Using Genomic Structural Equation Modeling to Model Joint Genetic Architecture of Complex Traits

Presented by:

Andrew D. Grotzinger & Elliot M. Tucker-Drob

DEPARTMENT OF PSYCHOLOGY

THE UNIVERSITY OF TEXAS AT AUSTIN

Paper:

Grotzinger, A. D., Rhemtulla, M., de Vlaming, R., Ritchie, S. J., Mallard, T. T., Hill, W. D, Ip, H. F., McIntosh, A. M., Deary, I. J., Koellinger, P. D., Harden, K. P., Nivard, M. G., & Tucker-Drob, E. M. (2018). **Genomic SEM provides insights into the multivariate genetic architecture of complex traits.** *bioRχiv.* https://www.biorxiv.org/content/early/2018/04/21/305029

Pervasive (Statistical) Pleiotropy Necessitates Methods for Analyzing Joint Genetic Architecture

Analysis of shared heritability in common disorders of the brain

The Brainstorm Consortium*†

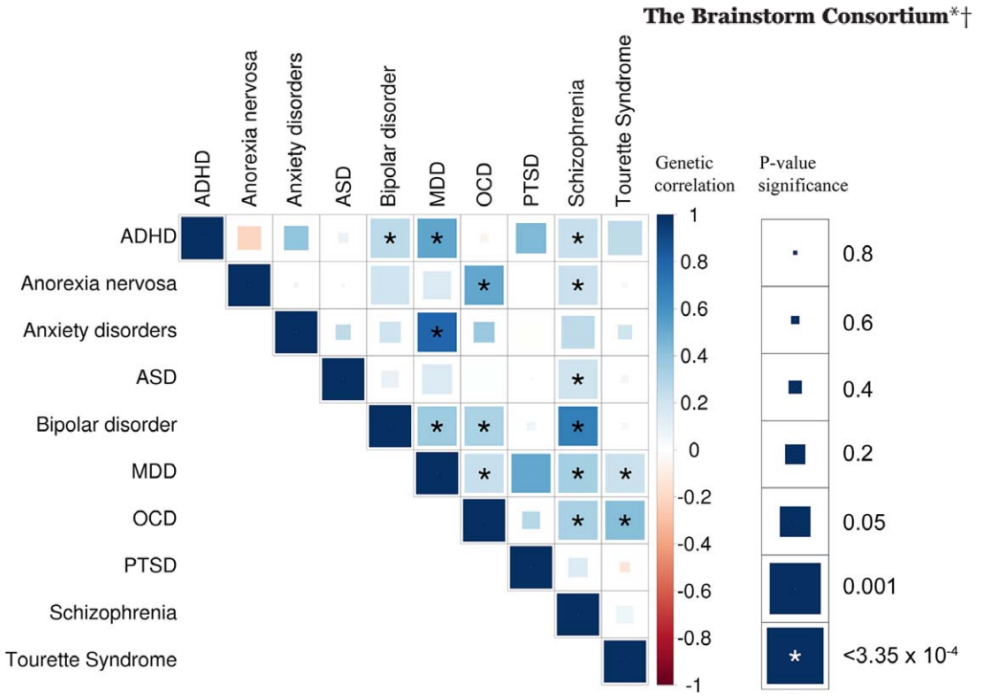


Fig. 1. Genetic correlations across psychiatric phenotypes. The color of each box indicates the magnitude of the correlation, and the size of the box indicates its significance (LDSC), with significant correlations filling each square completely. Asterisks indicate genetic correlations that are significantly different from zero after Bonferroni correction.

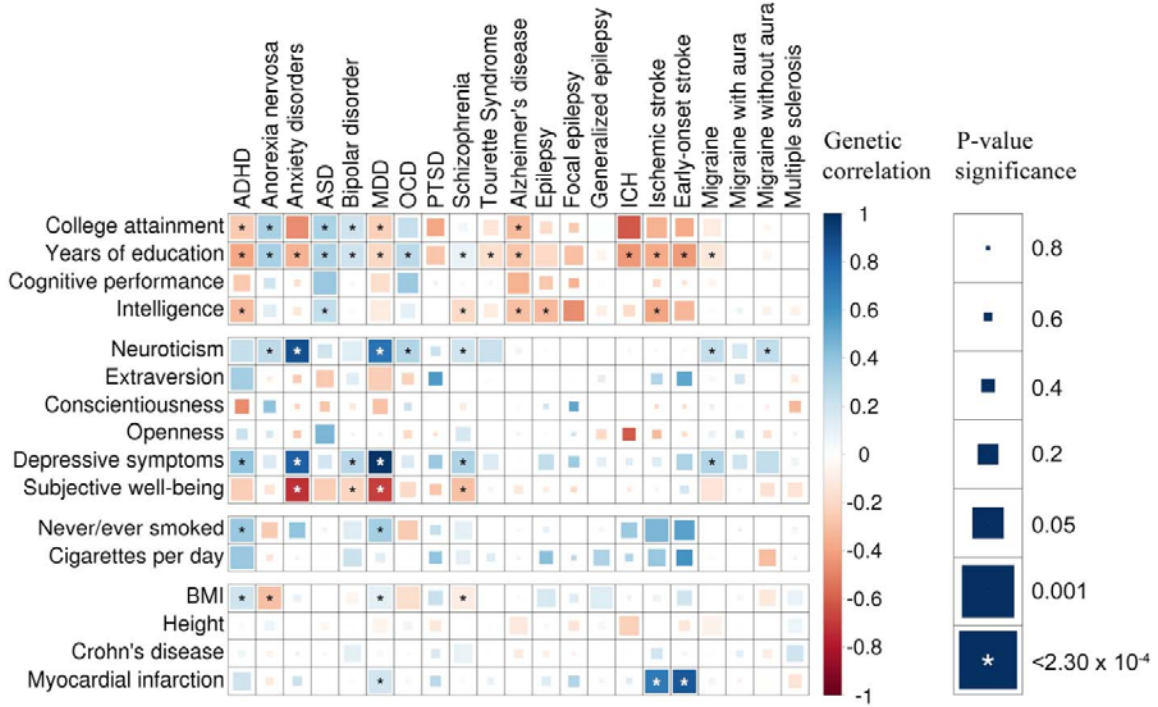


Fig. 4. Genetic correlations across brain disorders and behavioral-cognitive phenotypes. The color of each box indicates the magnitude of the correlation, and the size of the box indicates its significance (LDSC), with significant correlations filling each square completely. Asterisks indicate genetic correlations that are significantly different from zero after Bonferroni correction.

The Brainstorm Consortium, *Science* **360**, 1313 (2018) 22 June 2018

We have a genetic “Atlas.” Now what?

Genetic correlations as data to be modeled, not simply results by themselves

- What data-generating process gave rise to the correlations?
  - Are some more plausible than others?
- Can a high dimensional matrix of genetic correlations among phenotypes be closely approximated with low dimensional representation?

Incorporate joint genetic architecture into multivariate GWAS

- Discovery on latent factors, or residuals of phenotypes after controlling for other phenotypes

Derive novel phenotypes for use in polygenic score analyses

- Polygenic Scores for internalizing psychopathology (e.g. depression, anxiety, neuroticism)
- Polygenic scores for anxiety unique of depression

Genomic Structural Equation Modeling

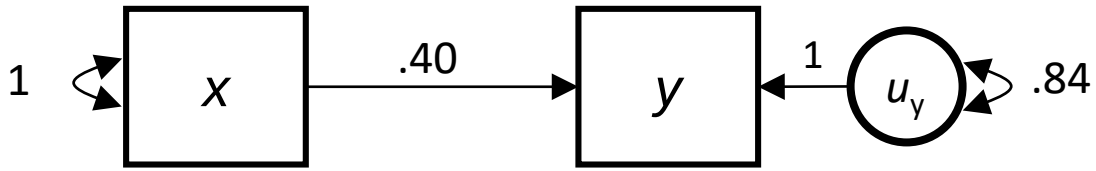
https://www.biorxiv.org/content/early/2018/04/21/305029

- Flexible method for modeling the joint genetic architecture of many traits
- Only requires conventional GWAS summary statistics
- Accommodates varying and unknown amounts of sample overlap
- Can incorporate models of joint genetic architecture into GWAS
  - to aid in multivariate discovery
  - to create polygenic scores for derived phenotypes
- Can be used to formalize Mendelian randomization across large constellations of SNPs and phenotypes
- Free, open source, self-contained R package

A Primer: How does SEM model covariances?

