Twin-based direction of causation (DOC) modeling

Nathan Gillespie
Virginia Institute for Psychiatric and Behavior Genetics (VIPBG)
Virginia Commonwealth University

nathan.gillespie@vcuhealth.org

When making a causal inference e.g. the impact of cannabis (CAN) use on SZ, the **primary difficulty** is to determine the degree to which the CAN-SZ association stems from either known or unknown confounding that influences the both the RISK and an OUTCOME.

Gold Standard = randomized clinical trial (RCT):

1. The response of experimental participants 'assigned' to exposure is compared with responses of participants in a non-exposed control group
2. The 'assignment' of participants to exposure and control groups is 'random'
3. The manipulation of the exposure is controlled by the researcher

Application of RCTs to psychiatric outcomes is invariably impractical and unethical.

Need **alternative** means of inferring causality

Natural Experiments:

1. **Pre-Post Designs**

2. **Co-relative designs**
   Compares the decline in exposure-outcome associations with increasing control of genetic and shared environmental influences by using relatives with varying degrees of genetic relatedness e.g. pairs of first cousins, sib pairs including DZ and MZ twin pairs

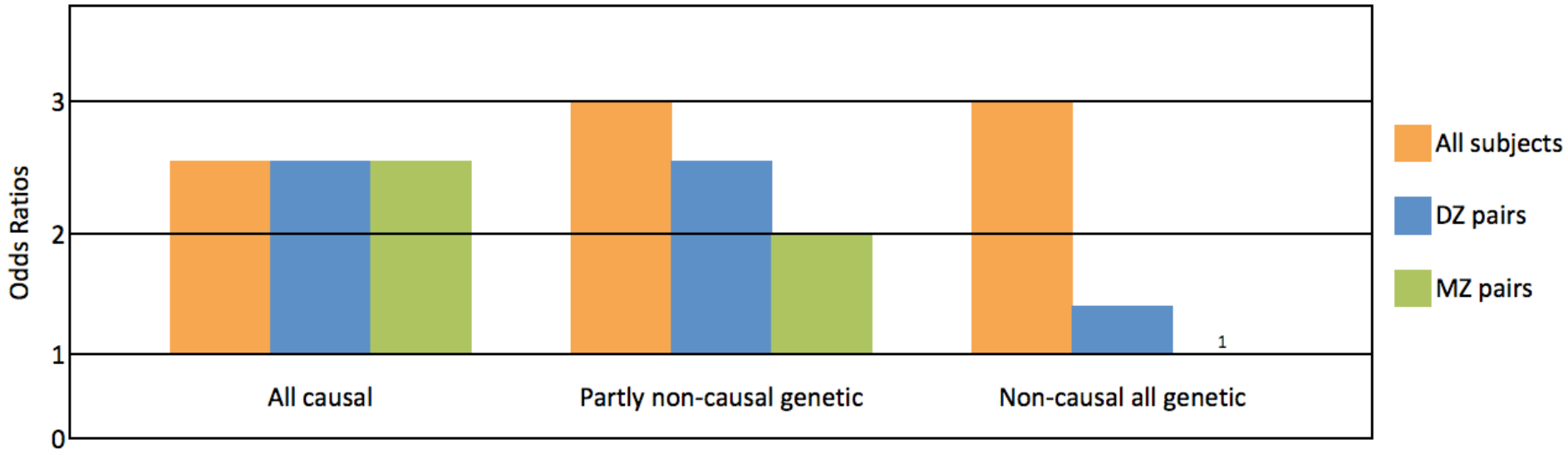
- Nicotine use causes major depression (Kendler, 1993)
- CSA causes adult psychiatric disorders (Kendler, 2000)
- Alcohol abuse-dependence causes lifetime adult major depression (Prescott, 2000)
- Cannabis use causes psychosis (McGrath, 2010)

In co-relative designs, if risk factor-outcome association is:

Causal
- Controlling for background & genetic effects makes no difference
- Estimates (ORs) are same

Partly due to G factors influencing risk & outcome
- Association strongest in entire sample (no control),
- Intermediate for DZs (full E, part G control)
- Lowest for MZ (full G & E control)

Entirely genetic
- MZ ORs approach = 1, DZ ORs are midway



3. Twin-based SEM DOC modeling

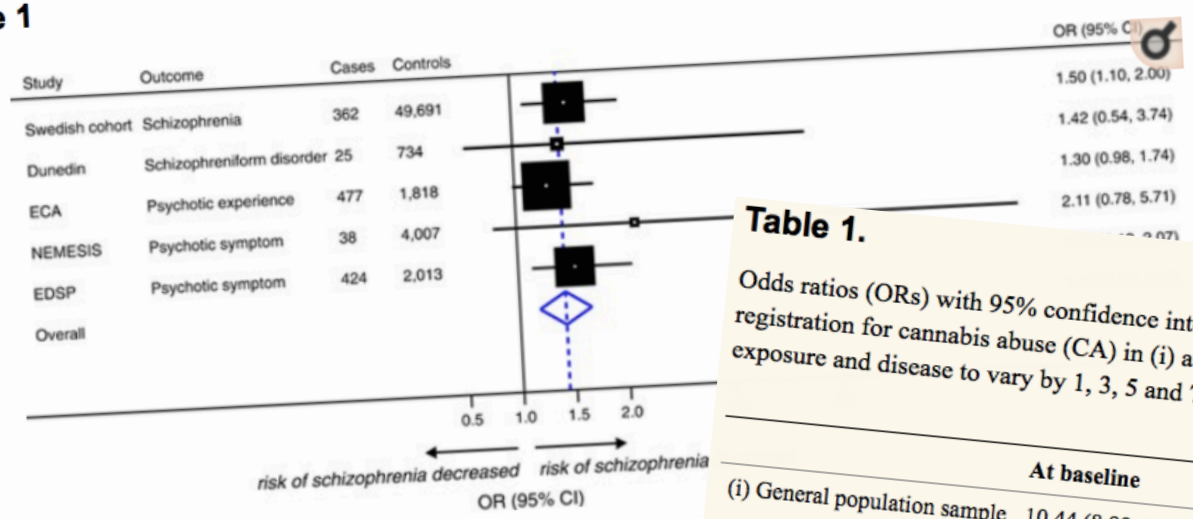
- simple bivariate DOC between measured phenotypes
- causal-common-contingency (CCC) modeling
- multiple indicator DOC modeling between latent phenotypes
- longitudinal DOC modeling & cross-panel designs

4. Instrumental Variable Analyses

- Mendelian Randomization

Warning: All causal inferences invariably rely on methodological assumptions that rarely produce unequivocal results e.g. **cannabis** & **schizophrenia**

Figure 1



Meta-analysis of prospective observational studies reporting an association
related disorders. Meta-analysis uses a random-effects model. Studies are s
schizophrenia and related outcomes). Odds ratios (ORs) and 95% confiden
psychotic symptoms for ever use of cannabis (compared with never use). I
Table S1. Dunedin, Dunedin Multidisciplinary Health & Development St
Developmental Stages of Psychopathology Study; NEMESIS, Netherlan
Cohort.

Table 1.

Odds ratios (ORs) with 95% confidence intervals (CIs) describing the risk of hospital diagnosis of schizophren
registration for cannabis abuse (CA) in (i) a general population sample and (ii) a co-relative sample, allowing th
exposure and disease to vary by 1, 3, 5 and 7 years

	At baseline	After 1year	After 3 years	After 5 years	After 7 years
(i) General population sample	10.44 (8.99–12.11) (n = 5456)	9.19 (7.89–10.70) (n = 5388)	7.69 (6.57–9.00) (n = 5306)	5.95 (5.04–7.02) (n = 5210)	4.24 (3.54–5.0 (n = 5119)
(ii) Co-relative sample					
First-cousin pairs	9.40 (8.12–10.87) (n = 2079)	8.37 (7.22–9.70) (n = 1836)	7.19 (6.19–8.35) (n = 1605)	5.85 (5.03–6.81) (n = 1343)	4.05 (3.46–4.73) (n = 989)
Paternal half-sibling pairs	9.15 (6.39–13.11) (n = 342)	8.42 (5.87–12.09) (n = 328)	7.82 (5.44–11.23) (n = 298)	6.12 (4.24–8.84) (n = 239)	4.18 (2.86–6.11) (n = 173)
Maternal half-siblings pairs	6.00 (4.20–8.58) (n = 245)	5.85 (4.04–8.46) (n = 226)	4.70 (3.23–6.84) (n = 188)	3.42 (2.32–5.05) (n = 146)	2.58 (1.72–3.85) (n = 118)
Full-sibling pairs	5.07 (4.17–6.16) (n = 728)	4.47 (3.66–5.47) (n = 635)	3.71 (3.02–4.55) (n = 546)	2.80 (2.27–3.46) (n = 441)	1.98 (1.59–2.48) (n = 346)
MZ twins (extrapolated)	3.92	3.38	3.31	2.63	1.67

Lancet Psychiatry. 2019 May;6(5):427-436. doi: 10.1016/S2215-0366(19)30048-3. Epub 2019 Mar 19.

The contribution of cannabis use to variation in the incidence of psychotic disorder across Europe (EU-GEI): a multicentre case-control study.

Di Forti M[1], Quattrone D[2], Freeman TP[3], Tripoli G[4], Gayer-Anderson C[5], Quigley H[4], Rodriguez V[4], Jongsma HE[6], Ferraro L[7], La Cascia C[7], La Barbera D[7], Tarricone I[8], Berardi D[8], Szöke A[9], Arango C[10], Tortelli A[11], Velthorst E[12], Bernardo M[13], Del-Ben CM[14], Menezes PR[15], Selten JP[16], Jones PB[17], Kirkbride JB[18], Rutten BP[19], de Haan L[12], Sham PC[20], van Os J[21], Lewis CM[22], Lynskey M[23], Morgan C[5], Murray RM[24]; EU-GEI WP2 Group.

nature neuroscience

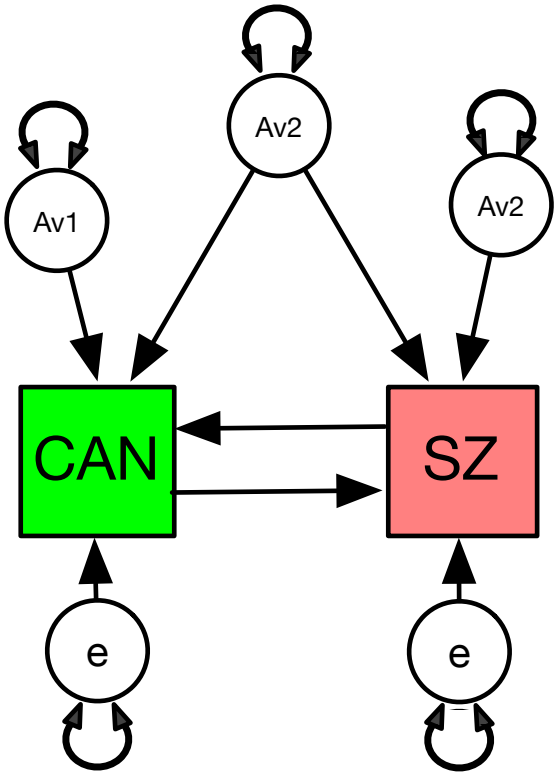
ARTICLES

https://doi.org/10.1038/s41593-018-0206-1

GWAS of lifetime cannabis use reveals new risk loci, genetic overlap with psychiatric traits, and a causal influence of schizophrenia

has been associated with substance use and risk-taking. Significant genetic correlations were found with 14 of 25 tested substance use and mental health-related traits, including smoking, alcohol use, schizophrenia and risk-taking. Mendelian randomization analysis showed evidence for a causal positive influence of schizophrenia risk on cannabis use. Overall, our study provides new insights into the etiology of cannabis use and its relation with mental health.

