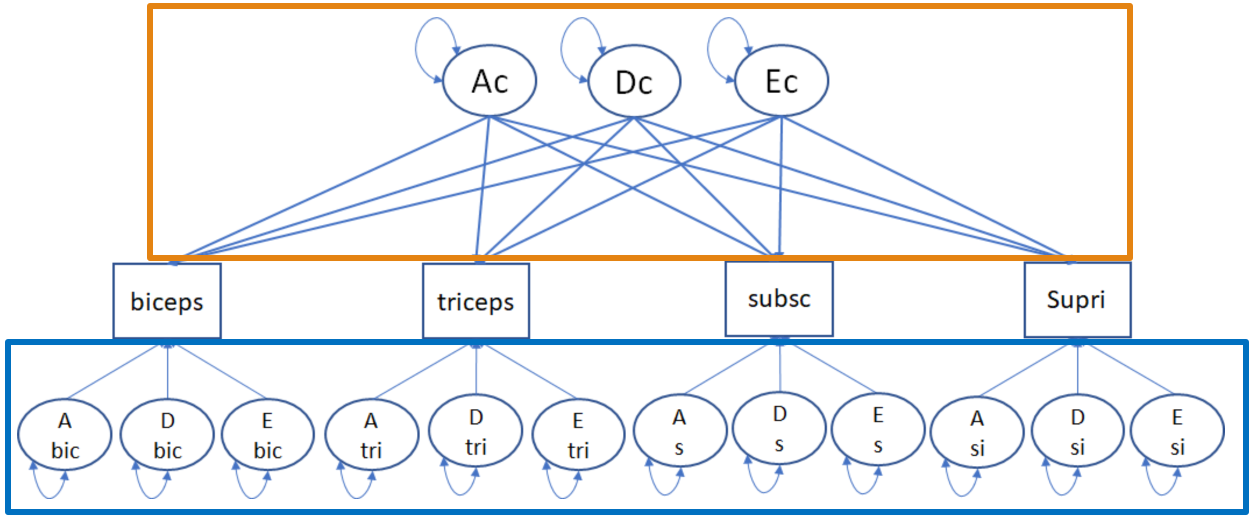
The independent pathway model

COMMON FACTOR MODEL APPLIED TO $\Sigma_A + \Sigma_D + \Sigma_E$

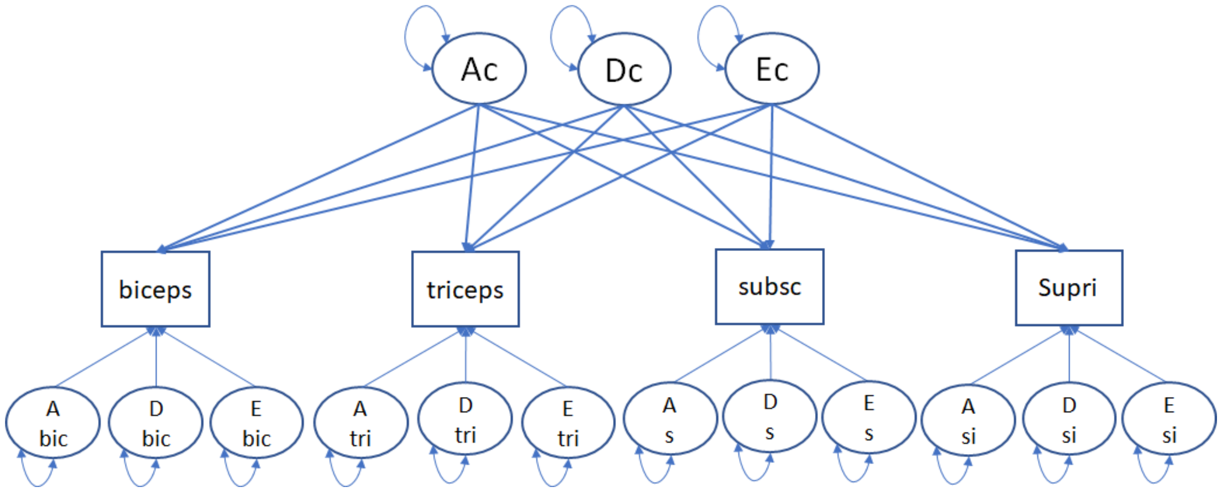
$$\Sigma_A = L_A L_A^t + T_A$$
$$\Sigma_D = L_D L_D^t + T_D$$
$$\Sigma_E = L_E L_E^t + T_E$$

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
	$\Sigma_A + \Sigma_D + \Sigma_E$	$.5*\Sigma_A + .25*\Sigma_D$
	$.5*\Sigma_A + .25*\Sigma_D$	$\Sigma_A + \Sigma_D + \Sigma_E$