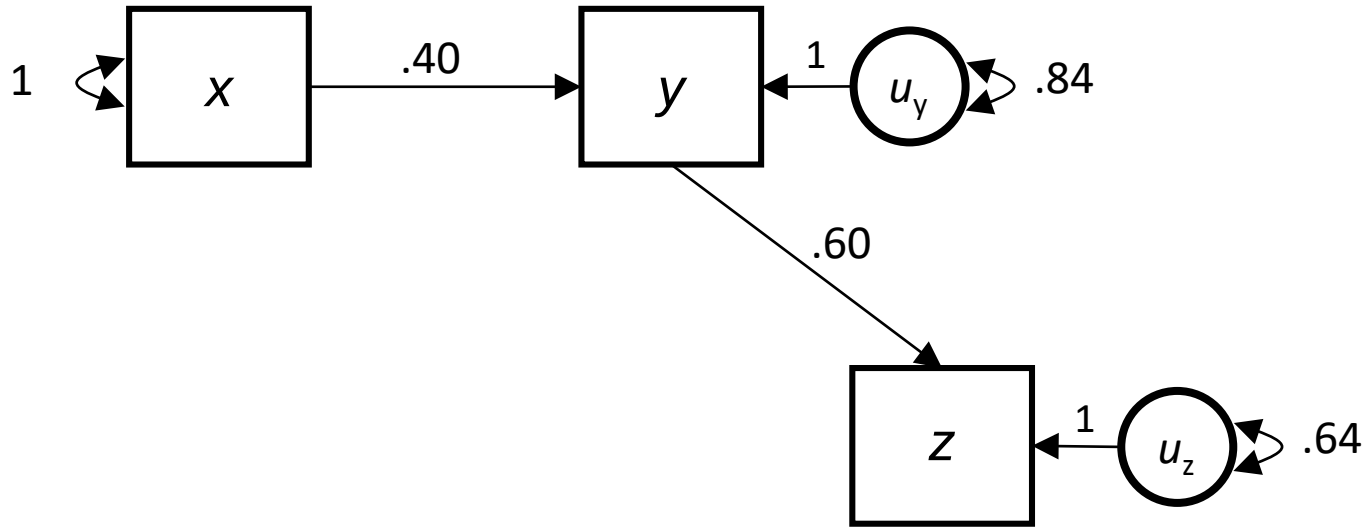
Structural Equation Modeling = structured covariance modeling

Imagine we knew the generating causal process



$y = .40\,x + u_y$ $\quad\quad$ $x \sim (0,1)$, $u_y \sim (0,.84)$

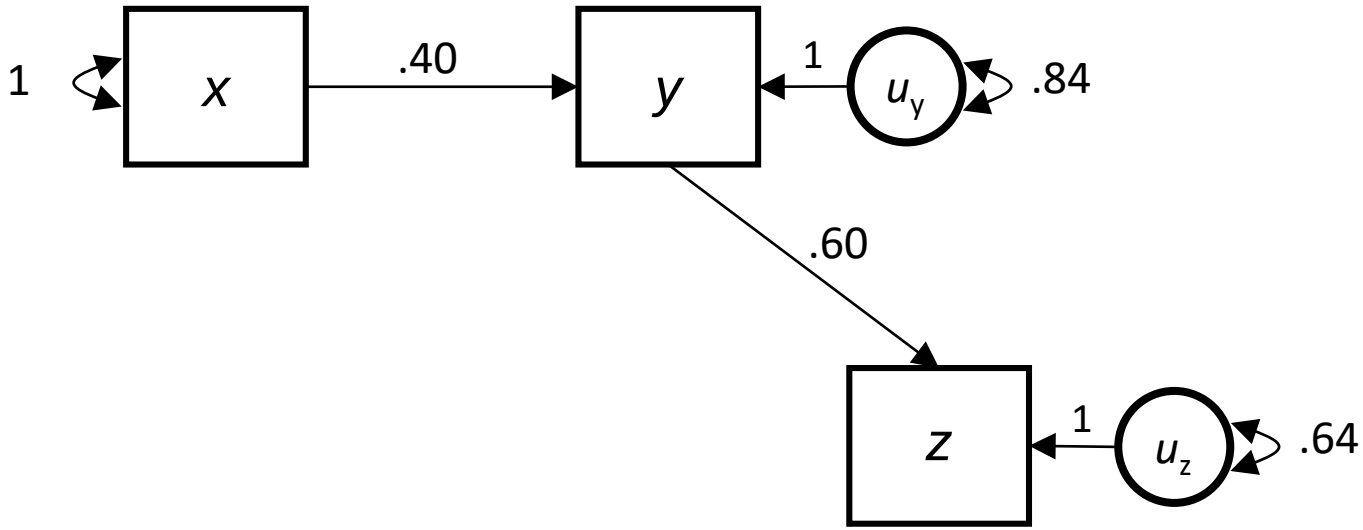
Imagine we knew the generating causal process



$y = .40\,x + u_y$ $\quad\quad$ $x \sim (0,1)$, $u_y \sim (0,.84)$

$z = .60\,y + u_z$ $\quad\quad$ $u_z \sim (0,.64)$

Imagine we knew the generating causal process



Implied covariance matrix
in the population

$\text{cov}(x,y,z)_{pop}$ =

1.00		
.40	1.00	
.24	.60	1.00

$y = .40\,x + u_y$ $\quad\quad$ $x \sim (0,1)$, $u_y \sim (0,.84)$

$z = .60\,y + u_z$ $\quad\quad$ $u_z \sim (0,.64)$

In practice, we only observe the sample data, and we propose a model

observed covariance matrix
in a sample

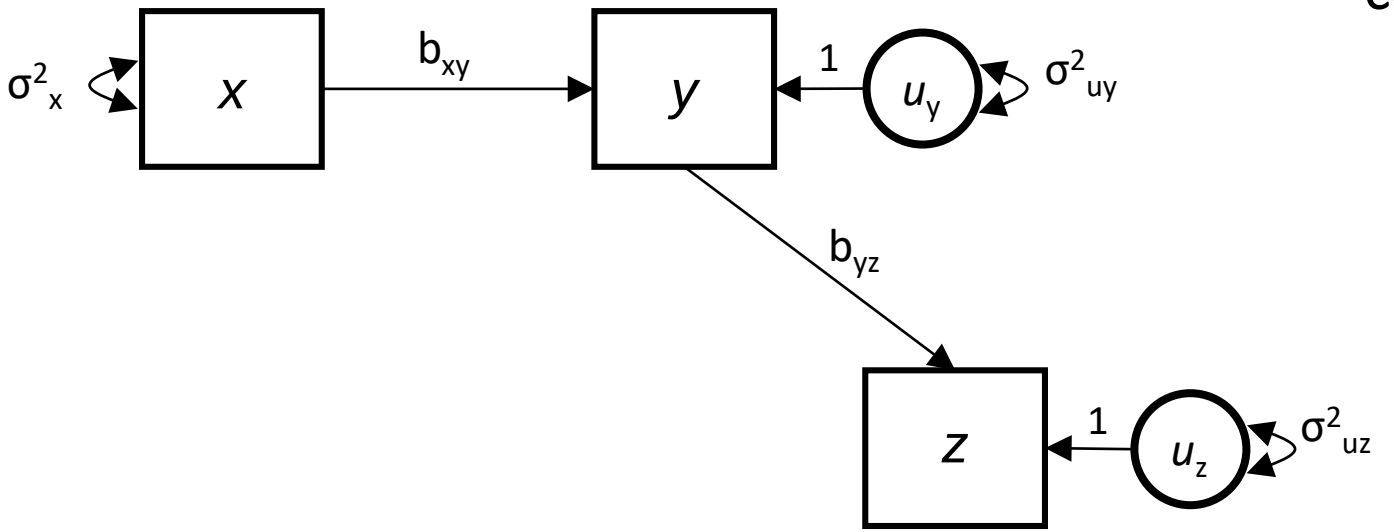
.94		
.33	1.02	
.27	.62	1.02

≈

covariance matrix
in population

1.00		
.40	1.00	
.24	.60	1.00

For the proposed model, estimate parameters from the data, and evaluate model fit to the data



$cov(x,y,z)_{sample}$ =

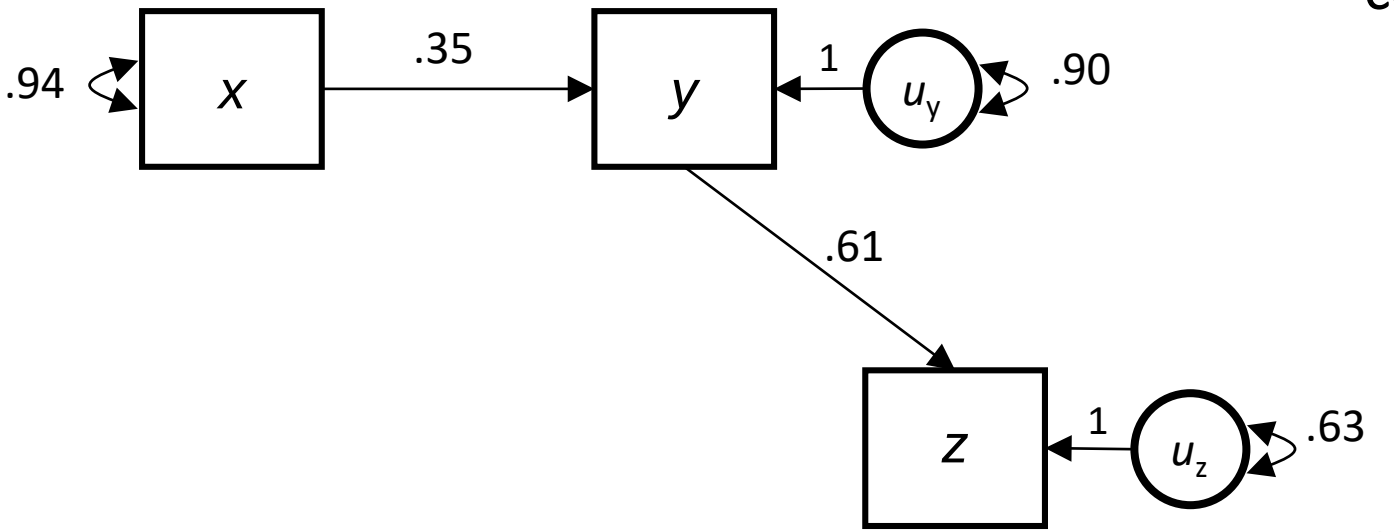
.94		
.33	1.02	
.27	.62	1.02

6 unique elements in the covariance matrix being modeled

5 free model parameters

1 df

For the proposed model, estimate parameters from the data, and evaluate model fit to the data