Based on our genome-wide association study (GWAS) meta-analysis of lifetime cannabis use we estimated a genome-wide genetic correlation of rg=0.25 (SE=0.03, p=5.81E-15) with schizophrenia risk, indicating that genetic risk factors for cannabis use and schizophrenia are positively correlated 5. This could be explained by pleiotropic, causal or reverse causal mechanisms. Mendelian randomization (MR) is an approach that uses



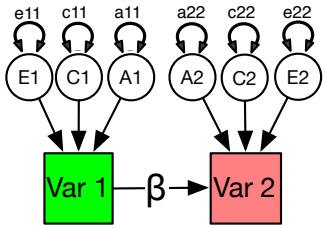
Solution? Triangulation

The confidence in a causal assumption should be based not on a single but multiple divergent methods each differing in their theoretical assumptions. This '**triangulation'** approach (Munafo, 2018) ought to provide stronger converging lines of evidence compared to any well-replicated single approach that ignores unacknowledged biases.

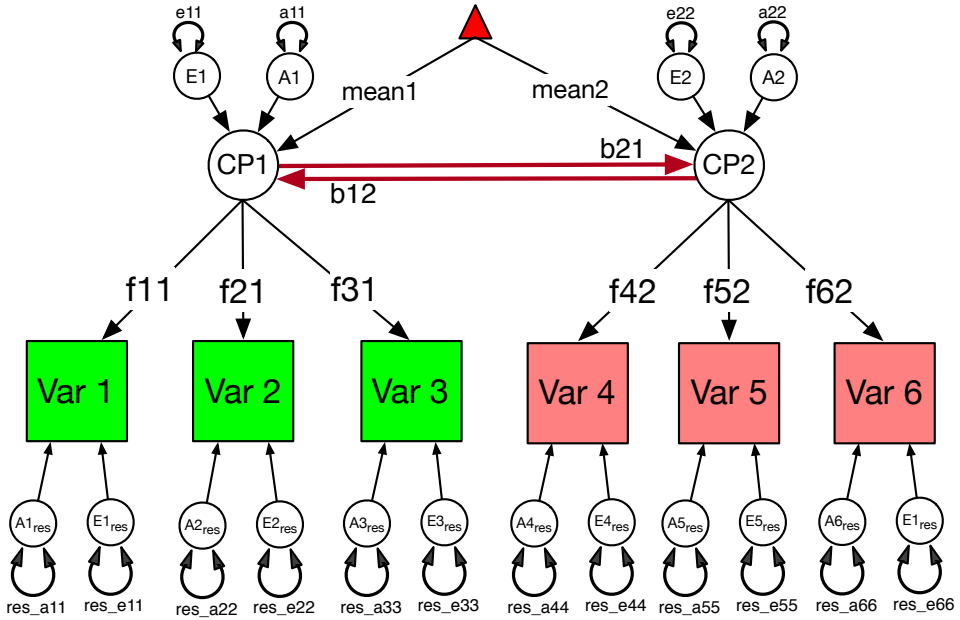
Twin-based SEM DOC modeling

Best natural experimental means of modeling DOC while controlling sources of genetic & environmental confounding

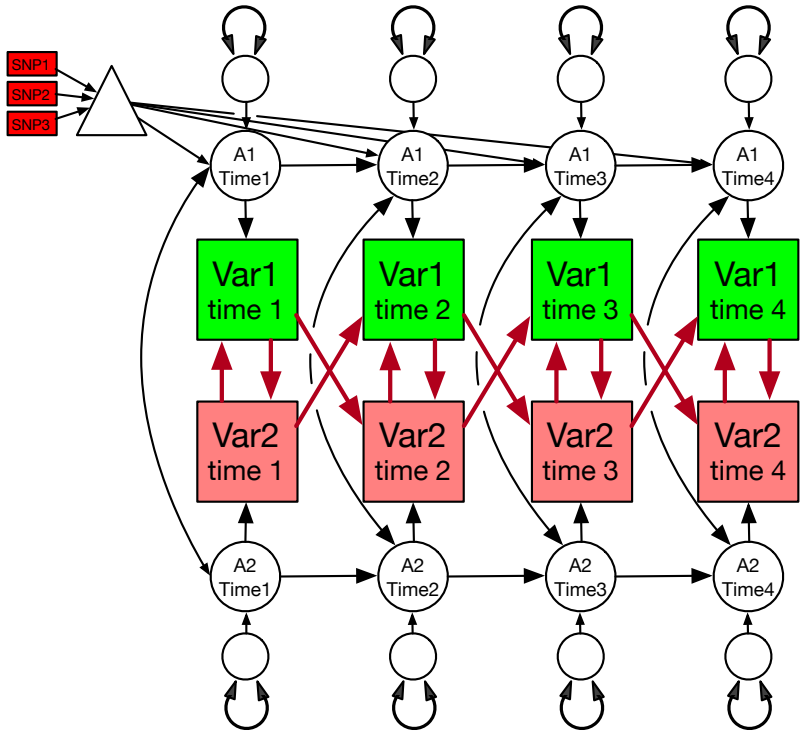
1. Easy and straightforward.



2. Innovative

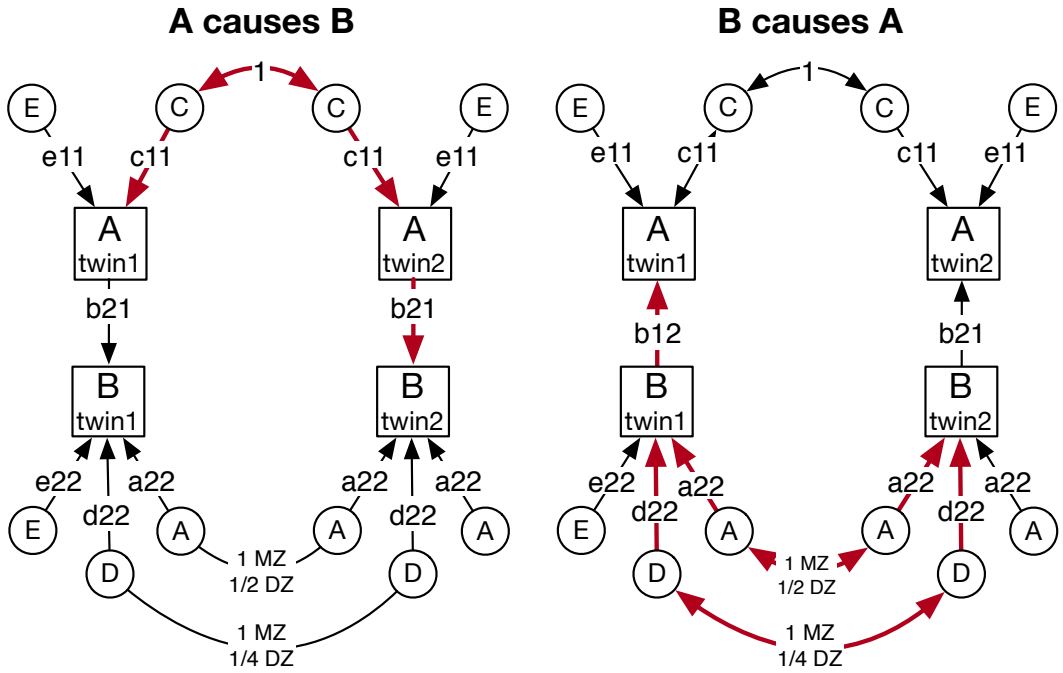


3. Modular & easily extended as data evolves.



Twin-based SEM DOC modeling

The ‘magic’ or power of twin-based SEM DOC modeling is based on the information from the cross-twin cross-trait covariance



Expected cross-twin cross-trait covariance

$rMZ = c_{11}^2 \times b_{21}$

$rDZ = c_{11}^2 \times b_{21}$

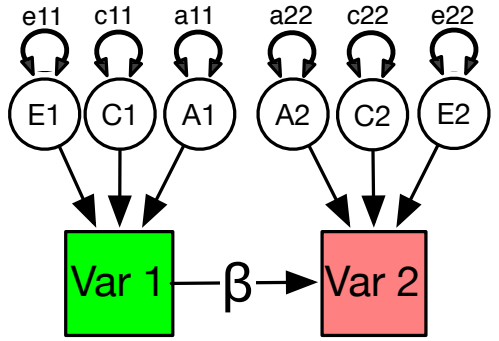
$rMZ = (a_{22}^2 + d_{22}^2) \times b_{12}$

$rDZ = (\frac{1}{2}a_{22}^2 + \frac{1}{4}d_{22}^2) \times b_{12}$

1. Members of a twin pair are not having any mutual effect on one another, i.e., sibling cooperation/rivalry, either within or across variables
2. Relationship between the two target variables is equivalent for twin 1 and twin 2
3. **Twin pair correlations are different between target variables**
4. There are no unmeasured variables influencing both measures and thereby inflate the correlations arising through the causal influence of one variable on the other

Twin-based SEM DOC modeling

Modeling causality at phenotypic level




```
a           <- mxMatrix( type="Symm", nrow=2, ncol=2, free=c(T,F,F,T), values=c(0.5,0,0,0.4),labels=c("a11","a21","a21","a22"), name="a")
c           <- mxMatrix( type="Symm", nrow=2, ncol=2, free=c(T,F,F,T), values=c(0.2,0,0,0.1),labels=c("c11","c21","c21","c22"), name="c")
e           <- mxMatrix( type="Symm", nrow=2, ncol=2, free=c(T,F,F,T), values=c(0.6,0,0,0.7),labels=c("e11","e21","e21","e22"), name="e")


I           <- mxMatrix( type="Iden", nrow=2, ncol=2, name="I")
B           <- mxMatrix( type="Full", nrow=2, ncol=2, byrow=T, free=c(F,F,T,F), values=c(0,0,0.6,0),
                         labels=c("b11"   ,"b2to1",
                                  "b1to2","b22"), name="B")

A           <- mxAlgebra( (solve(I-beta) %&% a, name = "A")
C           <- mxAlgebra( (solve(I-beta) %&% c, name = "C")
E           <- mxAlgebra( (solve(I-beta) %&% e, name = "E")

covMZ   <- mxAlgebra( expression= rbind(cbind(A+C+E , A+C  ),
                                        cbind(A+C   , A+C+E)),          name="expCovMZ" )
covDZ   <- mxAlgebra( expression= rbind(cbind(A+C+E     , 0.5%x%A+C),
                                        cbind( 0.5%x%A+C, A+C+E    )),  name="expCovDZ" )
```

```
IdenMatrix 'I'

$labels: No labels assigned.

$values
     [,1] [,2]
[1,]    1    0
[2,]    0    1
```

```
FullMatrix 'B'

$labels
     [,1]    [,2]
[1,] "b11"   "b2to1"
[2,] "b1to2" "b22"

$values
     [,1] [,2]
[1,]  0.0    0
[2,]  0.6    0
```

```
SymmMatrix 'a'

$labels
     [,1]  [,2]
[1,] "a11" "a21"
[2,] "a21" "a22"

$values
     [,1] [,2]
[1,]  0.5  0.0
[2,]  0.0  0.4
```