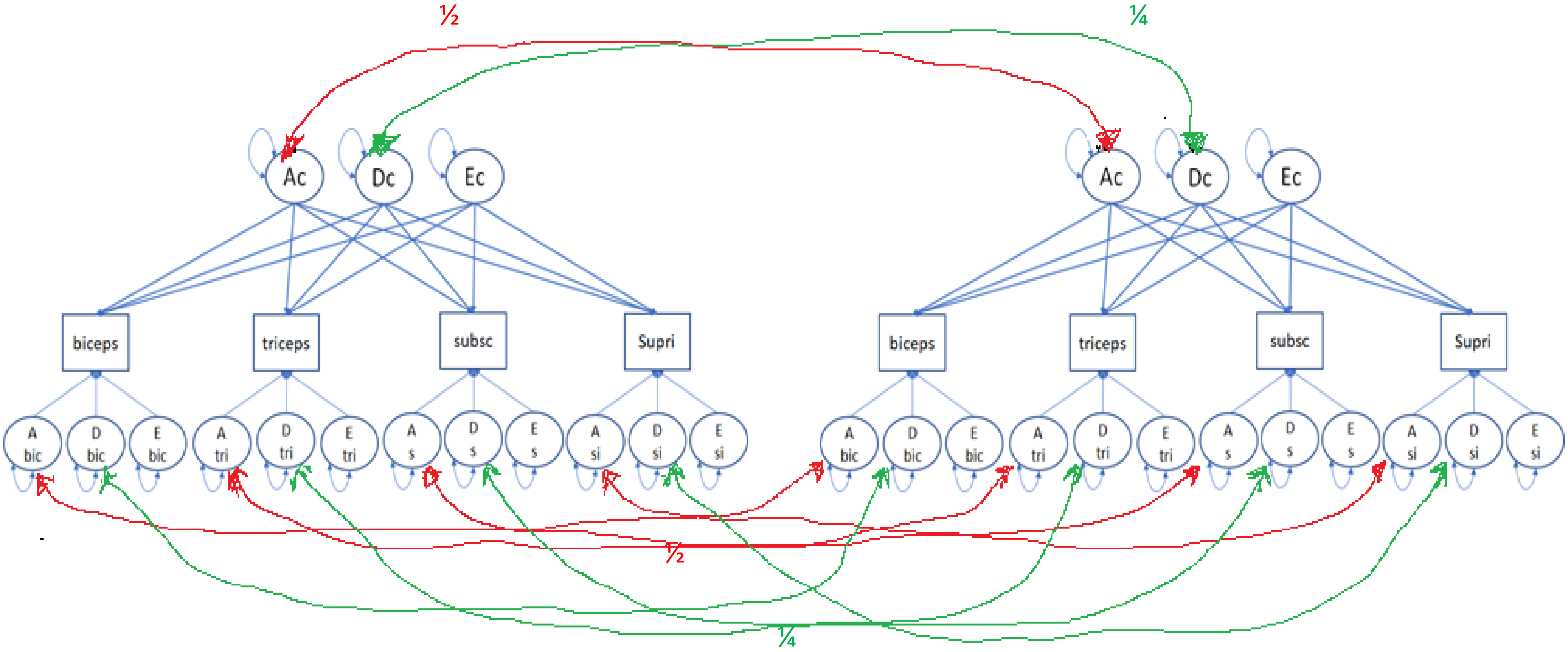
twin 1



twin 2

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

DZ twins

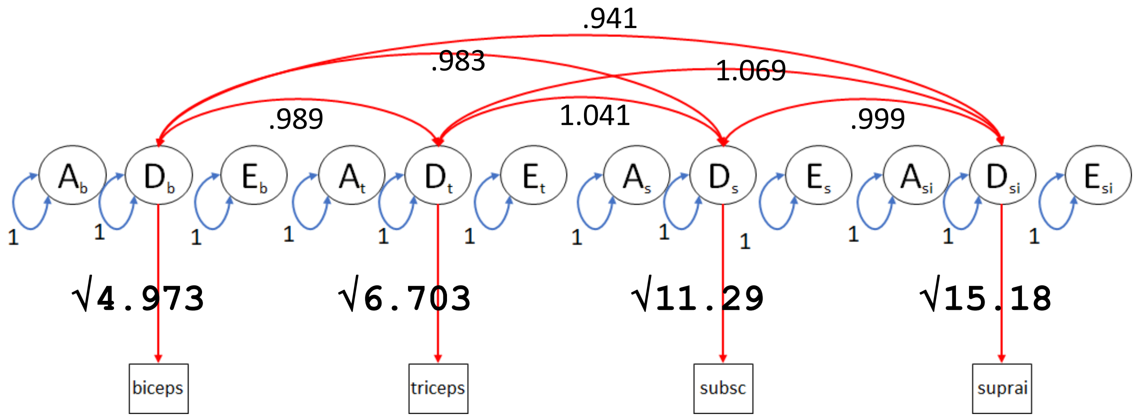

½
¼
Ac
Dc
Ec
Ac
Dc
Ec
biceps
triceps
subsc
Supri
biceps
triceps
subsc
Supri
A bic
D bic
E bic
A tri
D tri
E tri
A s
D s
E s
A si
D si
E si
A bic
D bic
E bic
A tri
D tri
E tri
A s
D s
E s
A si
D si
E si
½
¼

Estimating and comparing models

USING THE LIKELIHOOD RATIO TEST

Applying the IPM to the skinfold twin data ... in steps ... starting with Σ_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
       bic   tri     s    si
     4.973  6.703 11.29 15.18
```

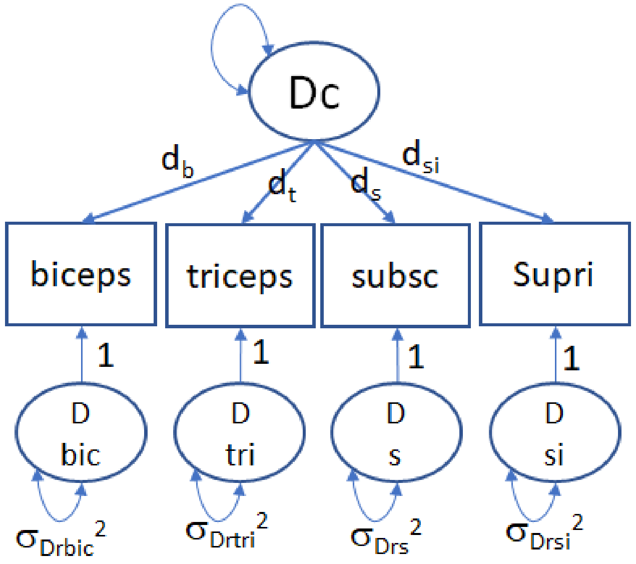


1) Fit the common factor model to Σ_D

2) Test $\sigma_{Drbic}^2 = \sigma_{Drtri}^2 = \sigma_{Drbic}^2 = \sigma_{Drbic}^2 = 0$

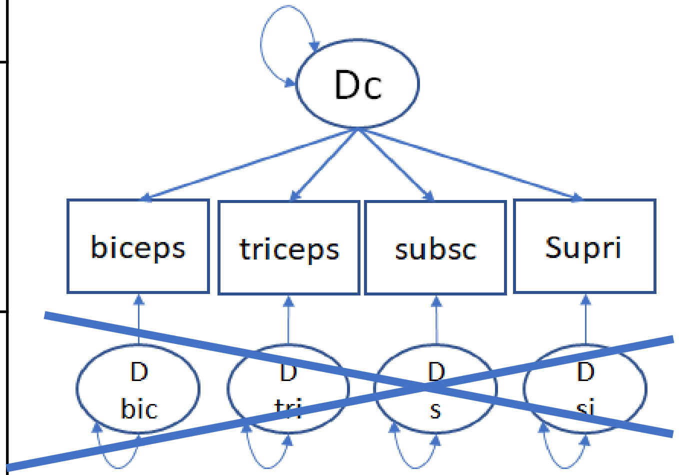
Q: Why test $\sigma_{Drbic}^2 = \sigma_{Drtri}^2 = \sigma_{Drs}^2 = \sigma_{Drsi}^2 = 0$?

Hint: What happens if the residual variances are zero



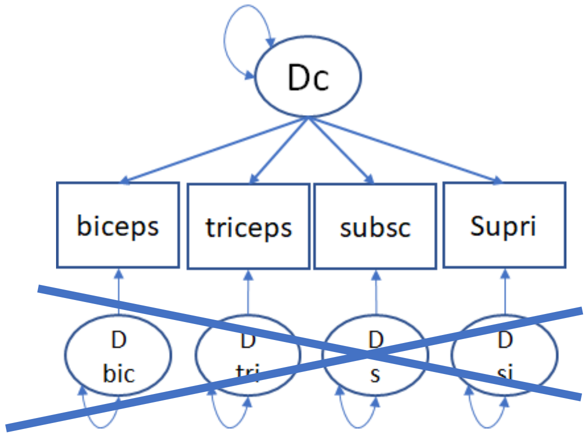
model4_1: Σ_A, Σ_D, and Σ_E - unconstrained covariance matrices,
model4_2: Σ_A, Σ_D, and Σ_E with Σ_D is subject to $\Sigma_D = L_D L_D^t + T_D$
model4_3: Σ_A, Σ_D, and Σ_E with Σ_D is subject to $\Sigma_D = L_D L_D^t$ (i.e., no residual variances $T_D = 0$)

Model	ADE model	df	-2lnL	LRT
model4_1 unconstrained	Σ_A, Σ_D, Σ_E	37	13308.8	
model4_2 D common factor model	Σ_A, $\Sigma_D = L_D L_D^t + T_D$, Σ_E	35	13309.6	4_1 vs 4_2 $T(\mathbf{2}) = 0.734$ $p = 0.693$
model4_3 D common factor model + $T_D = 0$	Σ_A, $\Sigma_D = L_D L_D^t$, Σ_E	31	13311.1	4_2 vs 4_3 $T(\mathbf{4}) = 1.567$ $p=0.816$



Model	ADE model	df	-2lnL	LRT
model4_1 unconstrained	$\Sigma_A, \Sigma_D, \Sigma_E$	37	13308.88	
model4_2 D common factor model	Σ_A, $\Sigma_D = L_D L_D^t + T_D$, Σ_E	35	13309.62	4_1 vs 4_2 $T(2) = 0.734$ $p = 0.693$
model4_3 D common factor model + $T_D = 0$	Σ_A, $\Sigma_D = L_D L_D^t$, Σ_E	31	13311.17	4_2 vs 4_3 $T(4) = 1.567$ $p = 0.816$
model4_4 E common factor model	Σ_A, $\Sigma_D = L_D L_D^t$, $\Sigma_E = L_E L_E^t + T_E$	29	13330.46	4_3 vs 4_4 $T(2) = 19.28$ $p<.0001$

model4_3



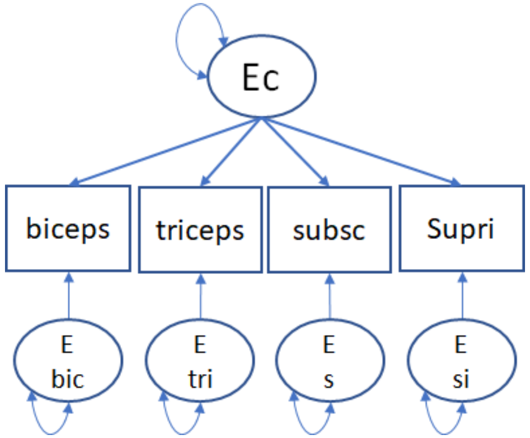
Expected cov: model4_1 Σ_E