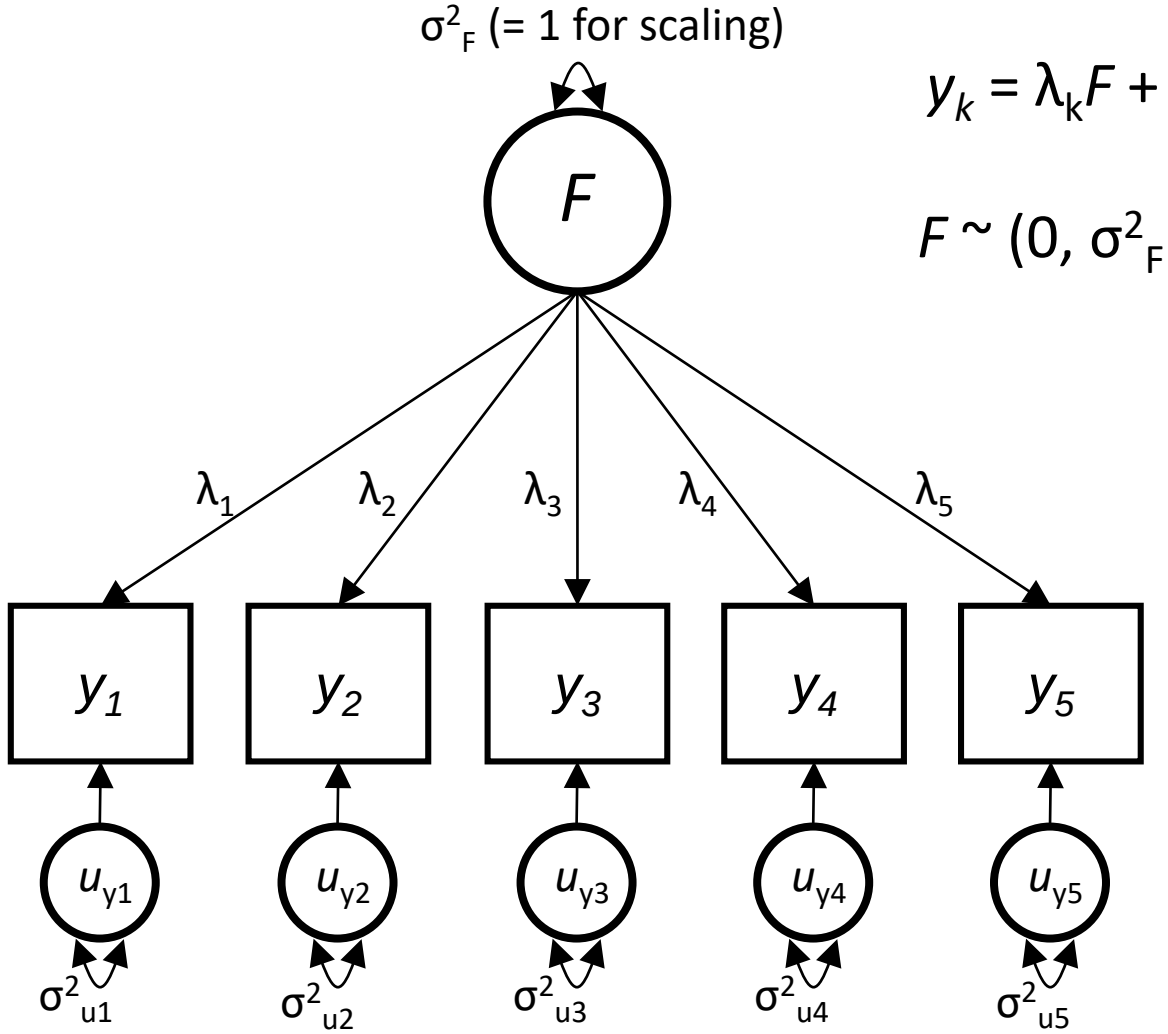
$\text{cov}(x,y,z)_{\text{sample}}$ =

.94		
.33	1.02	
.27	.62	1.02

$\text{cov}(x,y,z)_{\text{implied}}$ =

.94		
.33	1.03	
.20	.63	1.00

The model that we fit may include some variables for which we do not observe data

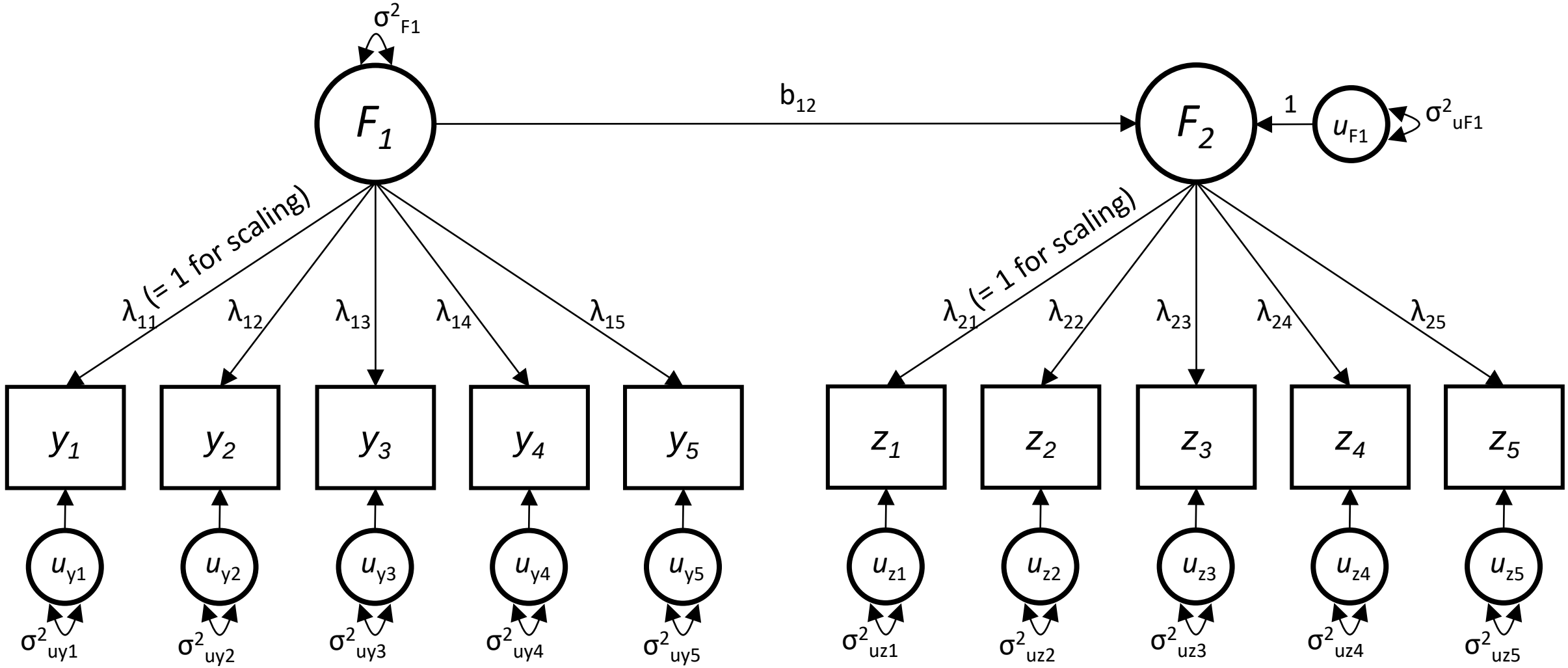


$y_k = \lambda_k F + u_k$

$F \sim (0, \sigma^2_F)$, $u_{yk} \sim (0, \sigma^2_{uk})$

F is unobserved.
Parameters are estimated from,
and fit is evaluated relative to,
the sample covariance matrix for y_1-y_k.