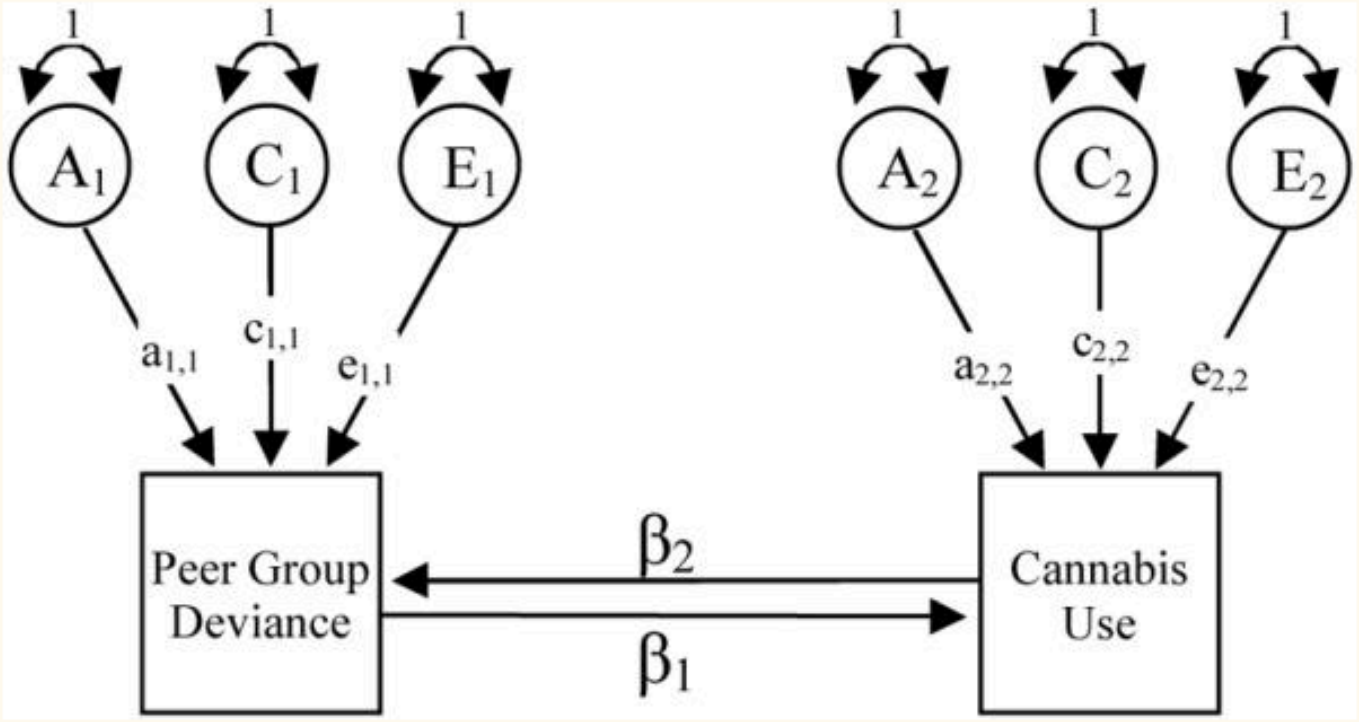


Figure 2

Modeling direction of causation between peer group deviance (PGD) and cannabis use (CU). This approach predicts the relationship between PGD and CU is explained by a reciprocal interaction at the phenotypic level. In the uni-directional PGD-to-CU and CU-to-PGD models, the β_2 and β_1 pathways are set to zero.

Note: A, C & E = latent additive genetic, shared and non-shared environmental effects for PGD and CU

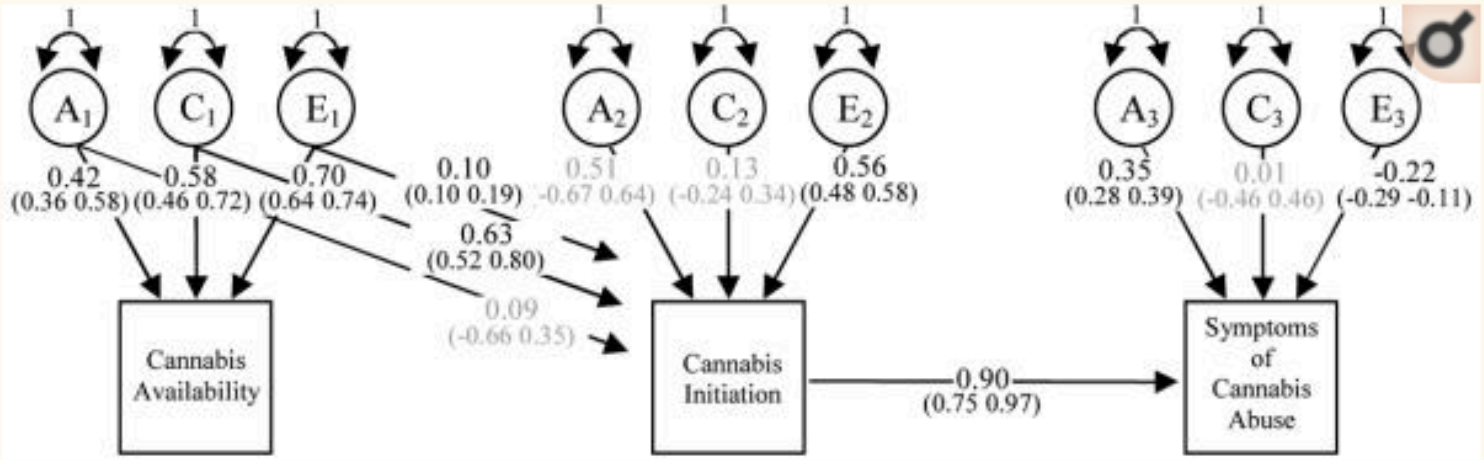


Figure 4

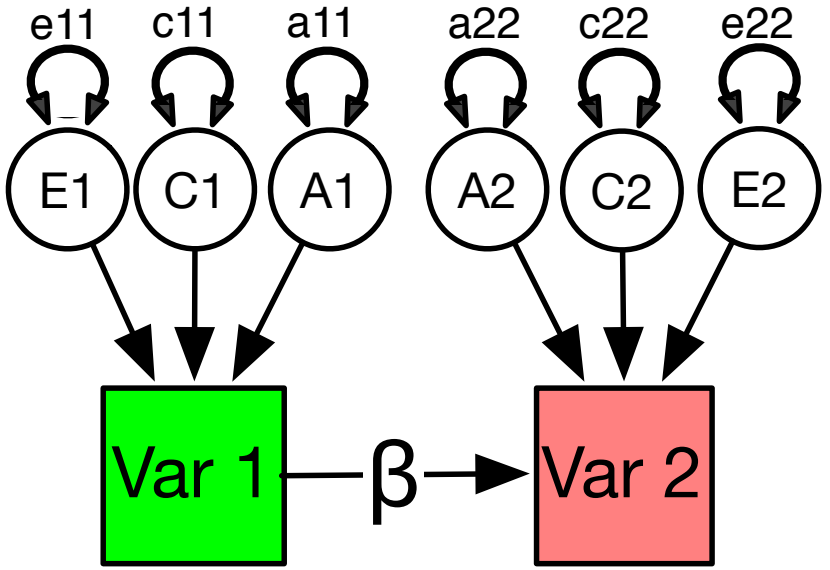
Best fitting model of cannabis availability, cannabis initiation, and the number of DSM-IV abuse symptoms. Model includes standardized path coefficients with 95% confidence intervals. Parameters with 95% confidence intervals spanning zero are shown in light grey. All others are in black.

Note: A=Addition genetic variance, C=Common or shared environmental variance, E=Non-shared environmental variance

DOC modeling

Problems with modeling DOC at the phenotypic level

Omission measurement error known to produce biased estimates of the causal parameters (Neale & Cardon, 1992)



Genet Epidemiol. 1994;11(6):503-22.

Depression and parental bonding: cause, consequence, or genetic covariance?

Neale MC[1], Walters E, Health AC, Kessler RC, Pérusse D, Eaves LJ, Kendler KS.

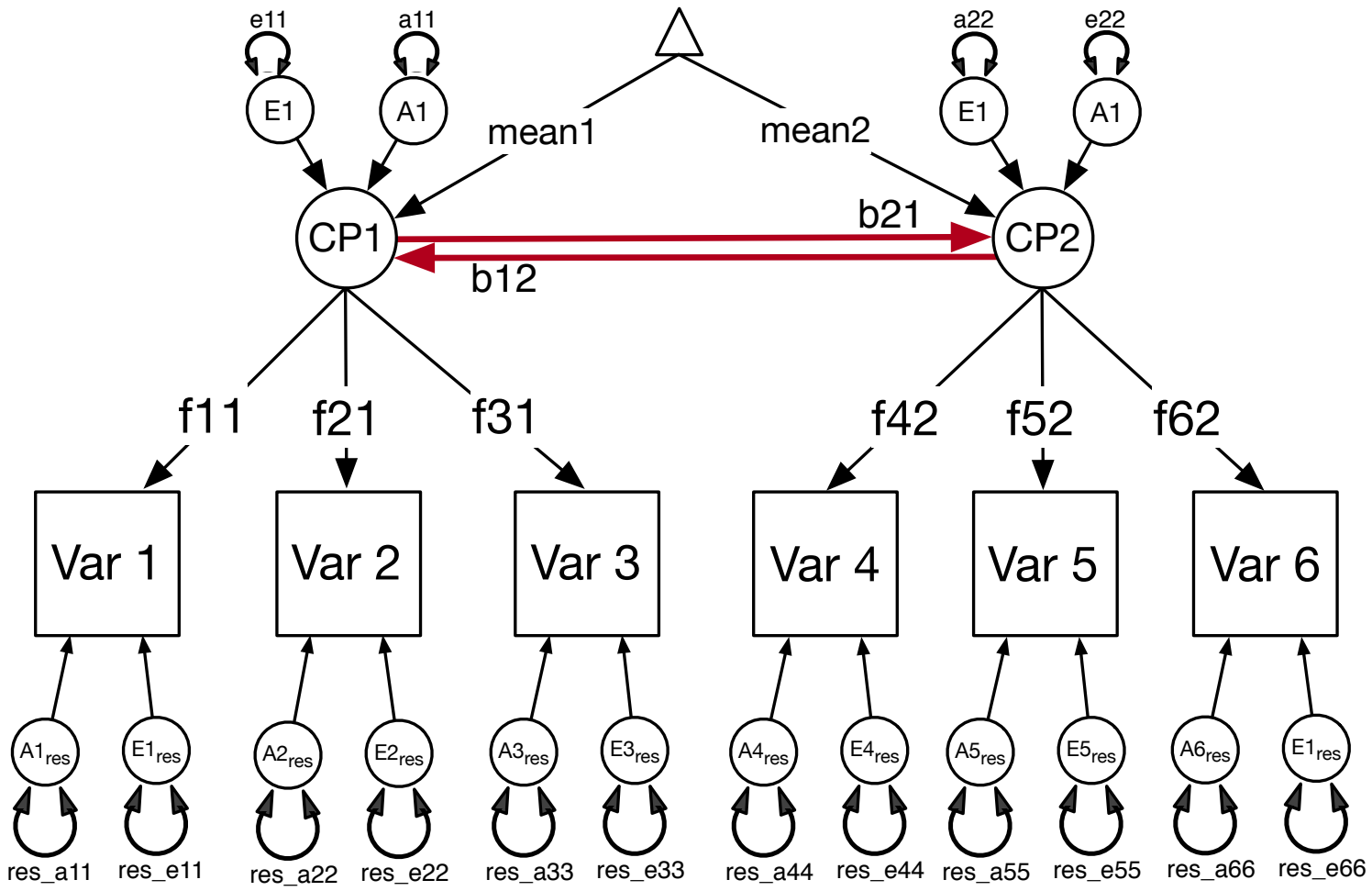
Behav Genet. 1993 Jan;23(1):29-50.

Testing hypotheses about direction of causation using cross-sectional family data.