```
      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
      bic   tri     s    si
    2.726 2.114 4.731 5.253
```

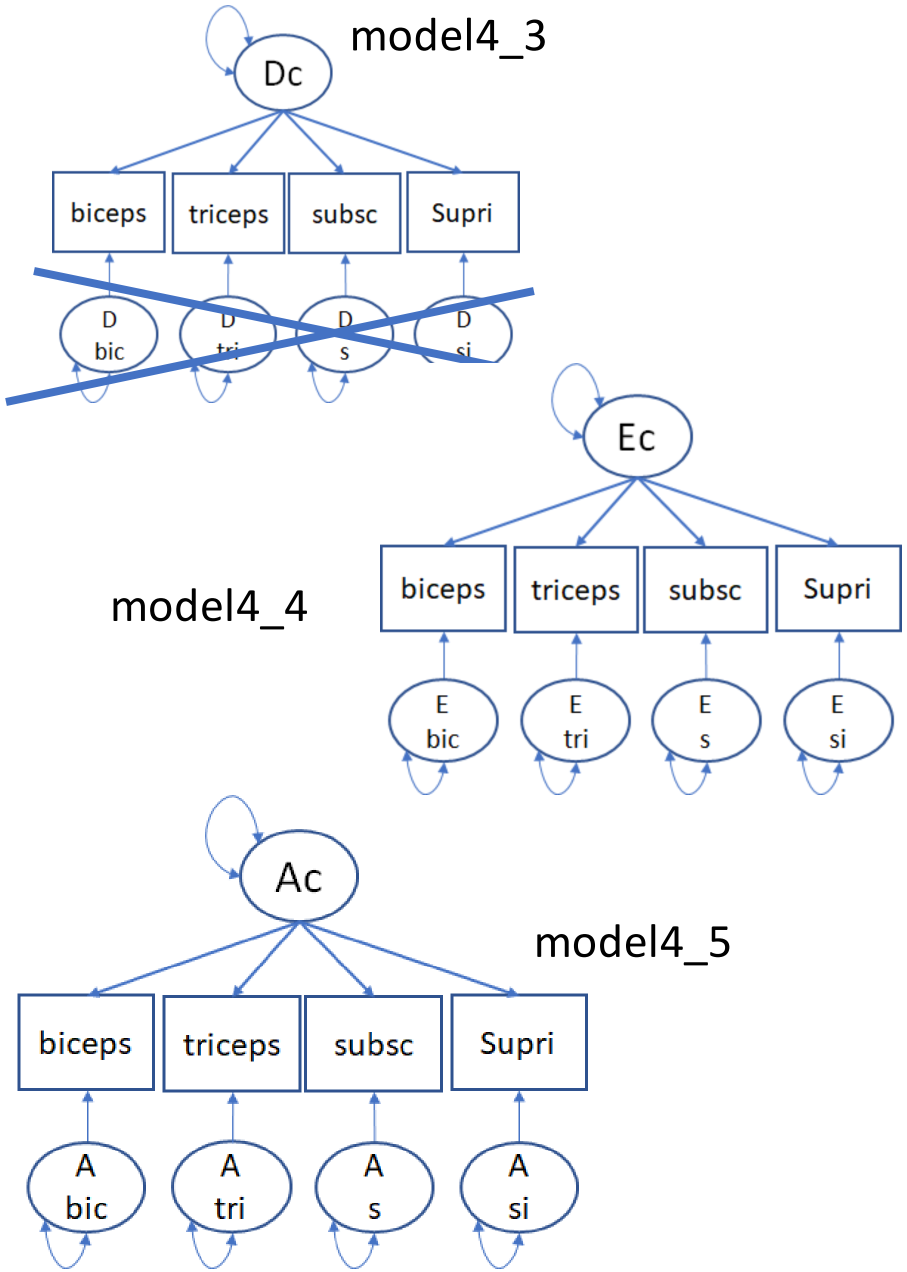
Expected cov: model4_4 $\Sigma_E = L_E L_E^t + T_E$