The model that we fit may include some variables for which we do not observe data



Genomic SEM uses these principles to fit structural equation models to genetic covariance matrices derived from GWAS summary statistics using 2 Stage Estimation

- Stage 1: Estimate Genetic Covariance Matrix and associated matrix of standard errors and their codependencies
  - We use LD Score Regression, but any method for estimating this matrix (e.g. GREML) and its sampling distribution can be used
- Stage 2: Fit a Structural Equation Model to the Matrices from Stage 1

Fitting Structural Equation Models to GWAS-Derived Genetic Covariance Matrices

R package: GenomicSEM

```
install.packages("devtools")
library(devtools)
install_github("MichelNivard/GenomicSEM")
library(GenomicSEM)
```

Start with GWAS Summary Statistics for the Phenotypes of Interest

- No need for raw data
- No need to conduct a primary GWAS yourself: Download them online!
  - sumstats for over 3700 phenotypes have been helpfully indexed at http://atlas.ctglab.nl/
  - sumstats for over 4000 UK Biobank phenotypes are downloadable at http://www.nealelab.is/uk-biobank

CHR	SNP	BP	A1	A2	INFO	OR	SE	P	Nca	Nco	MAF
8	rs62513865	101592213	T	C	0.957	1.01461	0.0153	0.3438	59851	113154	0.07330
8	rs79643588	106973048	A	G	0.999	1.02122	0.0136	0.1231	59851	113154	0.09200
8	rs17396518	108690829	T	G	0.980	1.00331	0.0080	0.6821	59851	113154	0.43500
8	rs6994300	102569817	A	G	0.466	0.88126	0.4243	0.7658	16823	25632	0.00556
8	rs138449472	108580746	A	G	0.734	0.97181	0.0598	0.6320	41253	79756	0.00852
8	rs983166	108681675	A	C	0.991	0.99144	0.0080	0.2784	59851	113154	0.43200

Prepare the data for LDSC: Munge

- Aligns allele sign across sumstats for all traits
- Computes z-statistics needed for LDSC
- Restricts to common SNPs (MAF>.01) on reference panel
- Function requires:
  1. names of the summary statistics files
  2. name of the reference file. Hapmap 3 SNPs *(downloadable on our wiki)* with the MHC region removed is standard
     (well-imputed and well-known LD structure)

  3. trait names that will be used to name the saved files

```
munge(c("scz.txt", "bip.txt", “mdd.txt",
"ptsd.txt","anx.txt"),
"w_hm3.noMHC.snplist",trait.names=c("scz",
"bip","mdd","ptsd","anx"))
```

Stage 1 Estimation: Multivariable LDSC

Create a genetic covariance matrix, S: an “atlas of genetic correlations”

Diagonal elements are (heritabilities)

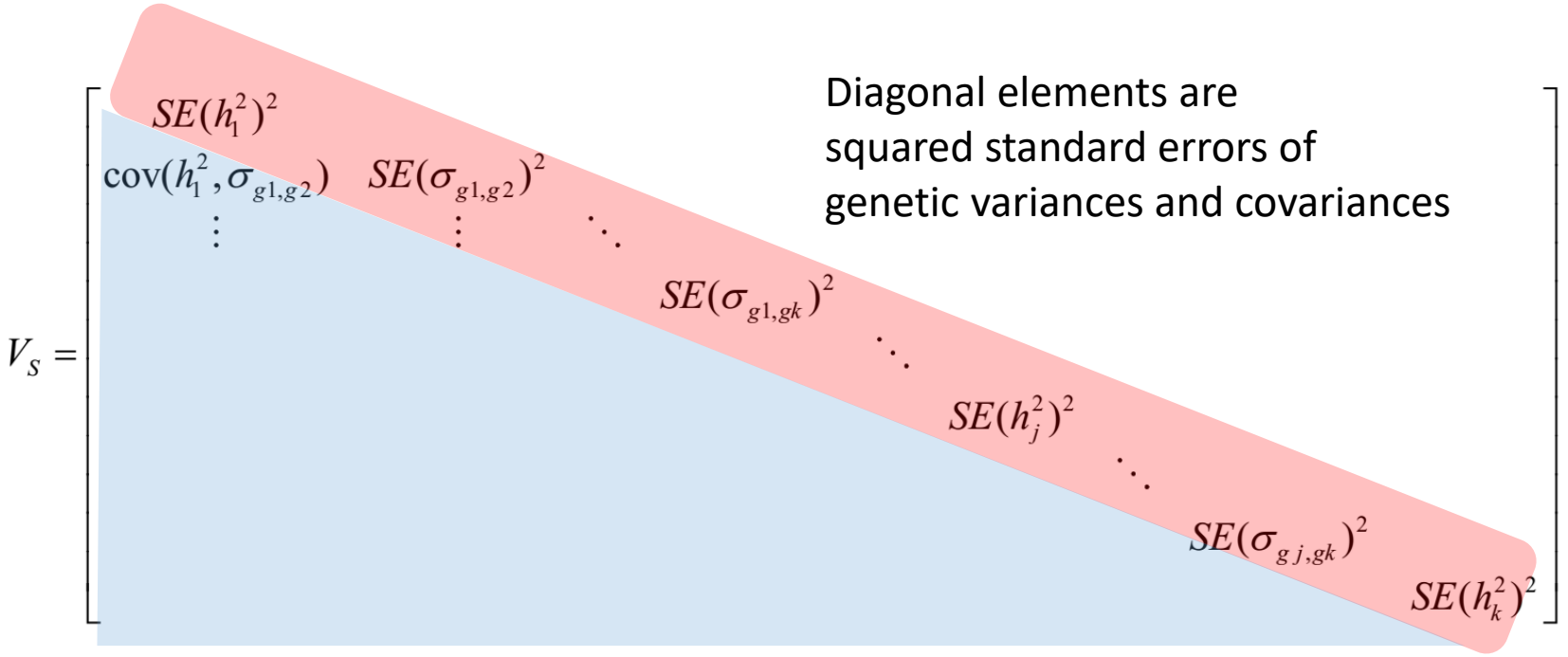
$$S = \begin{bmatrix} h_1^2 & \sigma_{g1,g2} & \cdots & \sigma_{g1,gk} \\ & h_2^2 & \cdots & \sigma_{g2,gk} \\ & & \ddots & \vdots \\ & & & h_k^2 \end{bmatrix}$$

Off-diagonal elements are coheritabilities

```
sumstats <- c("scz.sumstats.gz",
"bip.sumstats.gz",“EA.sumstats.gz")

#for case control phenotypes
sample.prev <- c(.39,.45,NA)
population.prev <- c(.01,.01,NA)

ld <- "eur_w_ld_chr/"

trait.names<-c("SCZ","BIP",“EA")

LDSCoutput <- ldsc(sumstats, sample.prev,
population.prev, ld, ld, trait.names)
```

Stage 1 Estimation: Multivariable LDSC

Also produced is a second matrix, V, of squared standard errors and the dependencies between estimation errors



Diagonal elements are squared standard errors of genetic variances and covariances

Off-diagonal elements are dependencies between estimation errors used to directly model dependencies that occur due to sample overlap from contributing GWASs

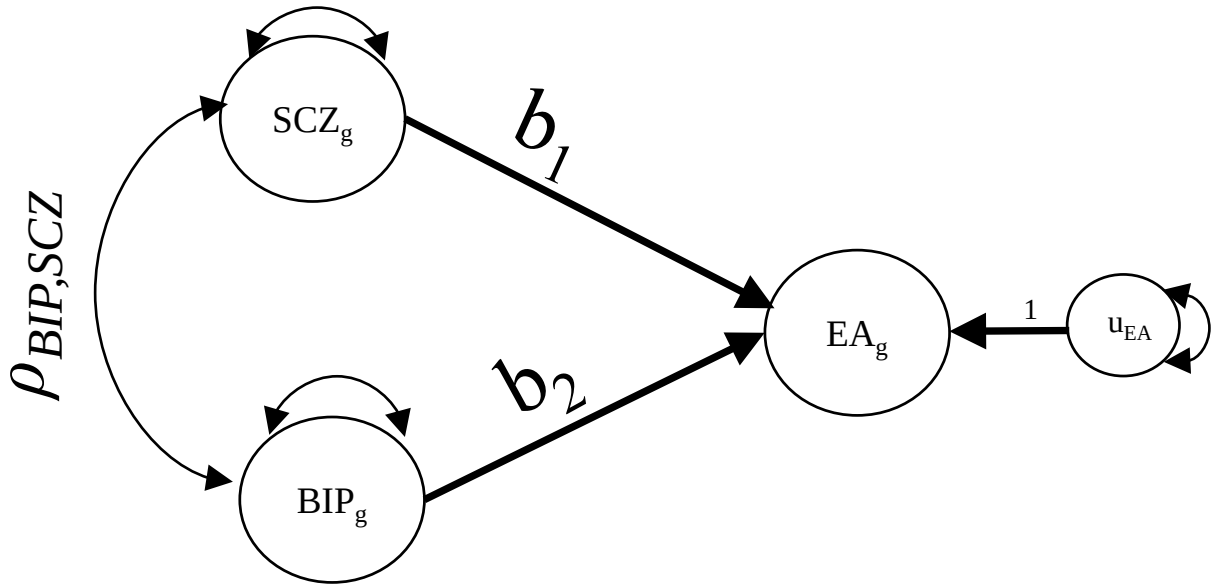
Stage 2 Estimation: Specify the SEM

Example: Genetic multiple regression

$$EA_g = b_1 \times SCZ_g + b_2 \times BIP_g + u$$

S =

SCZ		
.57	BIP	
.15	.27	EA



(df = 0, model parameters are a simply a transformation of the matrix)