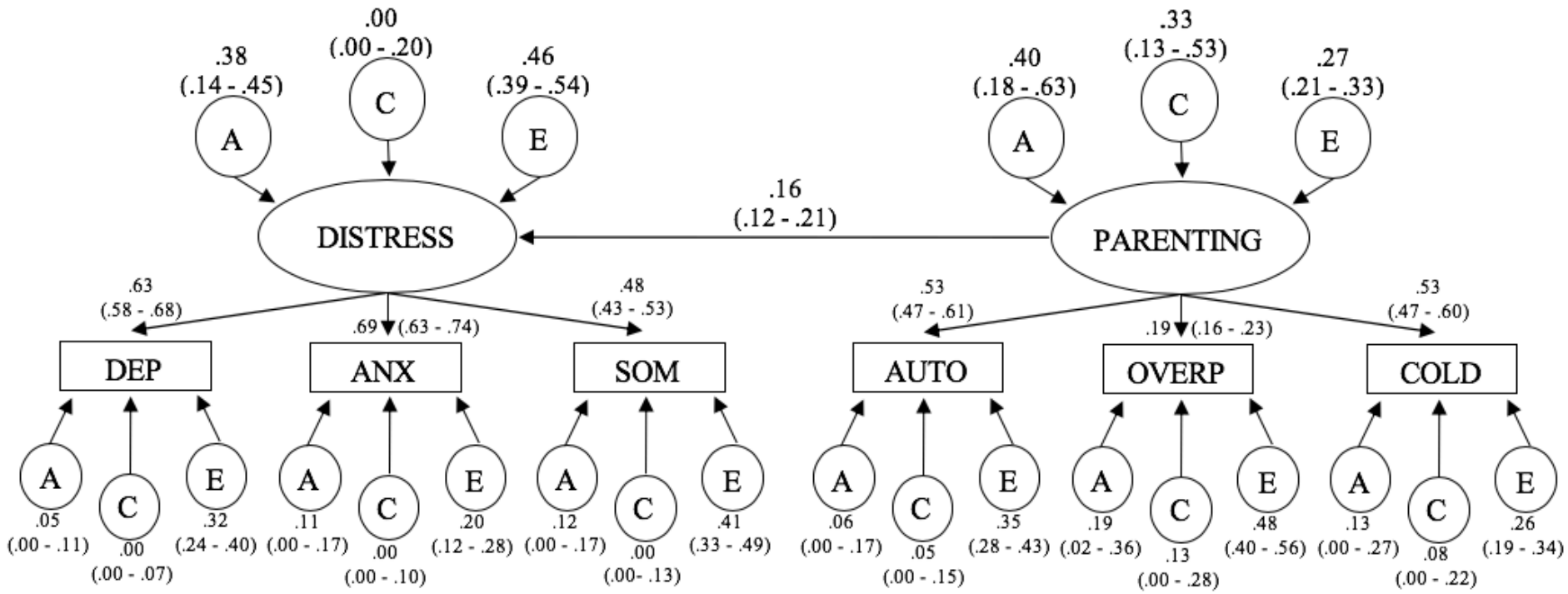
Heath AC[1], Kessler RC, Neale MC, Hewitt JK, Eaves LJ, Kendler KS.

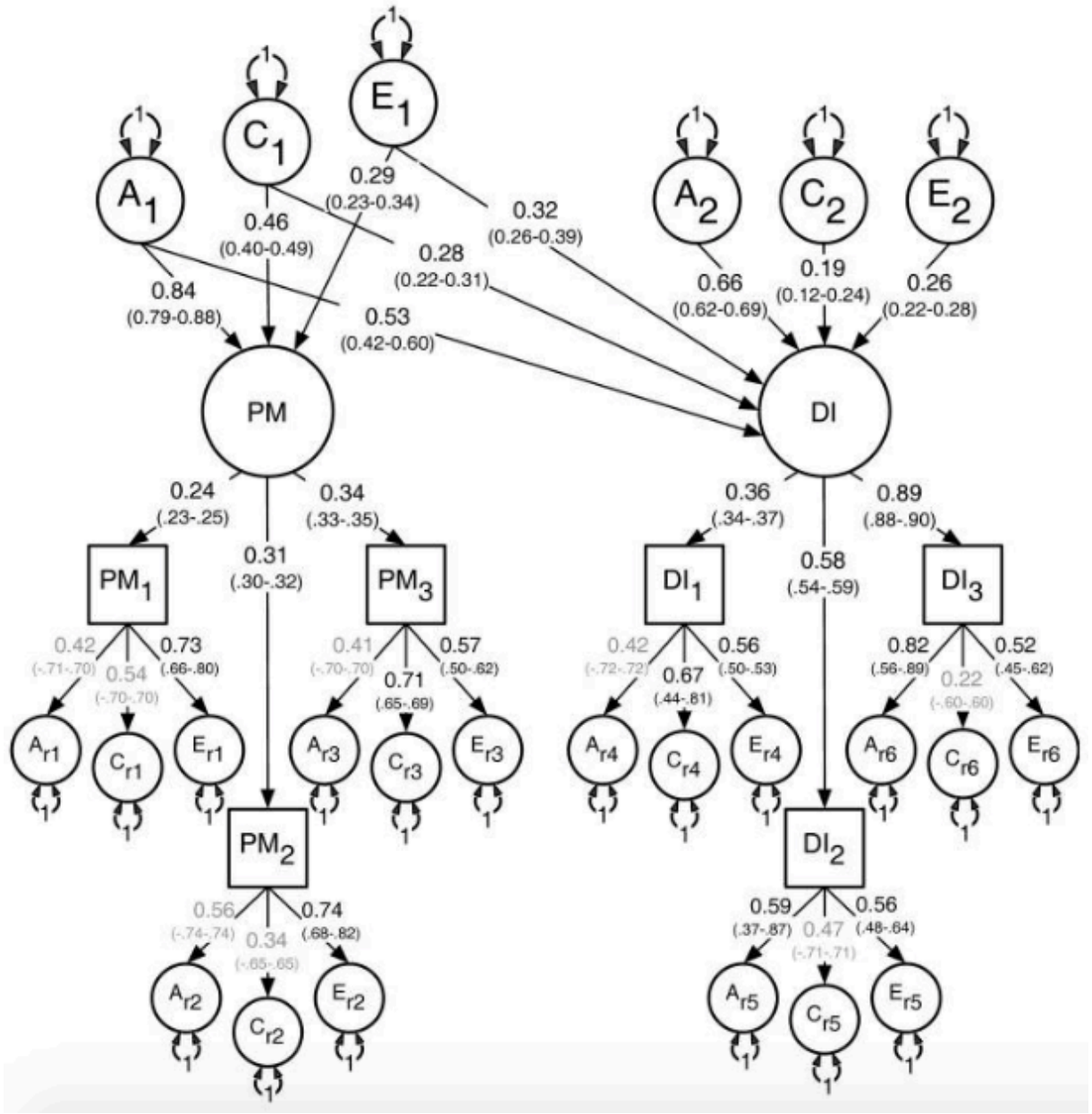
DOC modeling

Modeling DOC at the latent factor level

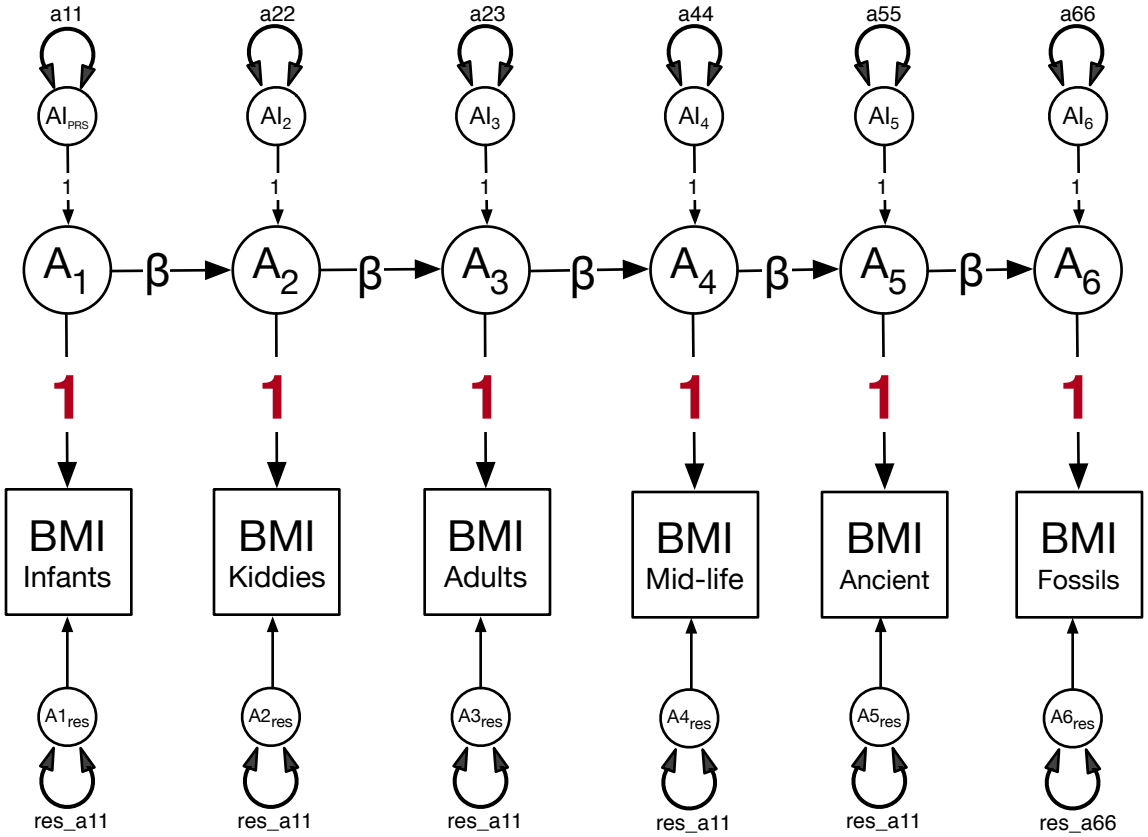


Best-fitting direction of causation model for the psychological distress and PBI parenting dimensions with standardized variance components and 95% confidence intervals.

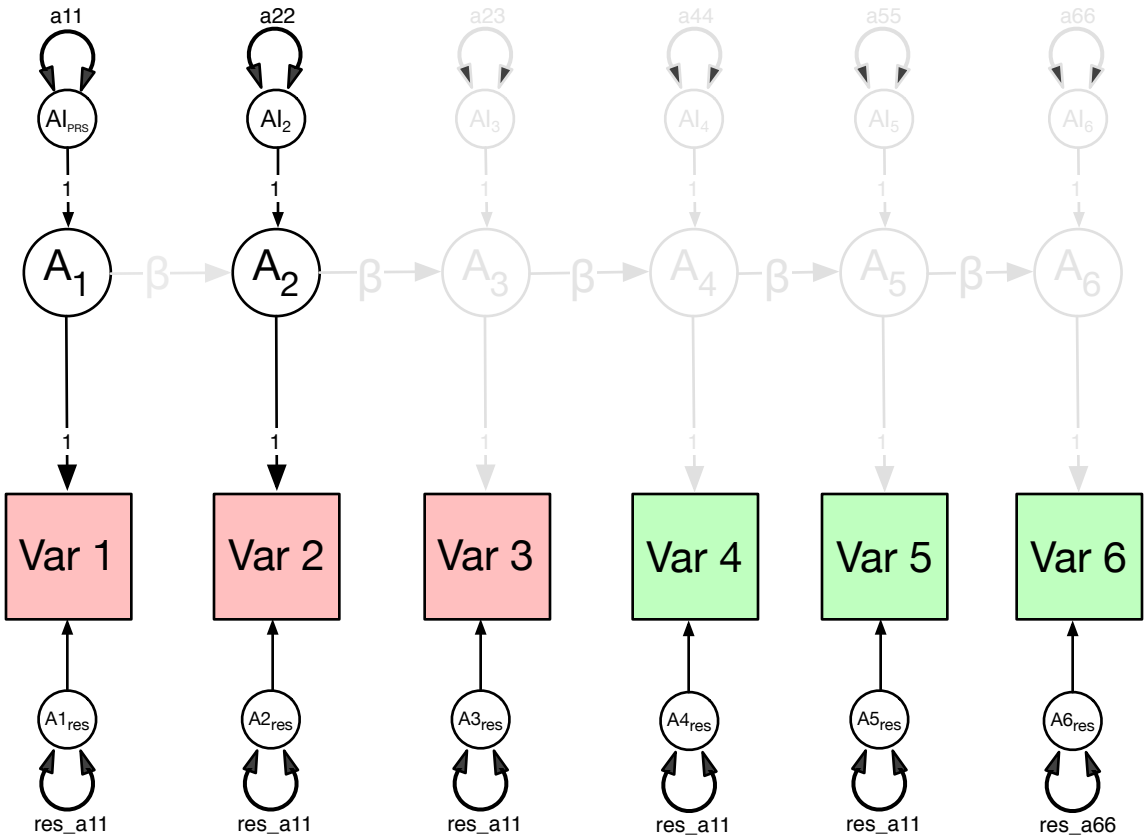



A1
C1
E1
A2
C2
E2
PM
DI
PM1
PM2
PM3
DI1
DI2
DI3
Ar1
Cr1
Er1
Ar2
Cr2
Er2
Ar3
Cr3
Er3
Ar4
Cr4
Er4
Ar5
Cr5
Er5
Ar6
Cr6
Er6
0.46 (0.40-0.49)
0.29 (0.23-0.34)
0.32 (0.26-0.39)
0.28 (0.22-0.31)
0.84 (0.79-0.88)
0.53 (0.42-0.60)
0.66 (0.62-0.69)
0.19 (0.12-0.24)
0.26 (0.22-0.28)
0.24 (.23-.25)
0.34 (.33-.35)
0.31 (.30-.32)
0.36 (.34-.37)
0.89 (.88-.90)
0.58 (.54-.59)
0.42 (-.71-.70)
0.54 (-.70-.70)
0.73 (.66-.80)
0.41 (-.70-.70)
0.71 (.65-.69)
0.57 (.50-.62)
0.42 (-.72-.72)
0.67 (.44-.81)
0.56 (.50-.53)
0.82 (.56-.89)
0.22 (-.60-.60)
0.52 (.45-.62)
0.56 (-.74-.74)
0.34 (-.65-.65)
0.74 (.68-.82)
0.59 (.37-.87)
0.47 (-.71-.71)
0.56 (.48-.64)
1