```
      bic   tri     s    si
bic 1.000 0.740 0.758 0.780
tri 0.740 1.000 0.771 0.794
s   0.758 0.771 1.000 0.814
si  0.780 0.794 0.814 1.000
      bic   tri     s    si
    2.739 2.123 4.726 5.230
```

model4_4

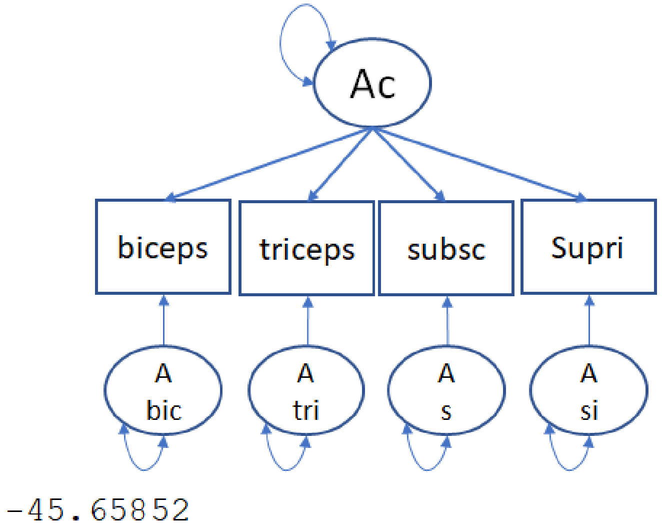


Model	ADE model	df	-2lnL	LRT
model4_1 unconstrained	$\Sigma_A, \Sigma_D, \Sigma_E$	37	13308.88	
model4_2 D common factor model	Σ_A, $\Sigma_D = L_D L_D^t + T_D$, Σ_E	35	13309.62	4_1 vs 4_2 $T(2) = 0.734$ $p = 0.693$
model4_3 D common factor model + $T_D = 0$	Σ_A, $\Sigma_D = L_D L_D^t$, Σ_E	31	13311.17	4_2 vs 4_3 $T(4) = 1.567$ $p=0.816$
model4_4 E common factor model	Σ_A, $\Sigma_D = L_D L_D^t$, $\Sigma_E = L_E L_E^t + T_E$	29	13330.46	4_4 vs 4_5 $T(2) = 19.28$ $p<0.0001$
model4_5 A common factor model	$\Sigma_A = L_A L_A^t + T_A$, $\Sigma_D = L_D L_D^t$, $\Sigma_E = L_E L_E^t + T_E$	27	13331.90	4_4 vs 4_5 $T(2) = 1.438$ $p=0.487$



Welcome to working with multivariate data! Output model4_5:

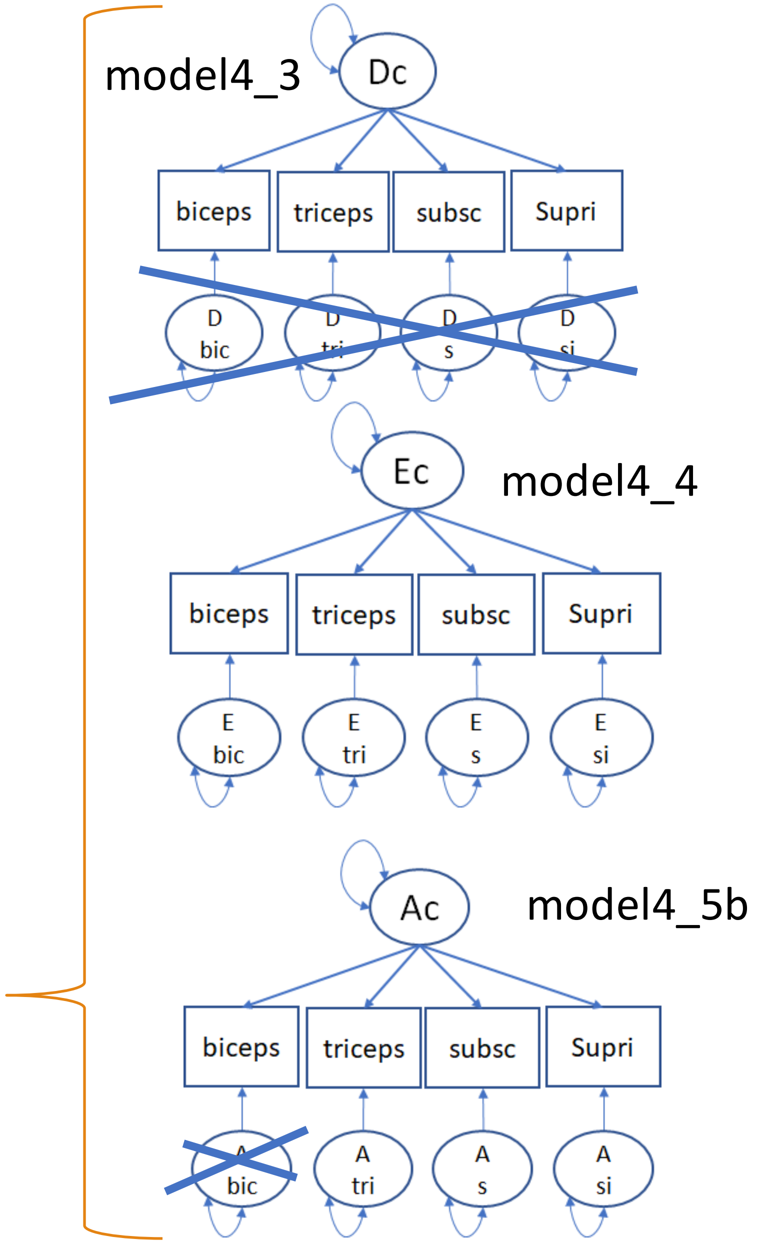
```
DiagMatrix 'TA'