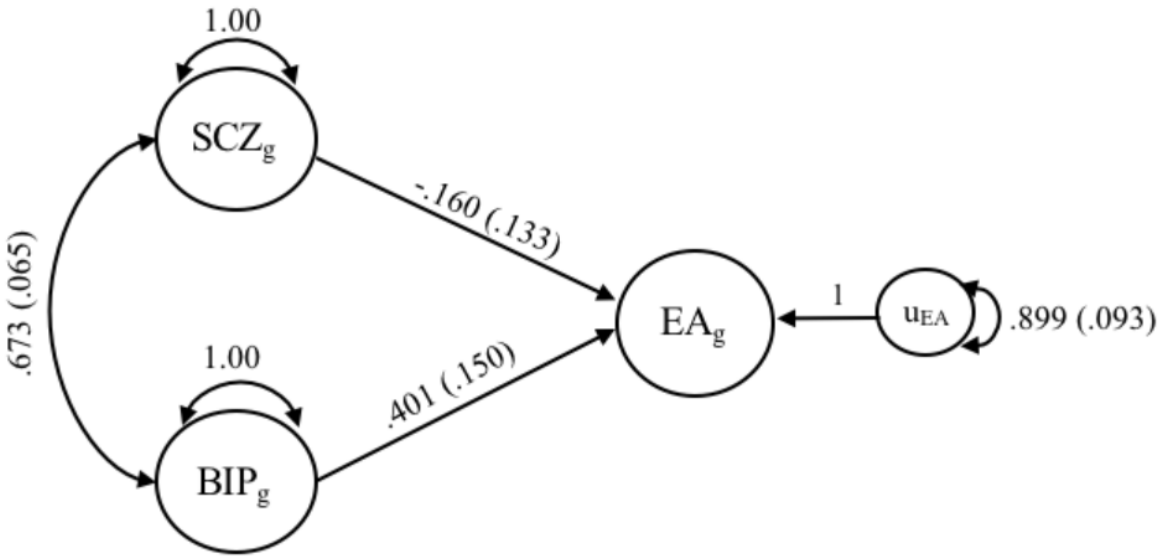
Stage 2 Estimation: Specify the SEM

```
REGmodel <- 'EA ~ SCZ + BIP
             SCZ~~BIP'

#run the model using the user defined function
REGoutput<-usermodel(LDSCoutput, model = REGmodel)

#print the output
REGoutput
```

RESULTS

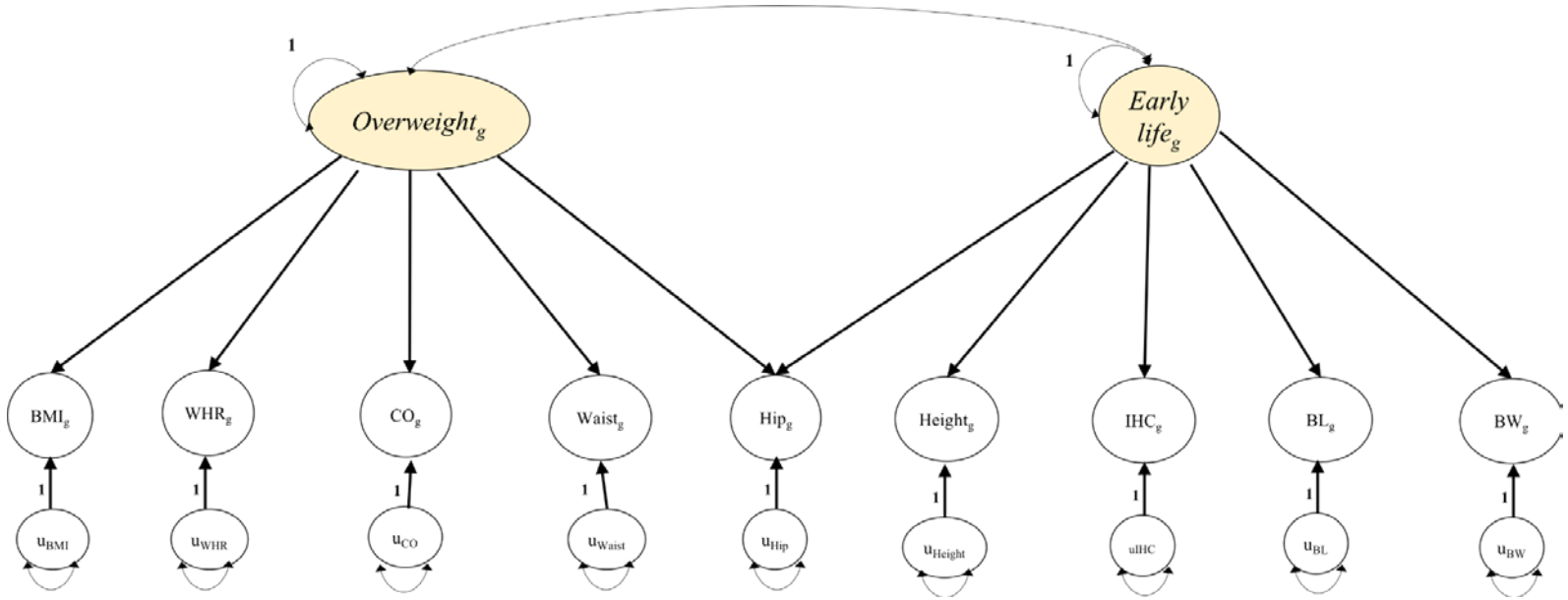
```
$results
   lhs op rhs Unstandardized_Estimate Unstandardized_SE Standardized_Est Standardized_SE
1   EA  ~ SCZ             -0.09305117       0.077550529       -0.1603337      0.13362500
2   EA  ~ BIP              0.31902692       0.119496510        0.4013460      0.15033041
3  SCZ ~~ BIP              0.12289914       0.011845099        0.6727780      0.06484279
10 SCZ ~~ SCZ              0.25020062       0.017482875        1.0000000      0.06987543
11 BIP ~~ BIP              0.13337232       0.013696265        1.0000000      0.10269196
12  EA ~~  EA              0.07582781       0.007838676        0.8998001      0.09301655
```

$EA_g = -.016 \times SCZ_g + \mathbf{.283} \times BIP_g + u$



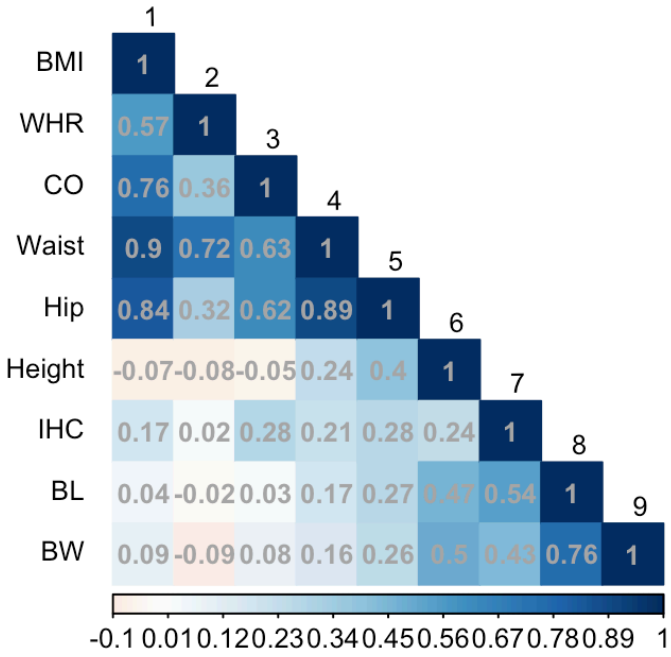
Example 2: Genetic Factor Analysis of Anthropometric Traits

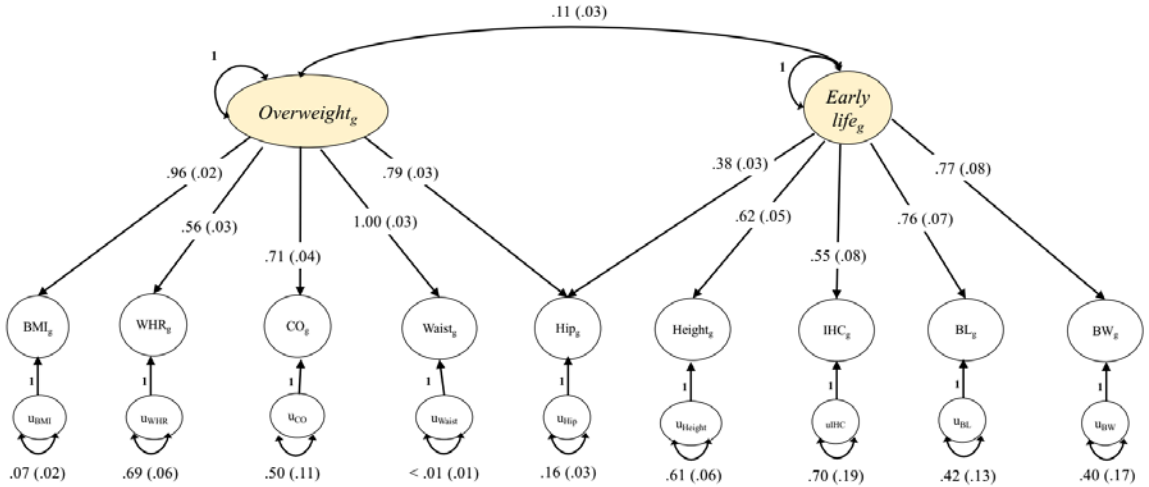
```
TwoFactor <- 'F1 =~ NA*BMI + WHR + CO + Waist + Hip
              F2 =~ NA*Hip + Height + IHC + BL + BW
               F1~~1*F1
               F2~~1*F2
               F1~~F2'

#run the model
Anthro<-usermodel(anthro, model = TwoFactor)

#print the results
Anthro
```