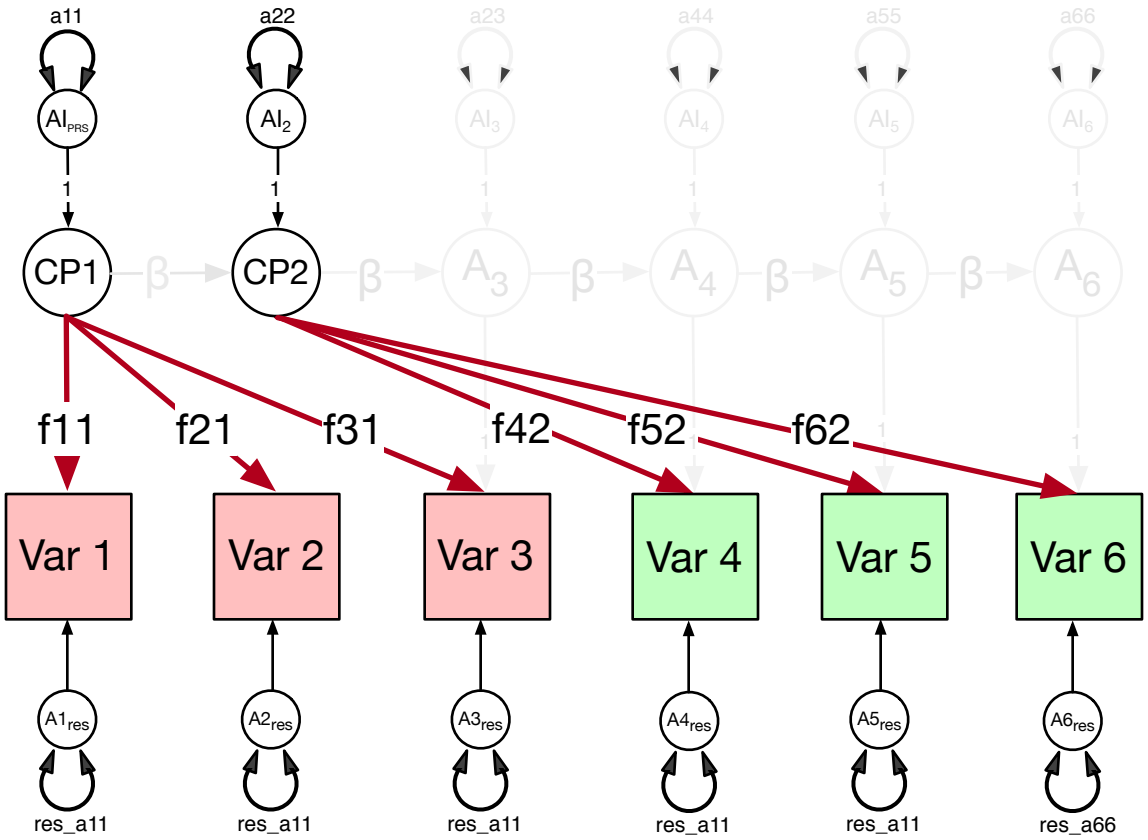
Specifying the model




```
lamba      <- mxMatrix(type="Iden", nrow=nFac, ncol=nFac, name = "lamba")

IdenMatrix 'lamba'
$values
     [,1] [,2] [,3] [,4] [,5] [,6]
[1,]    1    0    0    0    0    0
[2,]    0    1    0    0    0    0
[3,]    0    0    1    0    0    0
[4,]    0    0    0    1    0    0
[5,]    0    0    0    0    1    0
[6,]    0    0    0    0    0    1
$free: No free parameters
```

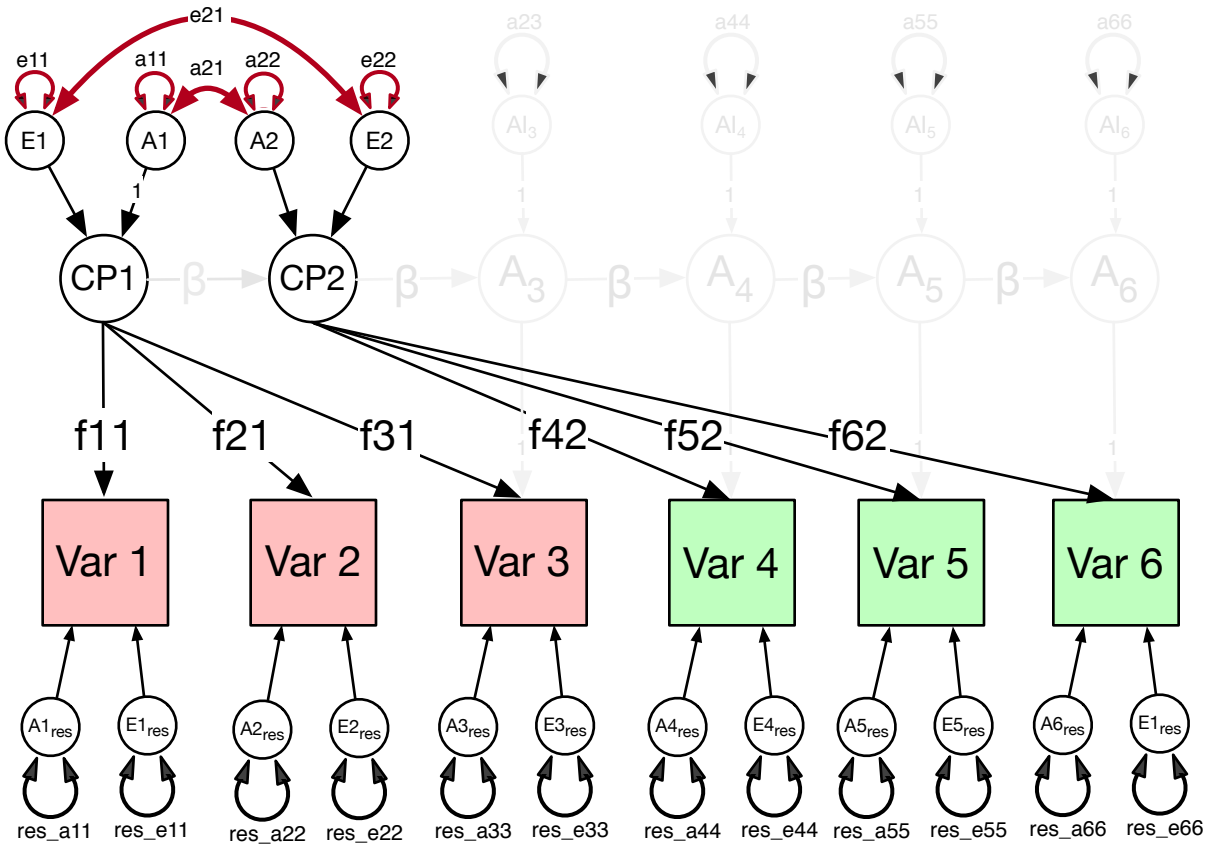


Specify the number of common pathways



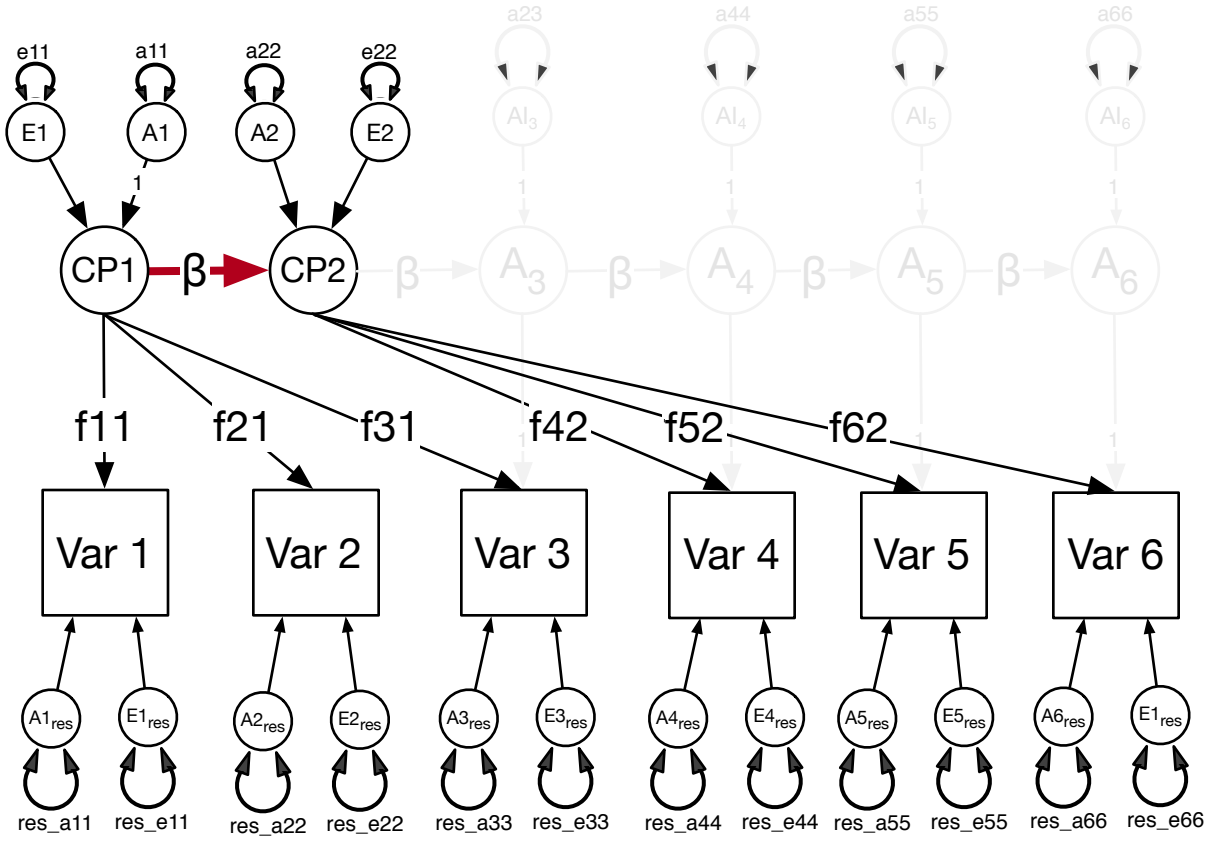
Specify the **factor loading matrix** i.e. factor loadings from the latent true scores (circles) to the observed phenotypes (squares)

```
lamba      <- mxMatrix(type="Full", nrow=6, ncol=2, free = Ffree, values = Fvals, labels=Flabs, name = "lamba")

FullMatrix 'lamba'

$free                $values              $labels
      [,1]  [,2]           [,1] [,2]           [,1]  [,2]
[1,]  TRUE FALSE    [1,] 0.85 0.00     [1,] "f11" NA
[2,]  TRUE FALSE    [2,] 0.85 0.00     [2,] "f21" NA
[3,]  TRUE FALSE    [3,] 0.85 0.00     [3,] "f31" NA
[4,] FALSE  TRUE    [4,] 0.00 0.85     [4,] NA    "f42"
[5,] FALSE  TRUE    [5,] 0.00 0.85     [5,] NA    "f52"
[6,] FALSE  TRUE    [6,] 0.00 0.85     [6,] NA    "f62"
```