$values
          [,1]     [,2]     [,3]      [,4]
[1,] -45.65852 0.000000 0.000000 0.0000000
[2,]   0.00000 1.890054 0.000000 0.0000000
[3,]   0.00000 0.000000 1.033179 0.0000000
[4,]   0.00000 0.000000 0.000000 0.9835559
```



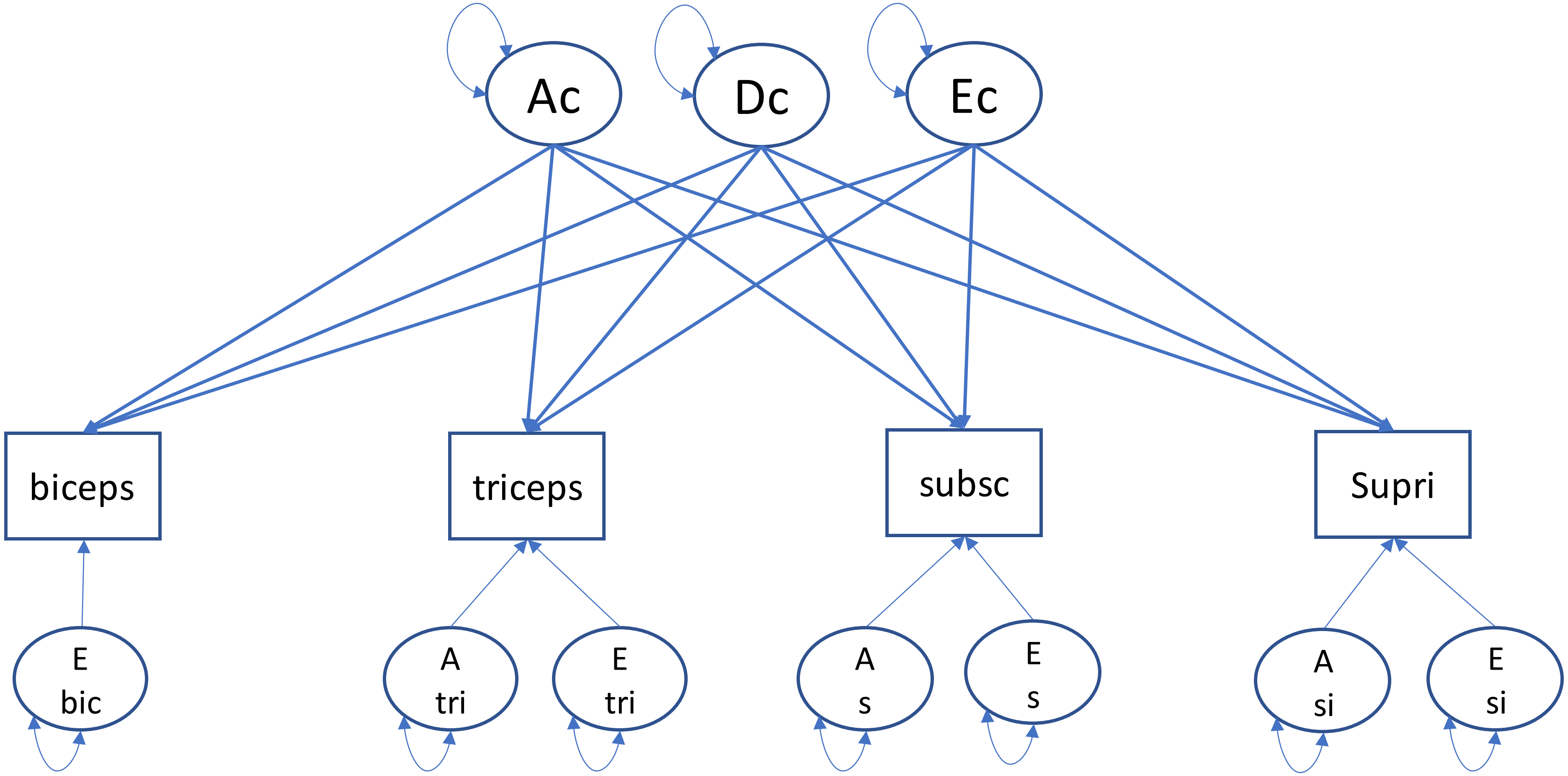
The single common A model fits, but the results cannot be trusted given the large negative residual ("Heywood case").

Possible solution: drop Abic

Model	ADE model	df	-2lnL	LRT
model 4_1 unconstrained	Σ_A, Σ_D, Σ_E	37	13308.88	
model 4_2 D common factor model	Σ_A, $\Sigma_D = L_D L_D^t + T_D$, Σ_E	35	13309.62	4_1 vs 4_2 T(2) = 0.734 *p*=0.693
model 4_3 D common factor model + $T_D = 0$	Σ_A, $\Sigma_D = L_D L_D^t$, Σ_E	31	13311.17	4_2 vs 4_3 T(4) = 1.567 *p*=0.816
model 4_4 E common factor model	Σ_A, $\Sigma_D = L_D L_D^t$, $\Sigma_E = L_E L_E^t + T_E$	29	13330.46	4_3 vs 4_4 T(2) = 19.28 *p*<0.0001
model 4_5a A common factor model	$\Sigma_A = L_A L_A^t + T_A$, $\Sigma_D = L_D L_D^t$, $\Sigma_E = L_E L_E^t + T_E$	27	13331.90	4_4 vs 4_5a T(2) = 1.438 *p*=0.487