Example 2: Genetic Factor Analysis of Anthropometric Traits

Genetic Correlation Matrix





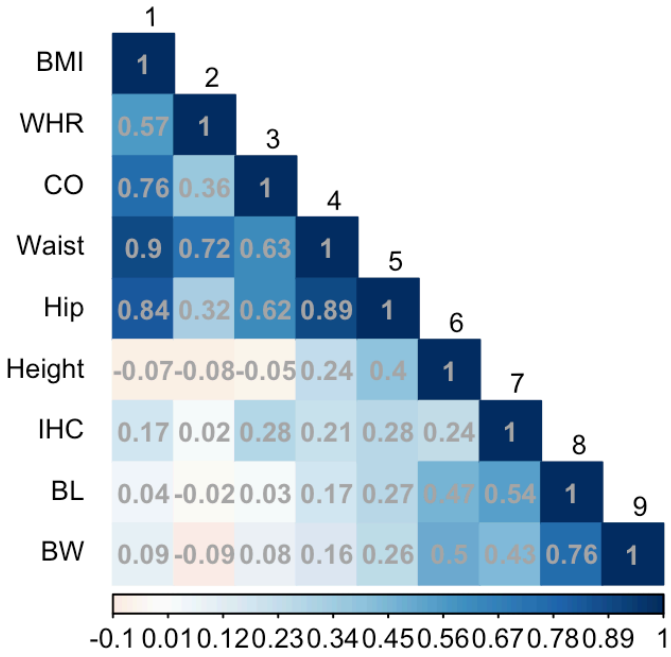
df = 25, CFI = .951, SRMR = .089

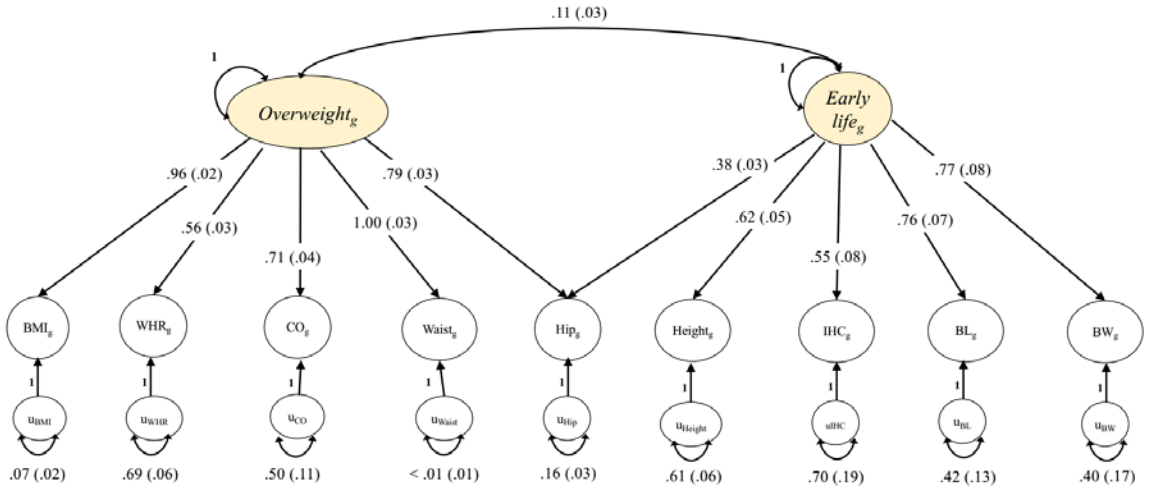
BMI = body mass index; WHR = waist-hip ratio; CO = childhood obesity; IHC = infant head circumference; BL = birth length; BW = birth weight.

sumstats from EGG and GIANT Consortia

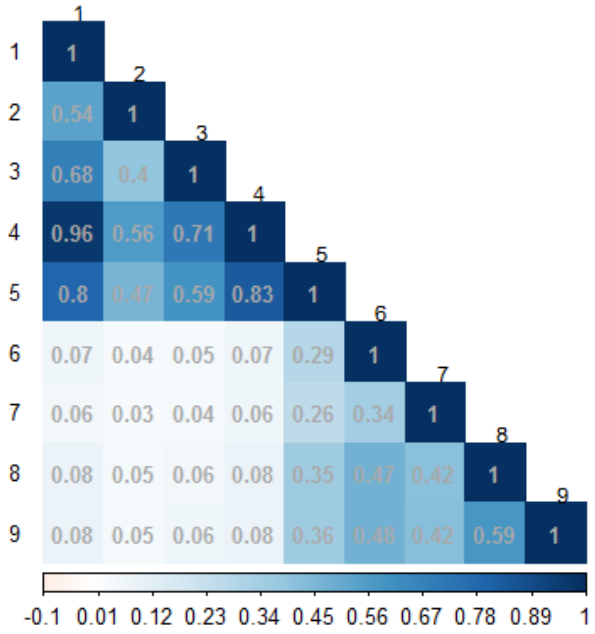
Example 2: Genetic Factor Analysis of Anthropometric Traits

Genetic Correlation Matrix





Model-Implied Matrix

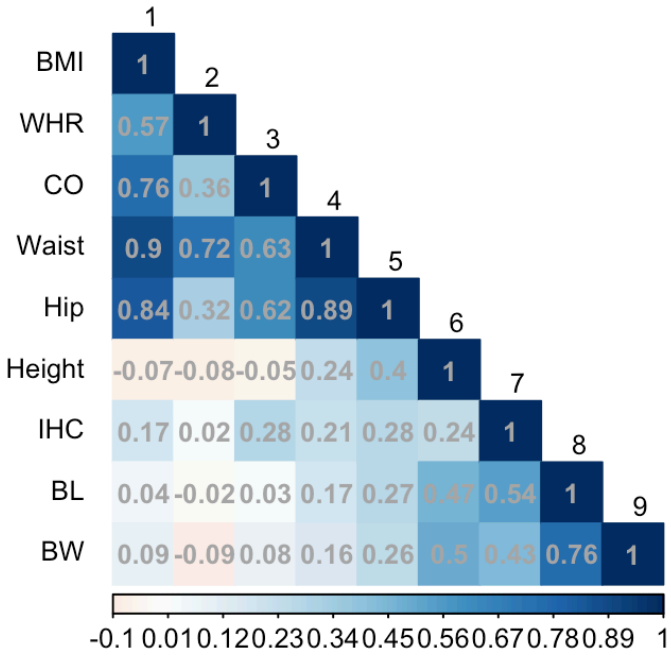


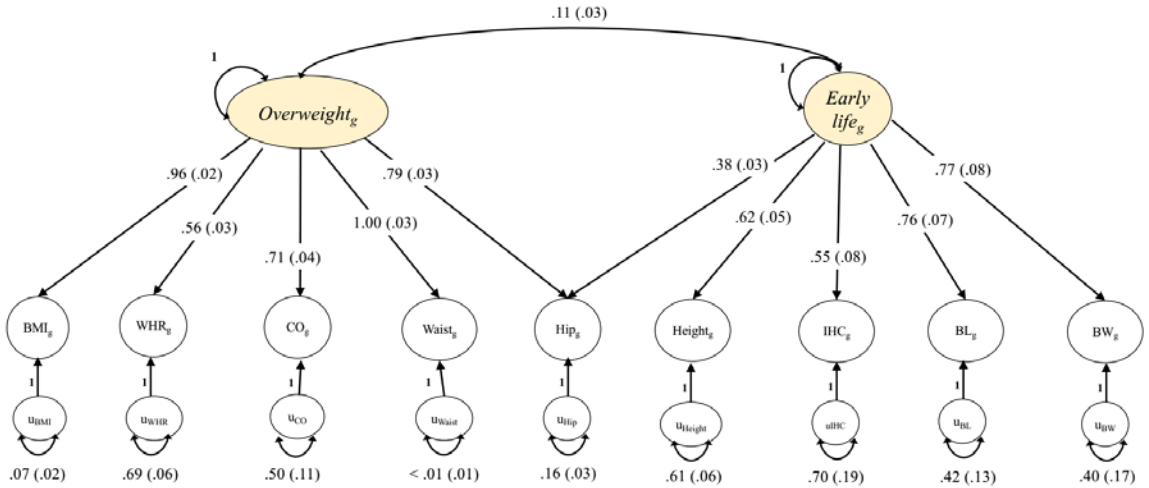
df = 25, CFI = .951, SRMR = .089

BMI = body mass index; WHR = waist-hip ratio; CO = childhood obesity; IHC = infant head circumference; BL = birth length; BW = birth weight.

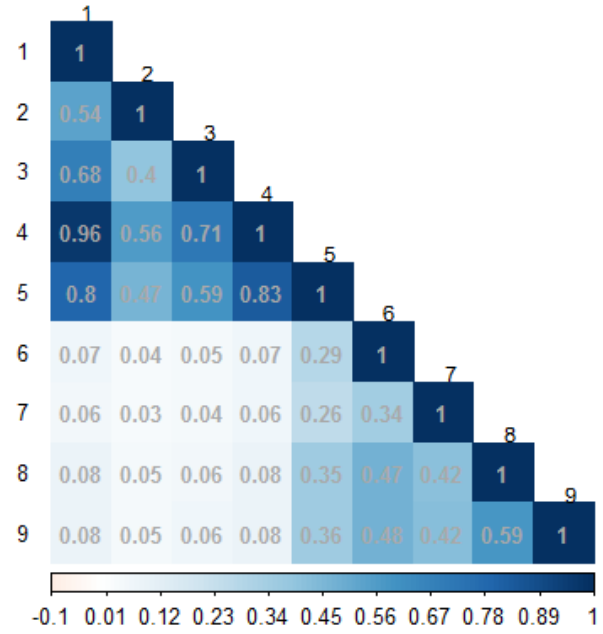
Example 2: Genetic Factor Analysis of Anthropometric Traits

Genetic Correlation Matrix



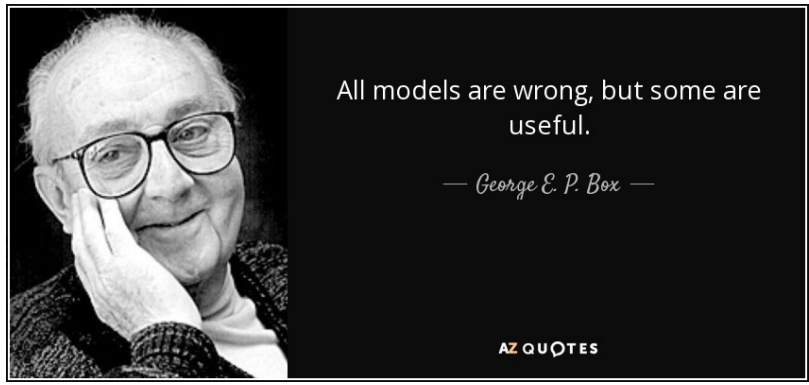


Model-Implied Matrix



df = 25, CFI = .951, SRMR = .089

BMI = body mass index; WHR = waist-hip ratio; CO = childhood obesity; IHC = infant head circumference; BL = birth length; BW = birth weight.