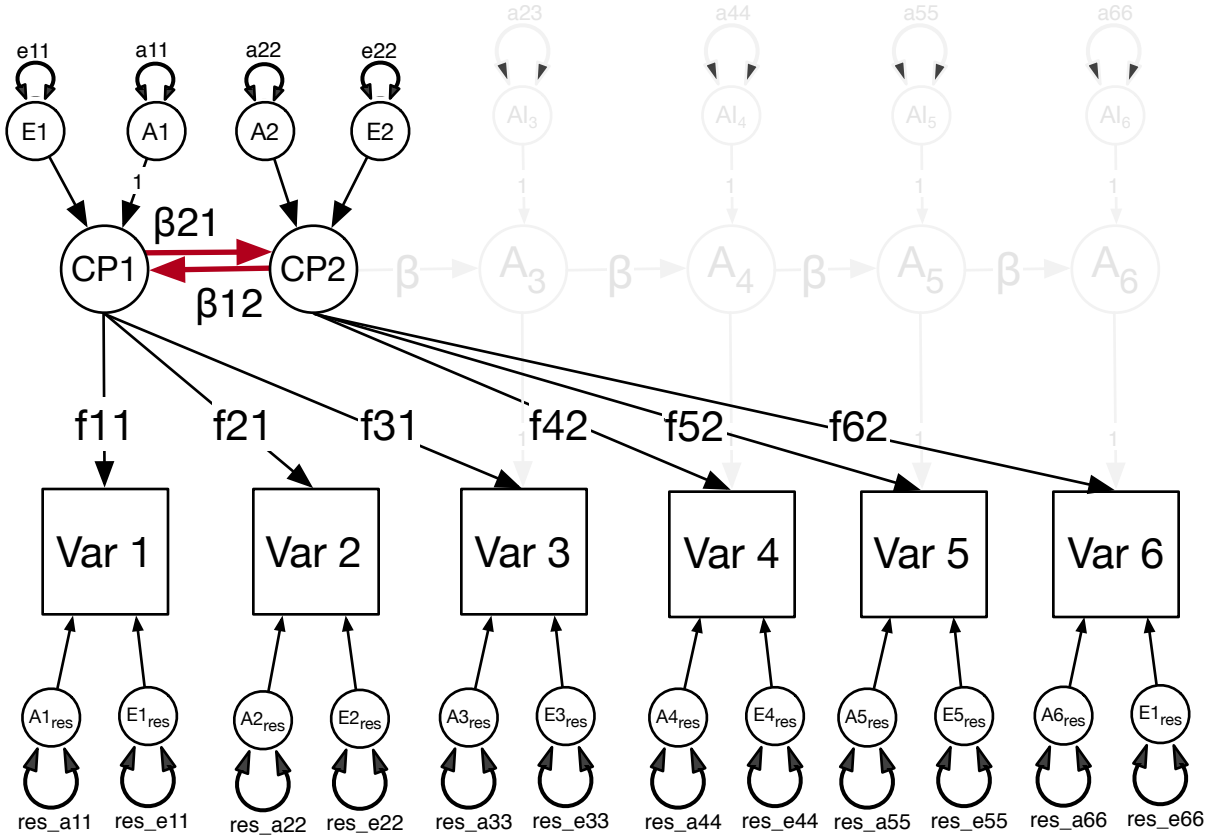
Specificy the **variance-covariance matrices for the common pathways** i.e. A, C & E variance-covariance 'psi' matrices for CP1 & CP2

```
aFree      <- c(T,T,T,T)
aVals      <- c(0.6,0.3,0.3,0.6)
psi_a      <- mxMatrix(type = "Symm", nrow = nFac,ncol=nFac, labels=c("a11","a21","a21","a22"), free=aFree, values=aVals, name="psi_a")

SymmMatrix 'psi_a'

$labels          $values          $free
     [,1]  [,2]       [,1] [,2]       [,1] [,2]
[1,] "a11" "a21" [1,]  0.6  0.3  [1,] TRUE TRUE
[2,] "a21" "a22" [2,]  0.3  0.6  [2,] TRUE TRUE
```



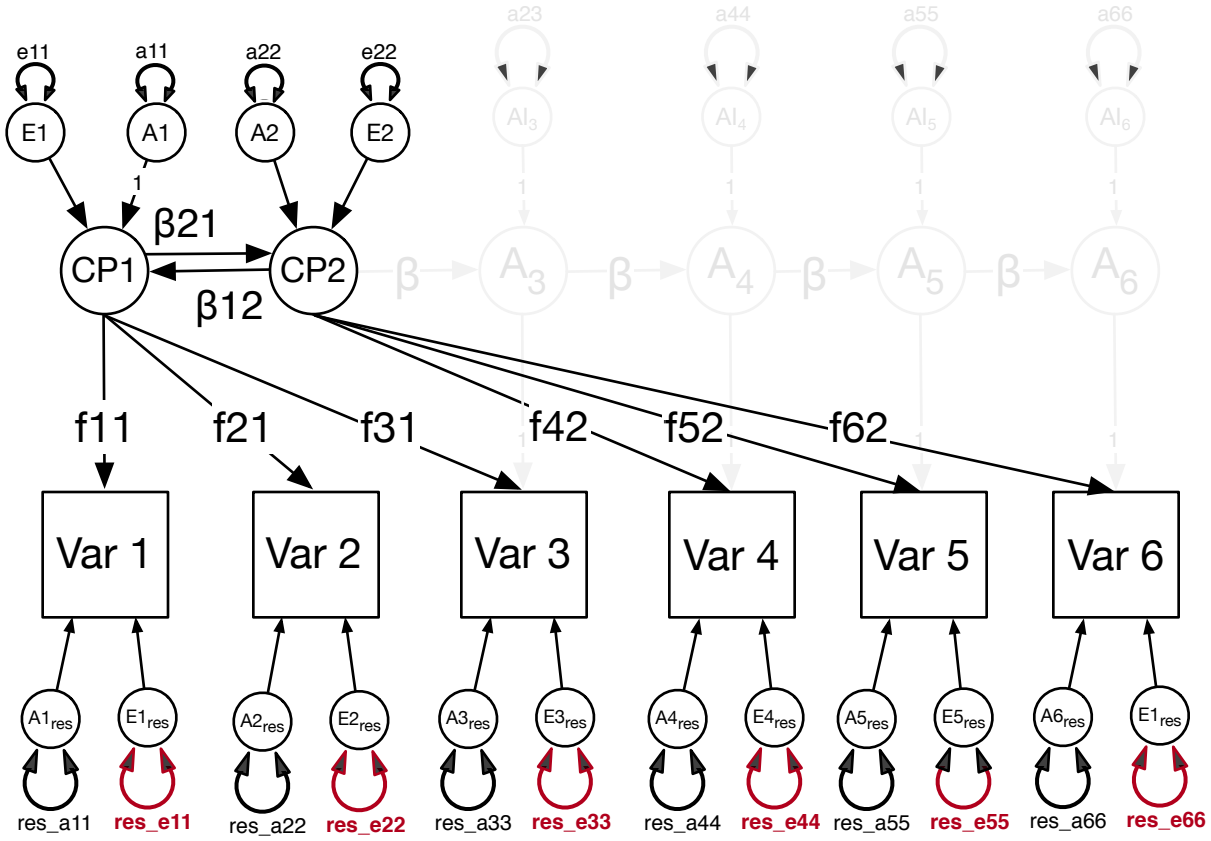
Specify the causal coefficients **between the common pathways** i.e. the beta matrix

```
beta       <- mxMatrix(type="Full", nrow=nFac, ncol=nFac, labels=Blabs, free=Bfree, values=Bvals, name = "beta" )

FullMatrix 'beta'

$labels          $values          $free
     [,1] [,2]        [,1] [,2]        [,1]  [,2]
[1,] NA   NA     [1,]  0.0    0    [1,] FALSE FALSE
[2,] "B"  NA     [2,]  0.6    0    [2,]  TRUE FALSE
```



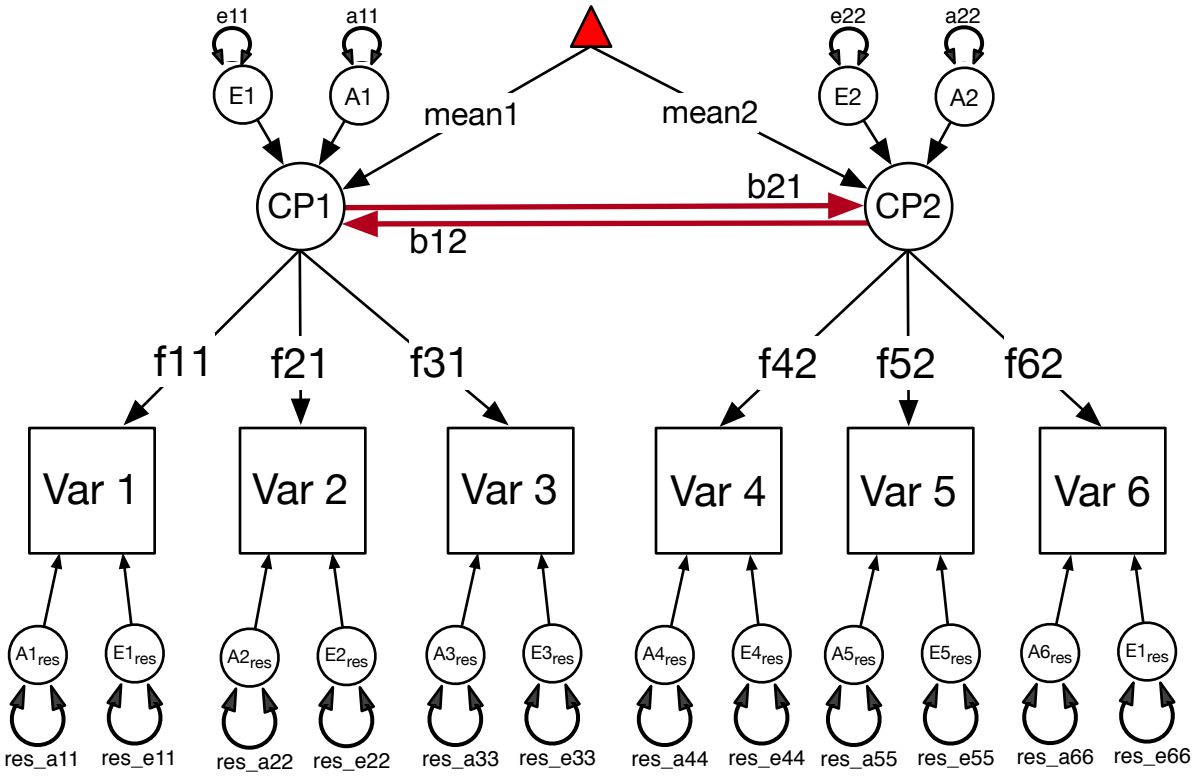
Specify the causal coefficients **between the common pathways** i.e. the beta matrix

```
FullMatrix 'beta'

$labels                $values          $free
     [,1]  [,2]              [,1] [,2]        [,1]  [,2]
[1,] NA    "B12"   [1,]  0.0  0.3  [1,] FALSE  TRUE
[2,] "B21" NA      [2,]  0.6  0.0  [2,]  TRUE FALSE
```