model 4_5b A common factor model + $T_A[1,1] = 0$	$\Sigma_A = L_A L_A^t + T_A$, $\Sigma_D = L_D L_D^t$, $\Sigma_E = L_E L_E^t + T_E$	26	13333.73	4_4 vs 4_5b T(3) = 3.275 *p*=0.351
				4_1 vs 4_5b T(11) = 24.852 *p*=.009



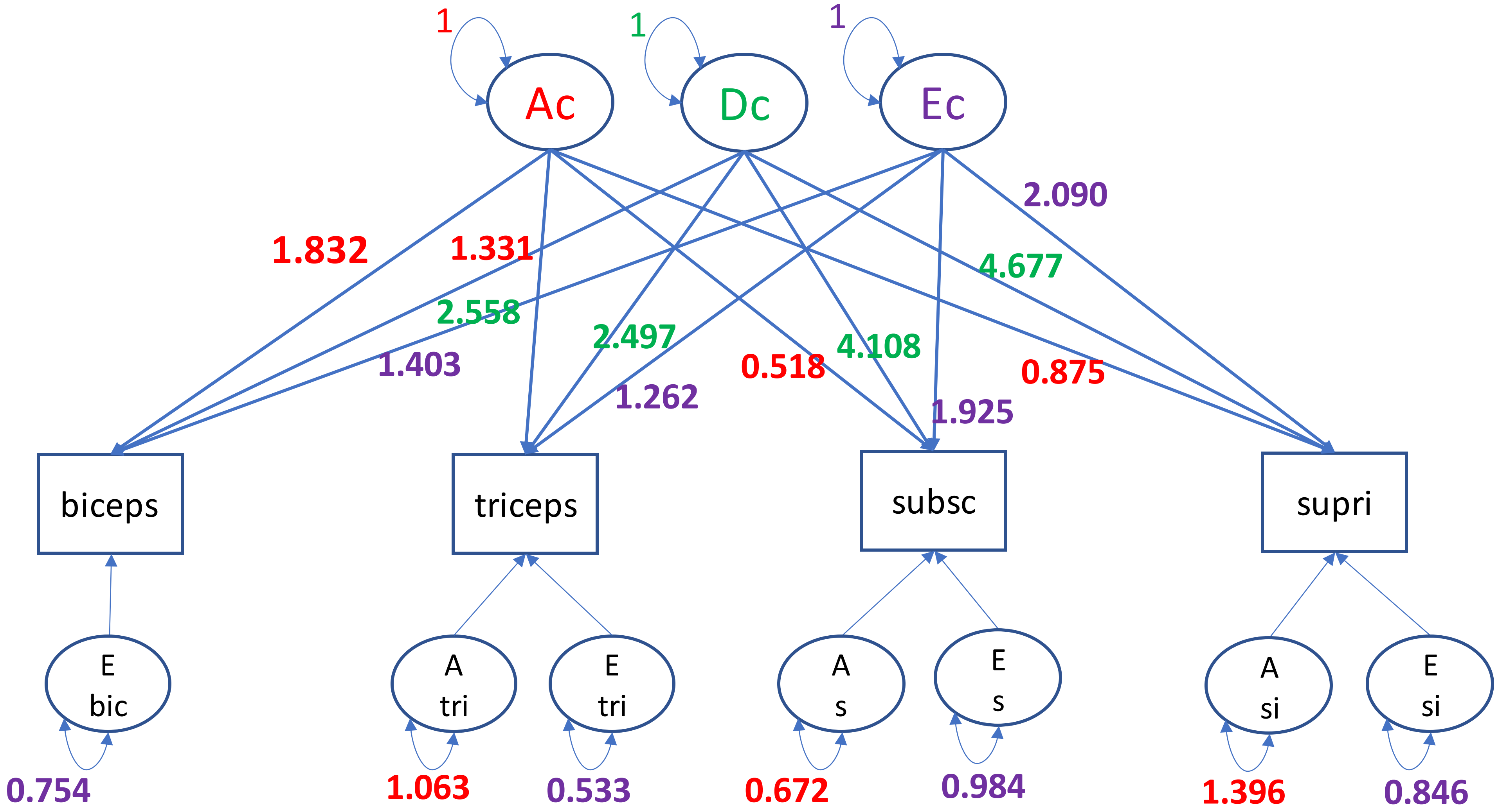
IPM model with a tweak

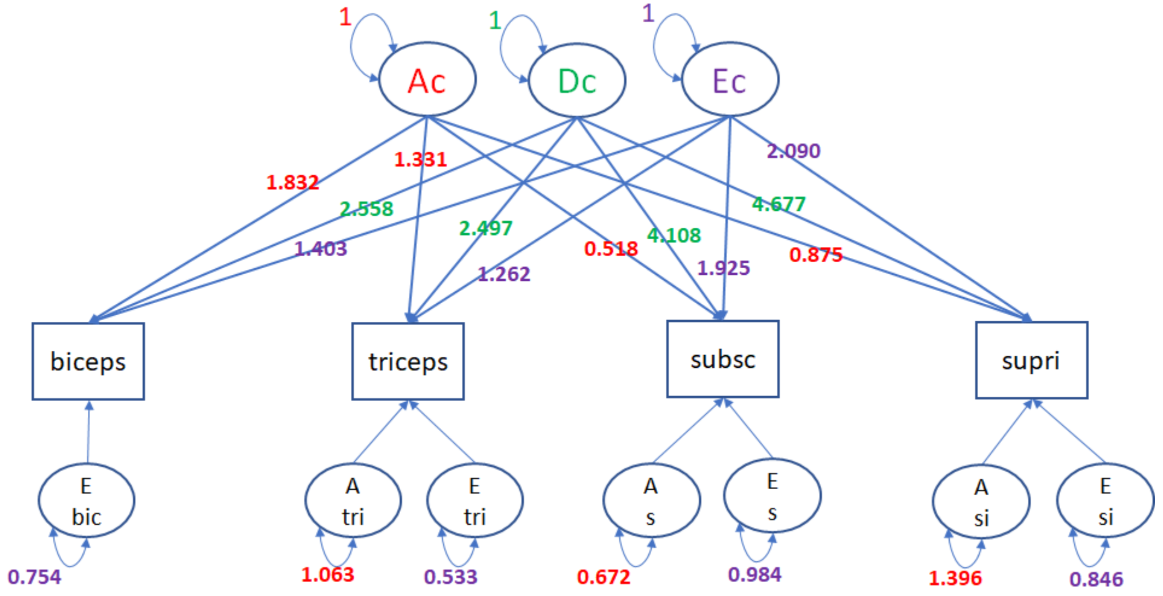


Tweaked IPM vs saturated ADE model: T(11) = 24.85 p = .0095

Interpreting and reporting results

IPM model with a tweak: Raw estimates





use the results to answers some concrete questions,
about, say, triceps:

Q. 1 what is the decomposition of variance of triceps:

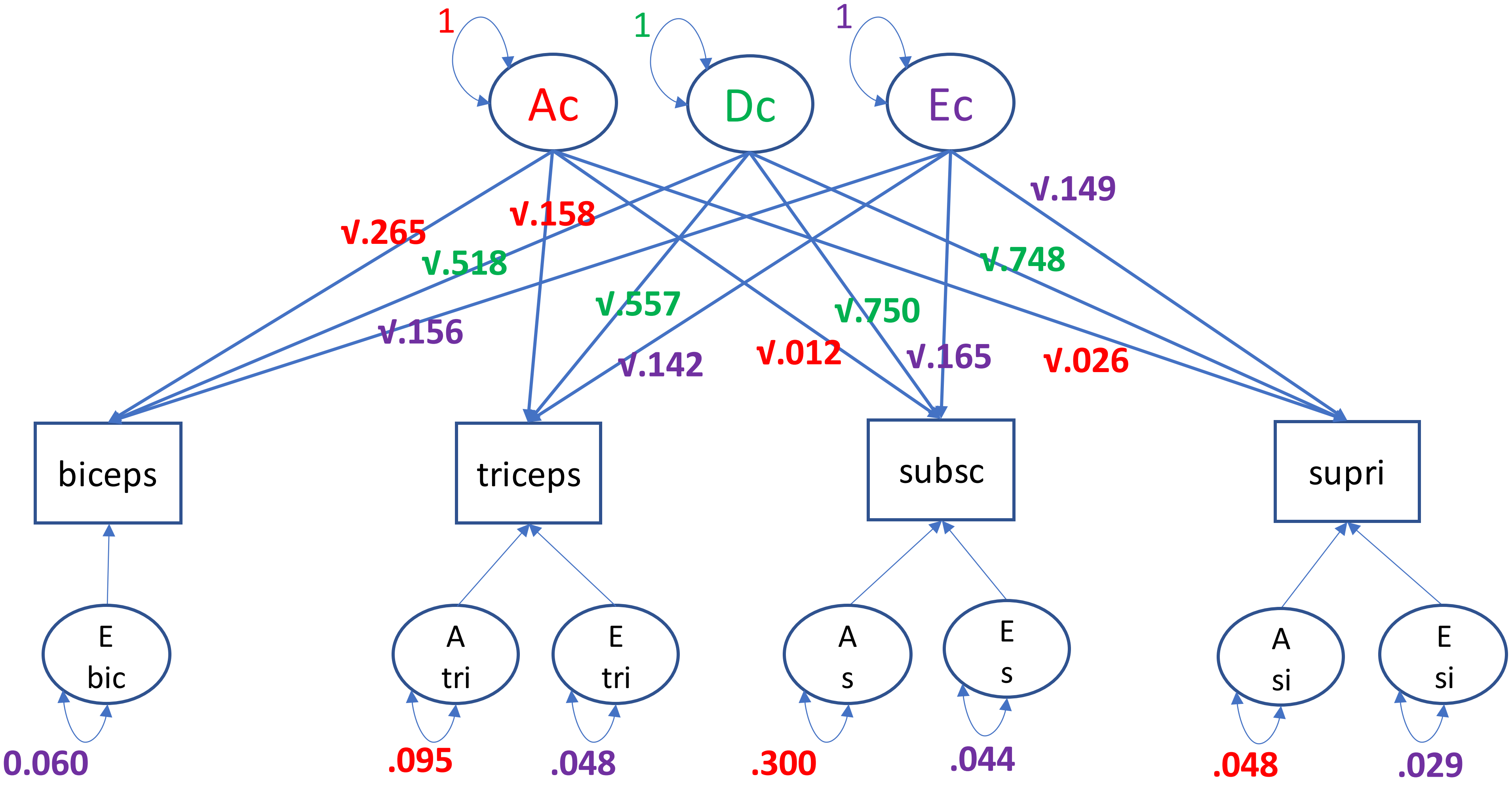
var(triceps) = (1.331^2 + 1.063) + (2.497^2) + (1.262^2+ 0.533) = 11.194
var(triceps) = (1.771+1.063) + (6.235) + (1.592+0.533) = 11.194

proportions = (1.771)/11.194+(1.063)/11.194 + (6.235)/11.194 + (1.592)/11.194 + (0.533)/11.194 = 1.00

proportions = 0.158 + 0.095 + 0.557 + 0.142 + 0.048 = 1.00

common A 15.8%
unique A 9.5%
common D 55.7%
common E 14.2%