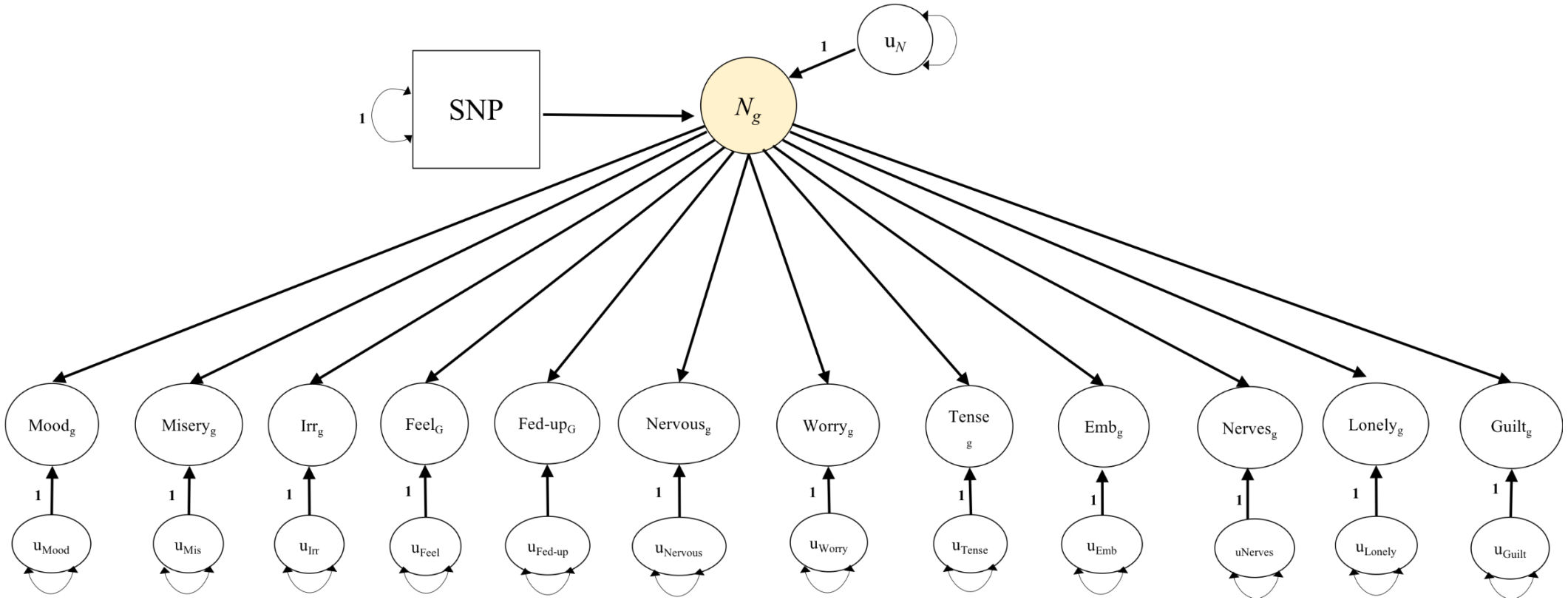


Incorporating Genetic Covariance Structure into Multivariate GWAS Discovery

Andrew

Example: Item level analysis of Neuroticism

- Univariate summary statistics for each of 12 individual items in UKB downloaded from Neale lab website.



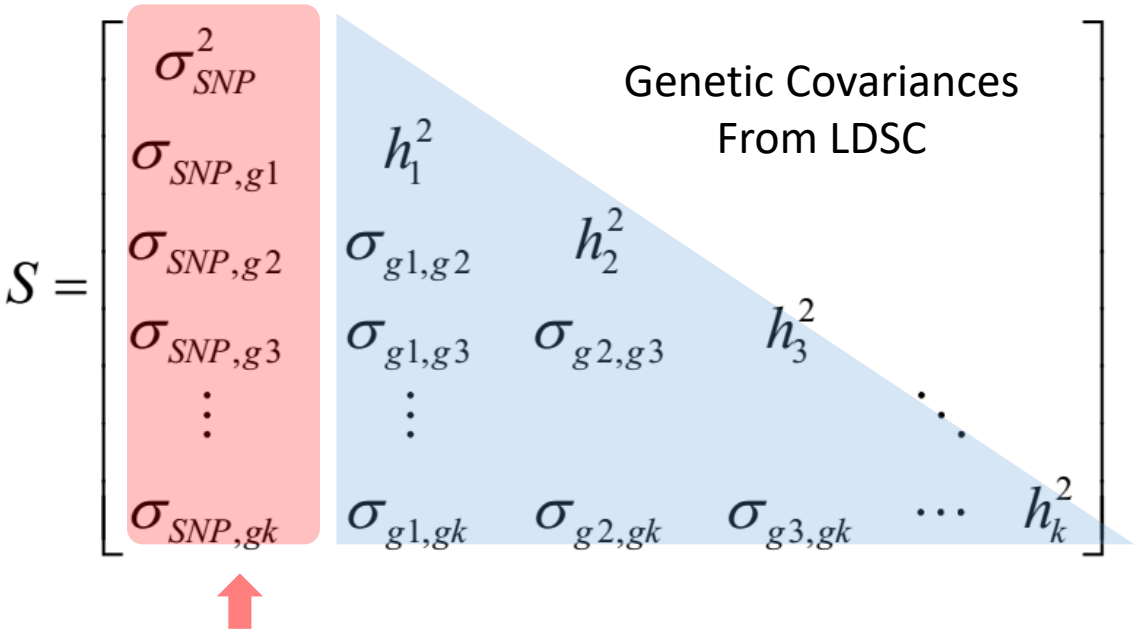
Prepare Summary Statistics:

- Aligns allele sign across sumstats for all traits
- Converts odds ratios and “linear probability model” coefficients into logistic regression coefficients
  - Converts corresponding standard errors
- Standardizes effect sizes to phenotypic variance = 1

```
ss=c("item1.txt","item2.txt","item3.txt", "item4.txt", "item5.txt", "item6.txt",
"item7.txt", "item8.txt", "item9.txt", "item10.txt", "item11.txt", "item12.txt")
refpan="reference.1000G.maf.0.005.txt"
items=c("N1","N2","N3","N4","N5","N6","N7","N8","N9",N10","N11","N12")
se.l=c(F,F,F,F,F,F,F,F,F,F,F,F)
lp=c(T,T,T,T,T,T,T,T,T,T,T,T)
propor<-c(.451,.427,.280,.556,.406,.237,.568,.171,.478,.213,.177,.283)

processed_sumstats <-
sumstats(files=ss,ref=refpan,trait.names=items,se.logit=se.l,linprob=lp,prop=propor)
```

Add SNP Effects to the “Atlas”