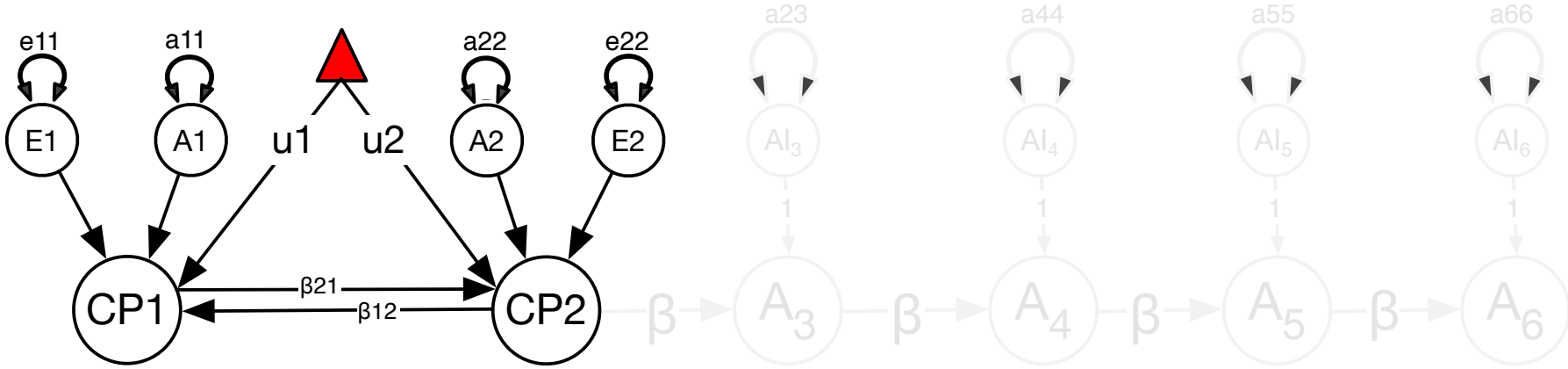


Specify the **residual genetic (A) & environmental (C & E) matrices** i.e. residual variances in the model not explained by the factor structure

```
epsilon_e  <- mxMatrix(type = "Diag", nrow=nv, ncol=nv, labels=res_e_labs, free=T, values=0.1, name="epsilon_e")

DiagMatrix 'epsilon_e'
$labels
     [,1]      [,2]      [,3]      [,4]      [,5]      [,6]
[1,] "res_e1" NA        NA        NA        NA        NA
[2,] NA       "res_e2"  NA        NA        NA        NA
[3,] NA       NA        "res_e3"  NA        NA        NA
[4,] NA       NA        NA        "res_e4"  NA        NA
[5,] NA       NA        NA        NA        "res_e5"  NA
[6,] NA       NA        NA        NA        NA        "res_e6"

$values
     [,1] [,2] [,3] [,4] [,5] [,6]
[1,]  0.1  0.0  0.0  0.0  0.0  0.0
[2,]  0.0  0.1  0.0  0.0  0.0  0.0
[3,]  0.0  0.0  0.1  0.0  0.0  0.0
[4,]  0.0  0.0  0.0  0.1  0.0  0.0
[5,]  0.0  0.0  0.0  0.0  0.1  0.0
[6,]  0.0  0.0  0.0  0.0  0.0  0.1
```

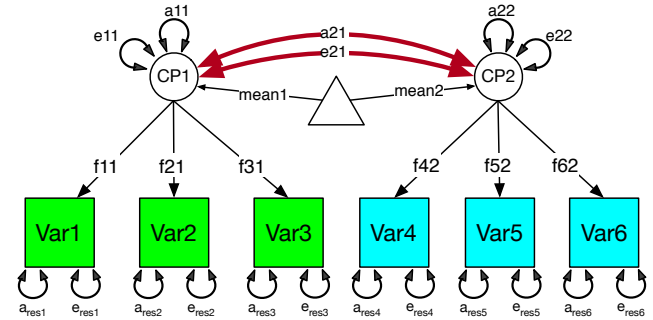


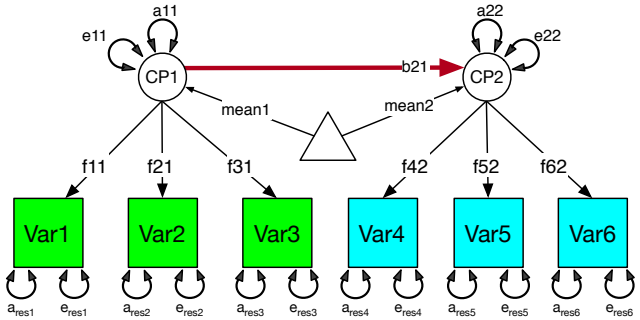
Specify the means of the latent CPs

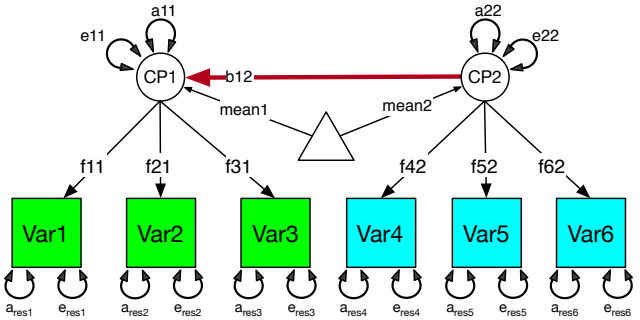


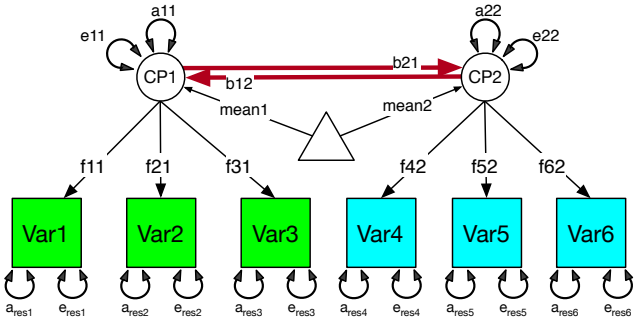

```
Mean      <- mxMatrix( type="Full", nrow=2, ncol=1, free=c(T,T), labels=c( "m1","m2"), values = c(0.01,0.30), name = "Mean" )
FacMean   <- mxAlgebra( solve(I-beta)  %*% Mean, name="FacMean" )
ManMean   <- mxAlgebra( t(lamba %*% FacMean ), name="ManMean" )
expMean   <- mxAlgebra( expression= cbind(ManMean,ManMean), name="expMean" )
```

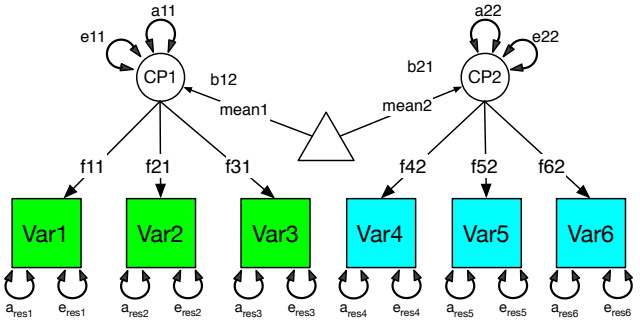
Varieties of causal and non-causal models





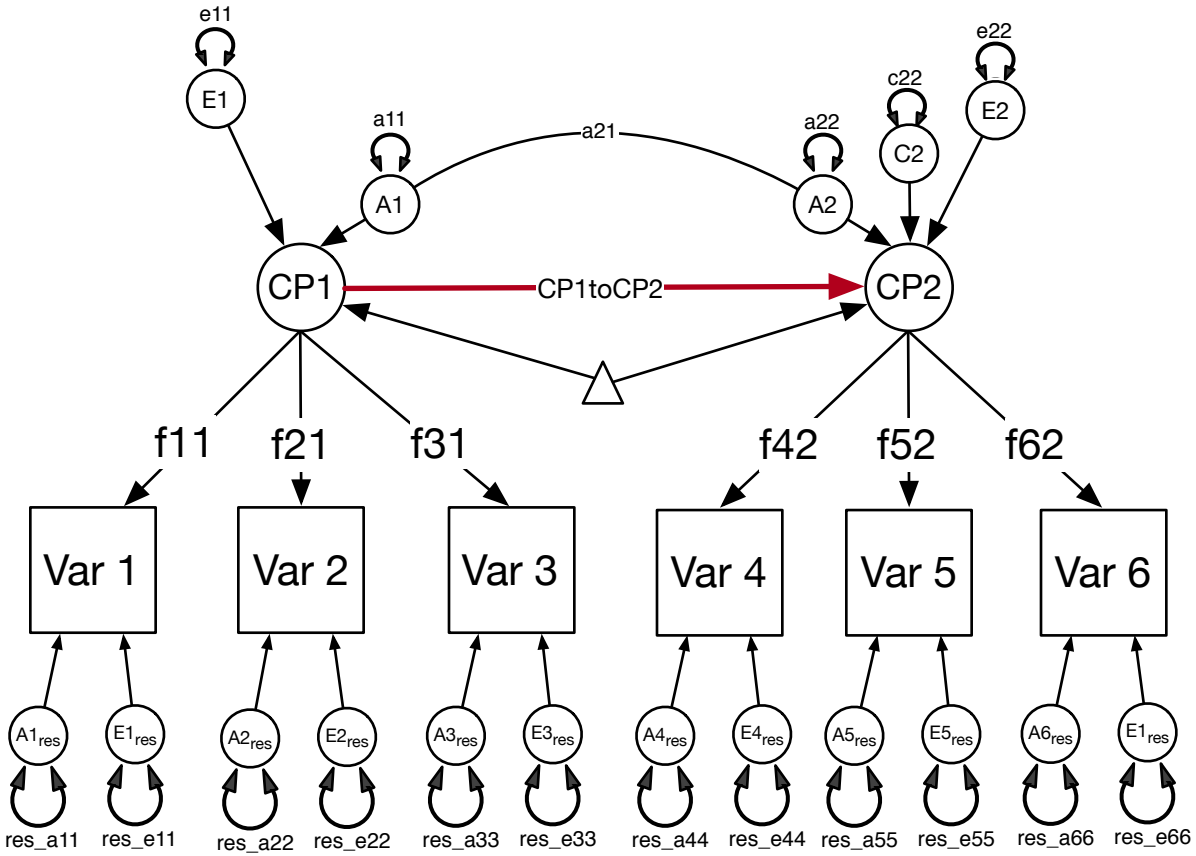







```
mxCheckIdentification(ACE_model, details=TRUE)
```

Hybrid model

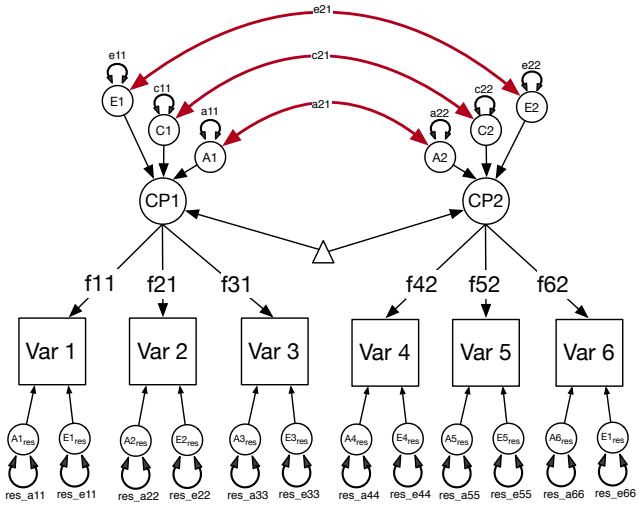


Tutorial

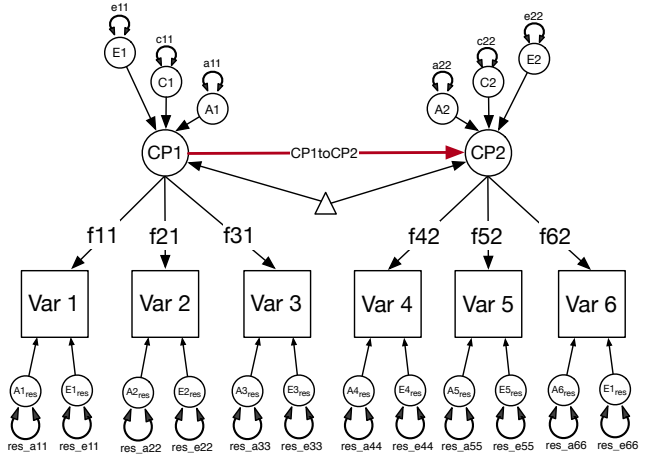
1. Run ACE variance components (non-causal) model = null hypothesis
2. Specify / test AE variance components (non-causal) model.
3. Specify / test CE variance components (non-causal) model
4. Compare ACE, AE & CE variance components (non-causal) models
   Extra credit: Re-run ACE model to estimate Cis on the A,C,E parameters. Which are non-significant?