unique E 4.8%

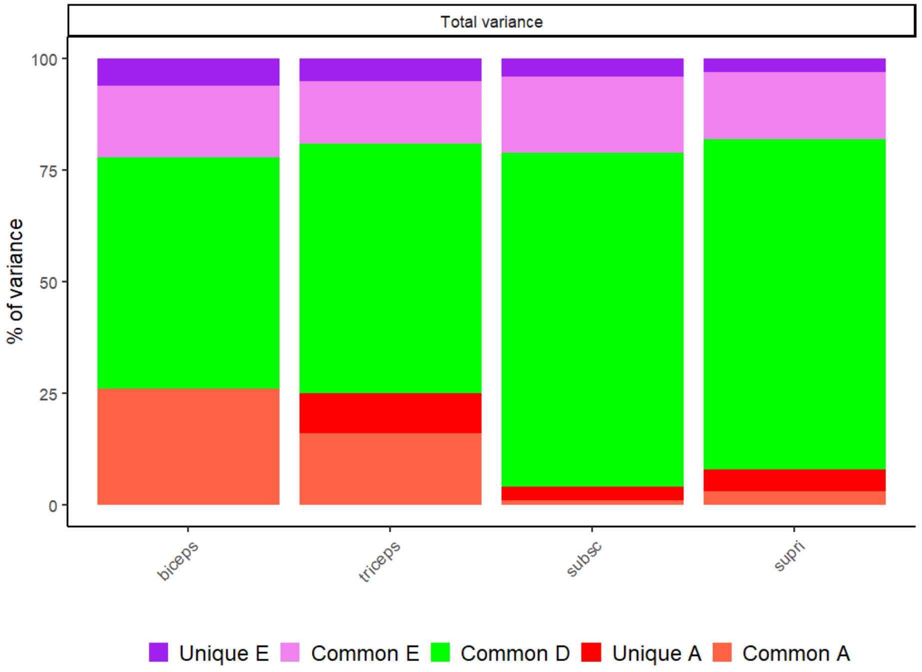
IPM model with a tweak: Standardized representation



Decomposition of variance of 4 phenotypes (standardized)

	Common A	Unique A	Common D	Common E	Unique E
biceps	0.265	0.000*	0.518	0.156	0.060
triceps	0.158	0.095	0.557	0.142	0.048
subsc	0.012	0.030	0.750	0.165	0.044
supri	0.026	0.048	0.748	0.149	0.029

* Note. This was fixed at zero.



	Total A (h_n^2)	Total D	Total E
biceps	.265 + .000 = .265	0.518	.156 + .060 = .216
triceps	.158 + .095 = .253	0.557	.142 + .048 = .190
subsc	.012 + .030 = .042	0.750	.165 + .044 = .209
supri	.026 + .048 = .074	0.748	.149 + .029 = .178

h_b^2:
.265 + .518 = .783
.253 + .557 = .810
.042 + .750 = .792
.074 + .748 = .822

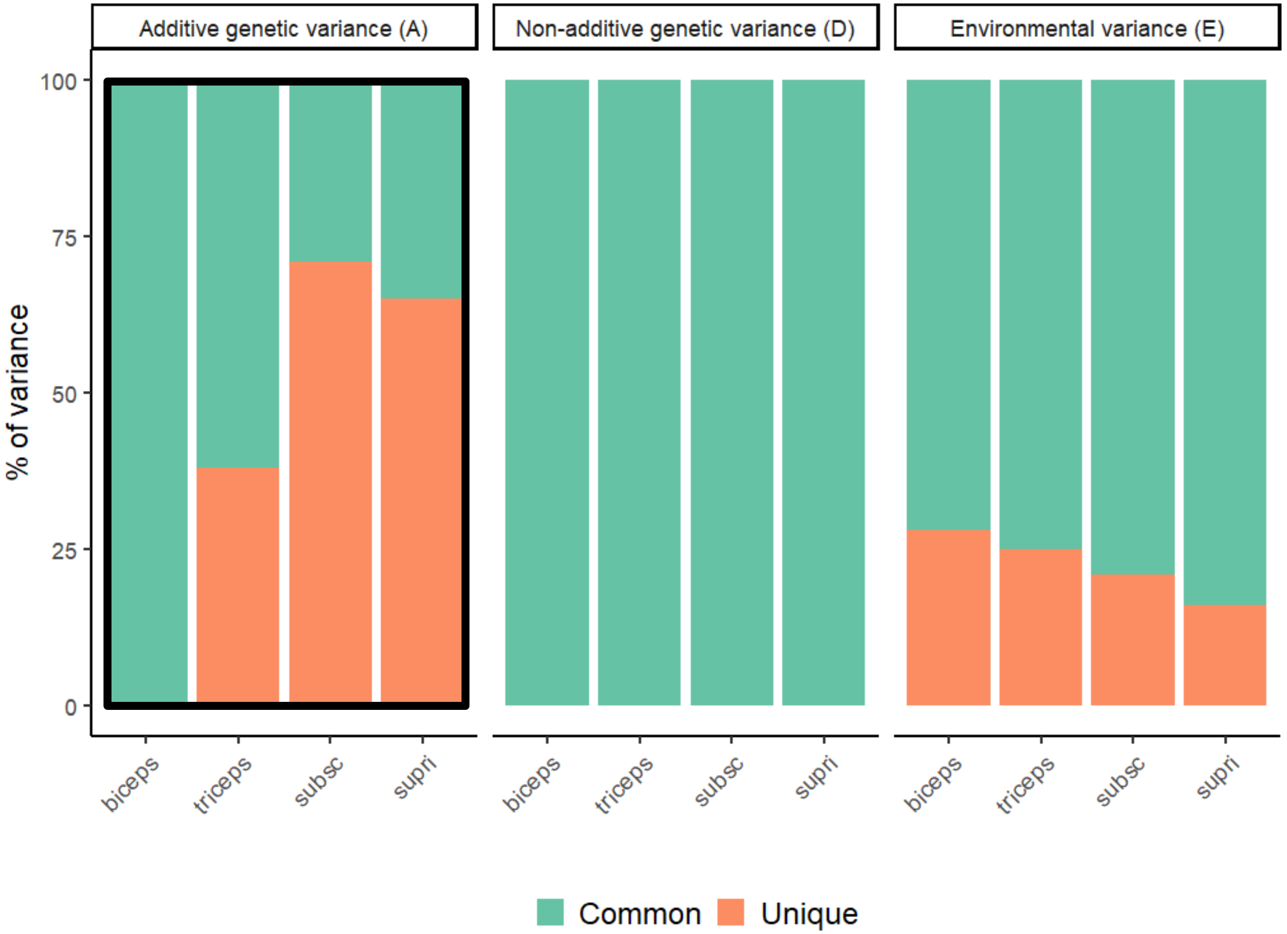
Decomposition of variance of 4 phenotypes (standardized)

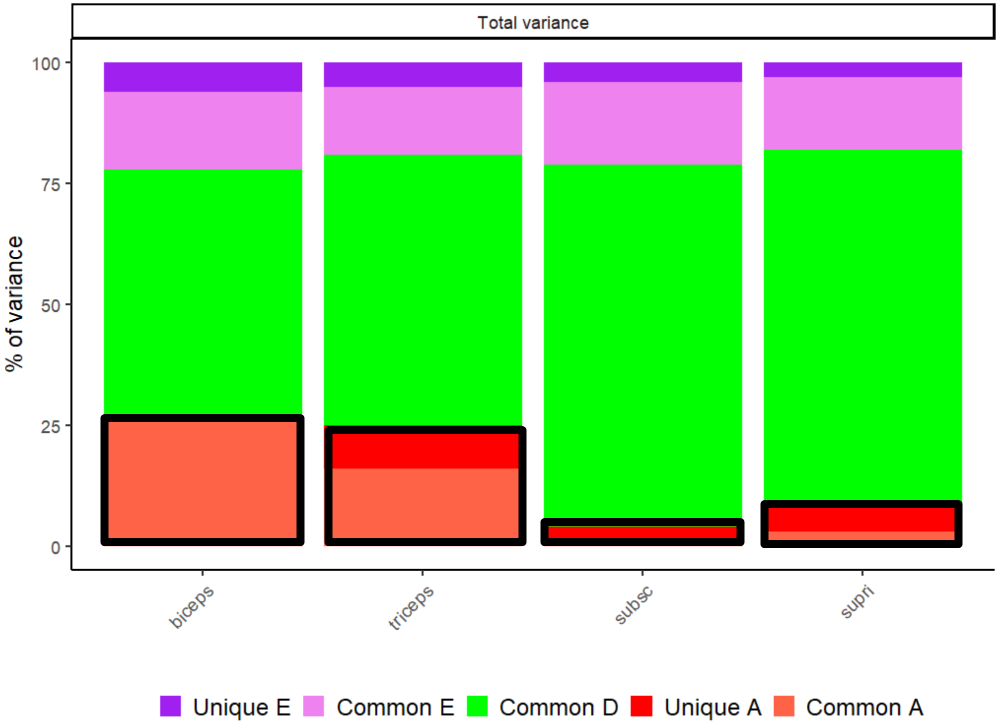
	Total A (h_n^2)	Total D	Total E
biceps	.265 + .000 = .265	0.518	.156 + .060 = .216
triceps	.158 + .095 = .253	0.557	.142 + .048 = .190
subsc	.012 + .030 = .042	0.750	.165 + .044 = .209
supri	.026 + .048 = .074	0.748	.149 + .029 = .178

	Common A	Unique A
biceps	100%	0%
triceps	62%	38%
subsc	29%	71%
supri	35%	65%

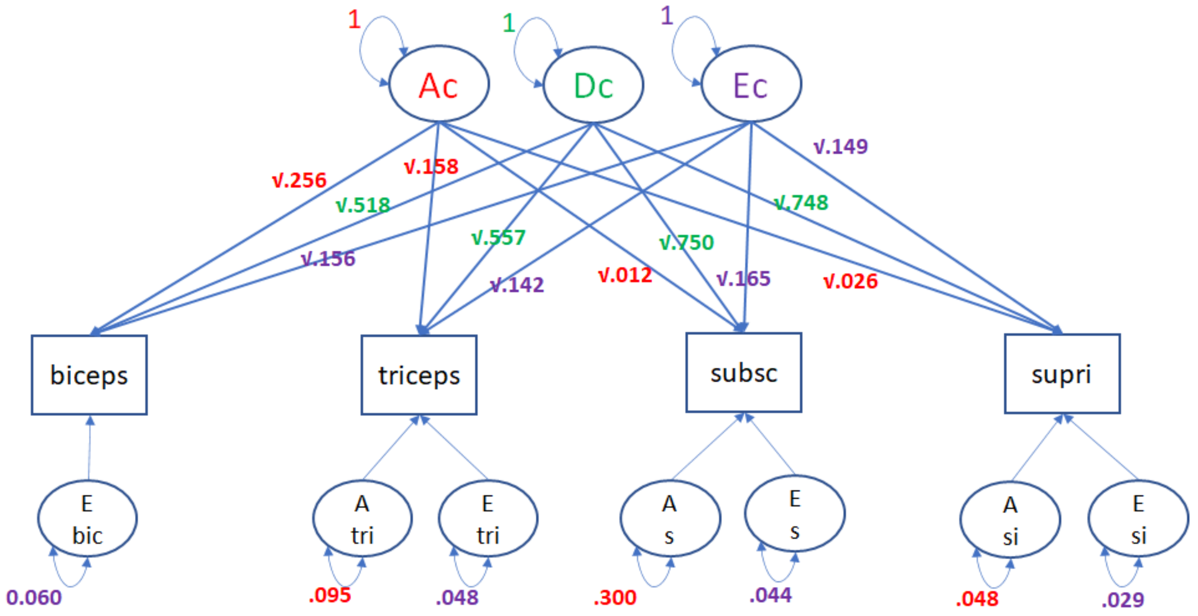
	Common D	Unique D
biceps	100%	0%
triceps	100%	0%
subsc	100%	0%
supri	100%	0%

	Common E	Unique E
biceps	72%	28%
triceps	75%	25%
subsc	79%	21%
supri	84%	16%


Additive genetic variance (A)
Non-additive genetic variance (D)
Environmental variance (E)
% of variance
100
75
50
25
0
biceps
triceps
subsc
supri
Common
Unique


Total variance
% of variance
100
75
50
25
0
biceps
triceps
subsc
supri
Unique E
Common E
Common D
Unique A
Common A

Expressing correlations in terms of the IPM



	Common A	Common D	Common E
biceps	0.265	0.518	0.156
triceps	0.158	0.557	0.142

corr(biceps, triceps) ≈ .89; in terms of the present model this is

A D E

$\sqrt{(.265 * .158)} + \sqrt{(.518 * .557)} + \sqrt{(.156 * .142)} = .89$

= = =

$0.2046 + 0.5371 + 0.1488 = .89$

$0.2046 / .89 = 0.23$ $0.5371 / .89 = 0.60$ $0.1488 / .89 = 0.17$