5. Run the 'CP1 to CP2' causal model (no correlated A, C or E effects)
6. Run the 'CP2 to CP1' causal model (no correlated A, C or E effects)
7. Run the 'CP1 to CP2' & 'CP2 to CP1' causal model (no correlated A, C or E effects) i.e. reciprocal interaction.
8. Compare the 3 causal models to the null hypothesis

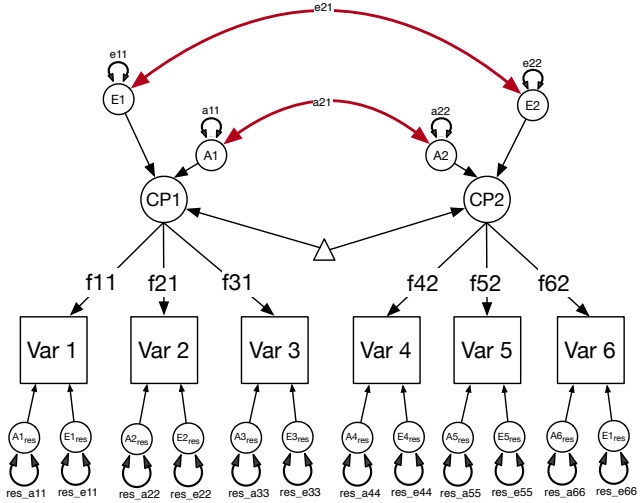
1


e21
e11
c21
e22
E1
c11
C1
a11
a21
a22
c22
E2
A1
A2
CP1
CP2
f11
f21
f31
f42
f52
f62
Var 1
Var 2
Var 3
Var 4
Var 5
Var 6
A1res
E1res
A2res
E2res
A3res
E3res
A4res
E4res
A5res
E5res
A6res
E1res
res_a11
res_e11
res_a22
res_e22
res_a33
res_e33
res_a44
res_e44
res_a55
res_e55
res_a66
res_e66

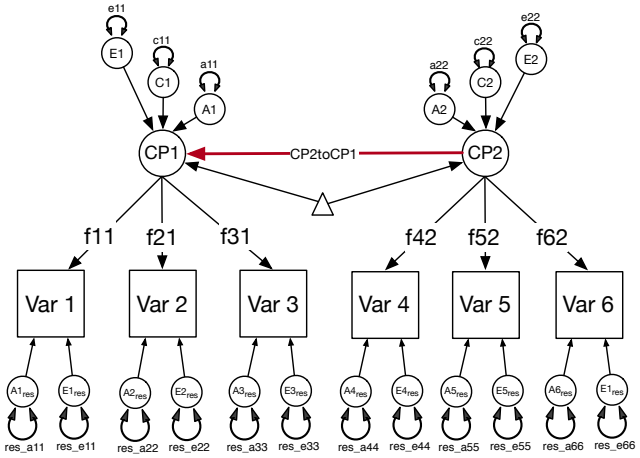
4


e11
E1
c11
C1
a11
A1
a22
A2
c22
C2
e22
E2
CP1
CP1toCP2
CP2
f11
f21
f31
f42
f52
f62
Var 1
Var 2
Var 3
Var 4
Var 5
Var 6
A1res
E1res
A2res
E2res
A3res
E3res
A4res
E4res
A5res
E5res
A6res
E1res
res_a11
res_e11
res_a22
res_e22
res_a33
res_e33
res_a44
res_e44
res_a55
res_e55
res_a66
res_e66

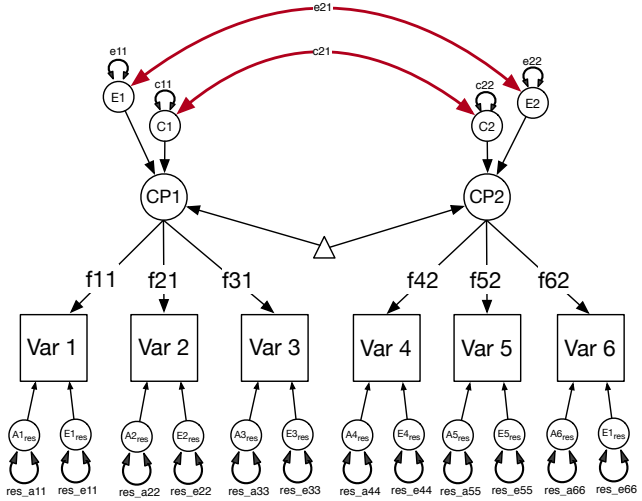
2


e21
e11
e22
E1
E2
a11
a21
a22
A1
A2
CP1
CP2
f11
f21
f31
f42
f52
f62
Var 1
Var 2
Var 3
Var 4
Var 5
Var 6
A1res
E1res
A2res
E2res
A3res
E3res
A4res
E4res
A5res
E5res
A6res
E1res
res_a11
res_e11
res_a22
res_e22
res_a33
res_e33
res_a44
res_e44
res_a55
res_e55
res_a66
res_e66

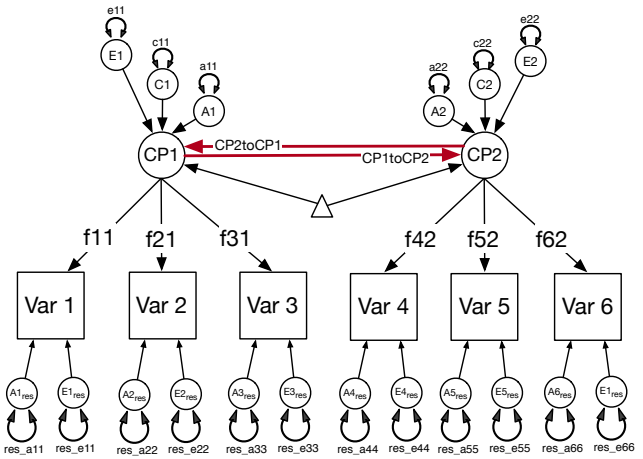
5

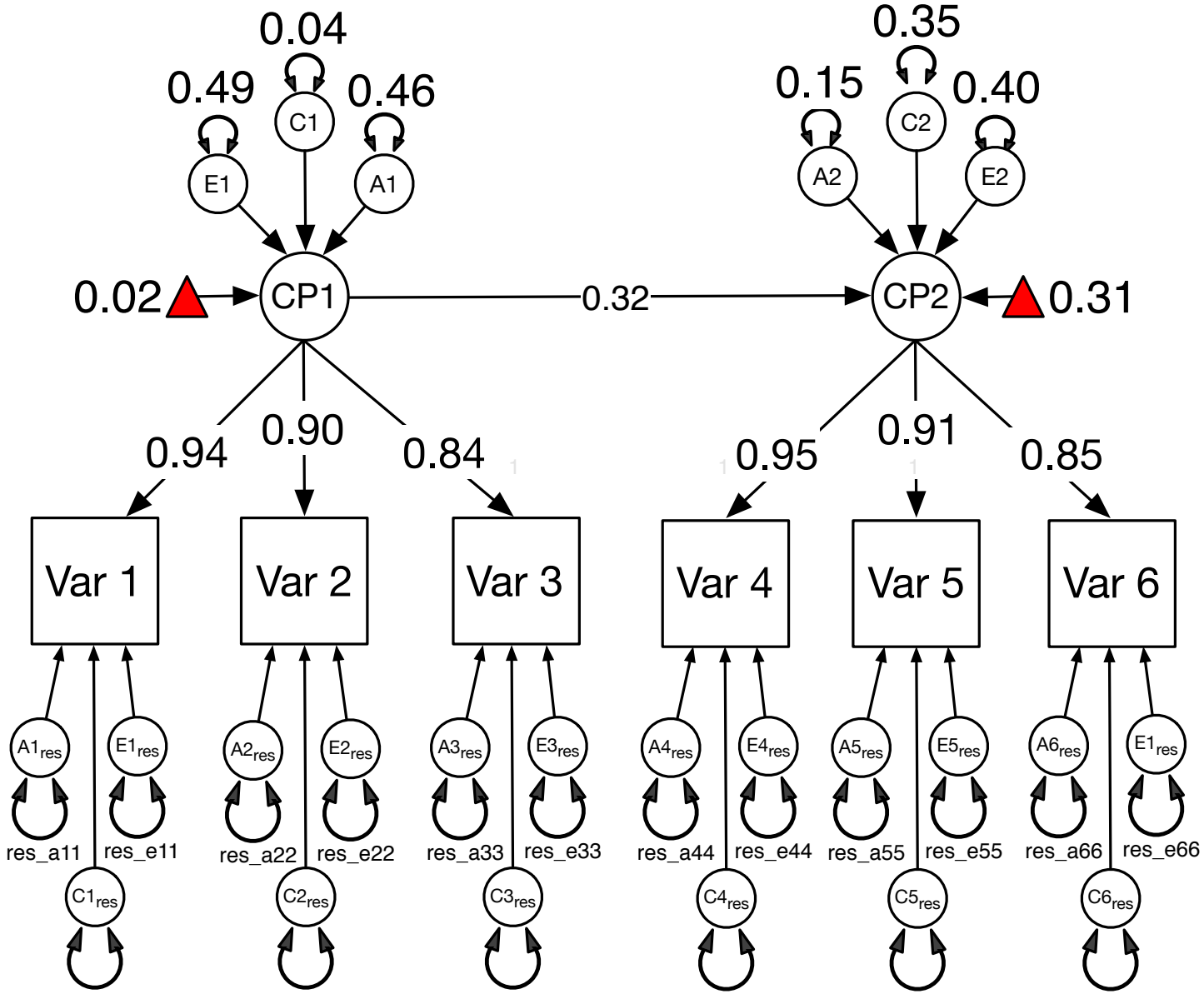

e11
E1
c11
C1
a11
A1
a22
A2
c22
C2
e22
E2
CP1
CP2toCP1
CP2
f11
f21
f31
f42
f52
f62
Var 1
Var 2
Var 3
Var 4
Var 5
Var 6
A1res
E1res
A2res
E2res
A3res
E3res
A4res
E4res
A5res
E5res
A6res
E1res
res_a11
res_e11
res_a22
res_e22
res_a33
res_e33
res_a44
res_e44
res_a55
res_e55
res_a66
res_e66

3


e21
e11
c21
e22
E1
c11
C1
c22
E2
CP1
CP2
f11
f21
f31
f42
f52
f62
Var 1
Var 2
Var 3
Var 4
Var 5
Var 6
A1res
E1res
A2res
E2res
A3res
E3res
A4res
E4res
A5res
E5res
A6res
E1res
res_a11
res_e11
res_a22
res_e22
res_a33
res_e33
res_a44
res_e44
res_a55
res_e55
res_a66
res_e66

6


e11
E1
c11
C1
a11
A1
a22
A2
c22
C2
e22
E2
CP1
CP2toCP1
CP1toCP2
CP2
f11
f21
f31
f42
f52
f62
Var 1
Var 2
Var 3
Var 4
Var 5
Var 6
A1res
E1res
A2res
E2res
A3res
E3res
A4res
E4res
A5res
E5res
A6res
E1res
res_a11
res_e11
res_a22
res_e22
res_a33
res_e33
res_a44
res_e44
res_a55
res_e55
res_a66
res_e66


0.04
0.49
0.46
C1
E1
A1
0.35
0.15
0.40
C2
A2
E2
0.02
CP1
0.32
CP2
0.31
0.94
0.90
0.84
0.95
0.91
0.85
Var 1
Var 2
Var 3
Var 4
Var 5
Var 6
A1res
E1res
A2res
E2res
A3res
E3res
A4res
E4res
A5res
E5res
A6res
E1res
res_a11
res_e11
res_a22
res_e22
res_a33
res_e33
res_a44
res_e44
res_a55
res_e55
res_a66
res_e66
C1res
C2res
C3res
C4res
C5res
C6res