Expand S to include SNP Effects

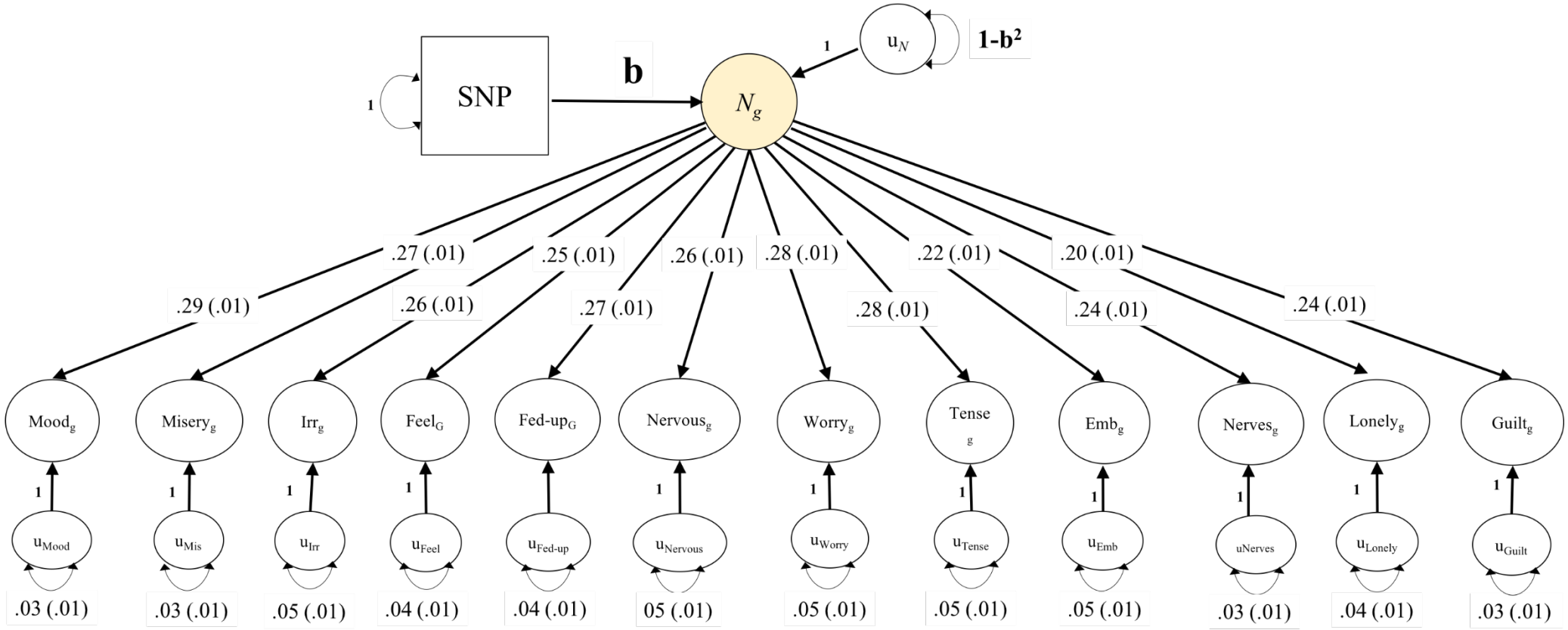


Betas from
GWAS sumstats
scaled to covariances
using MAFs

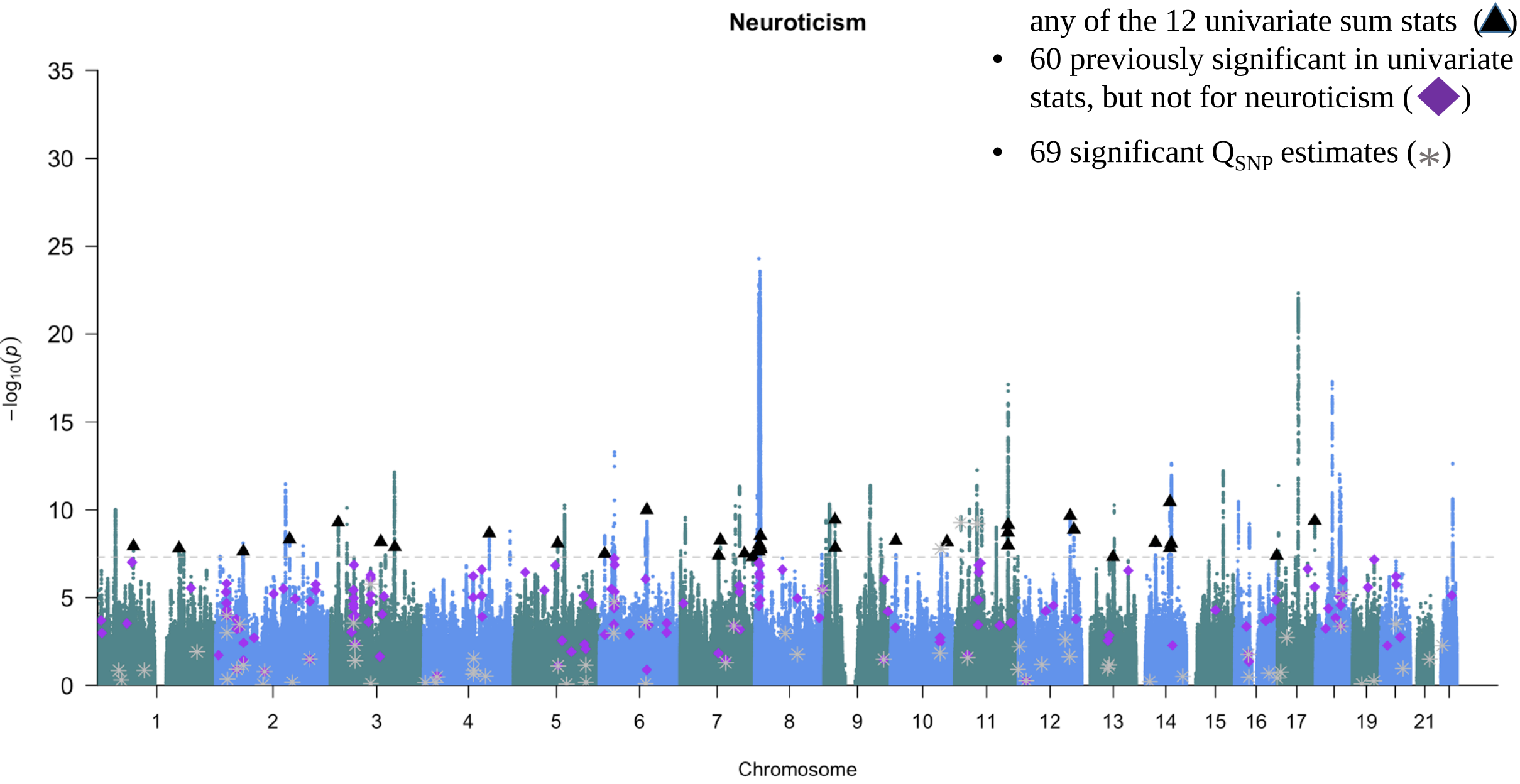
```
SNPcov<-
addSNPs(LDSCoutput,processed_sumstats)
```

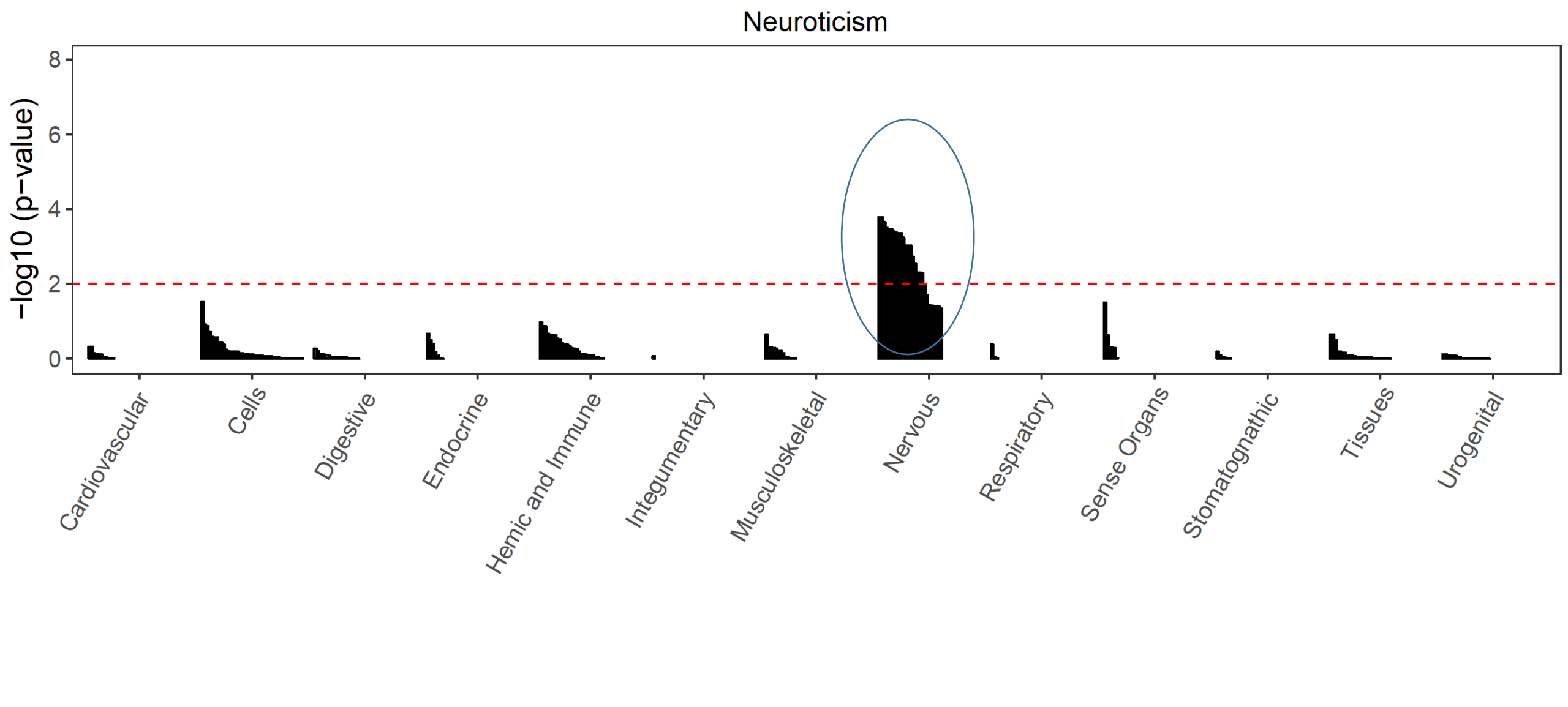
Run the model

```
NeurModel<-commonfactorGWAS(SNPcov)
```

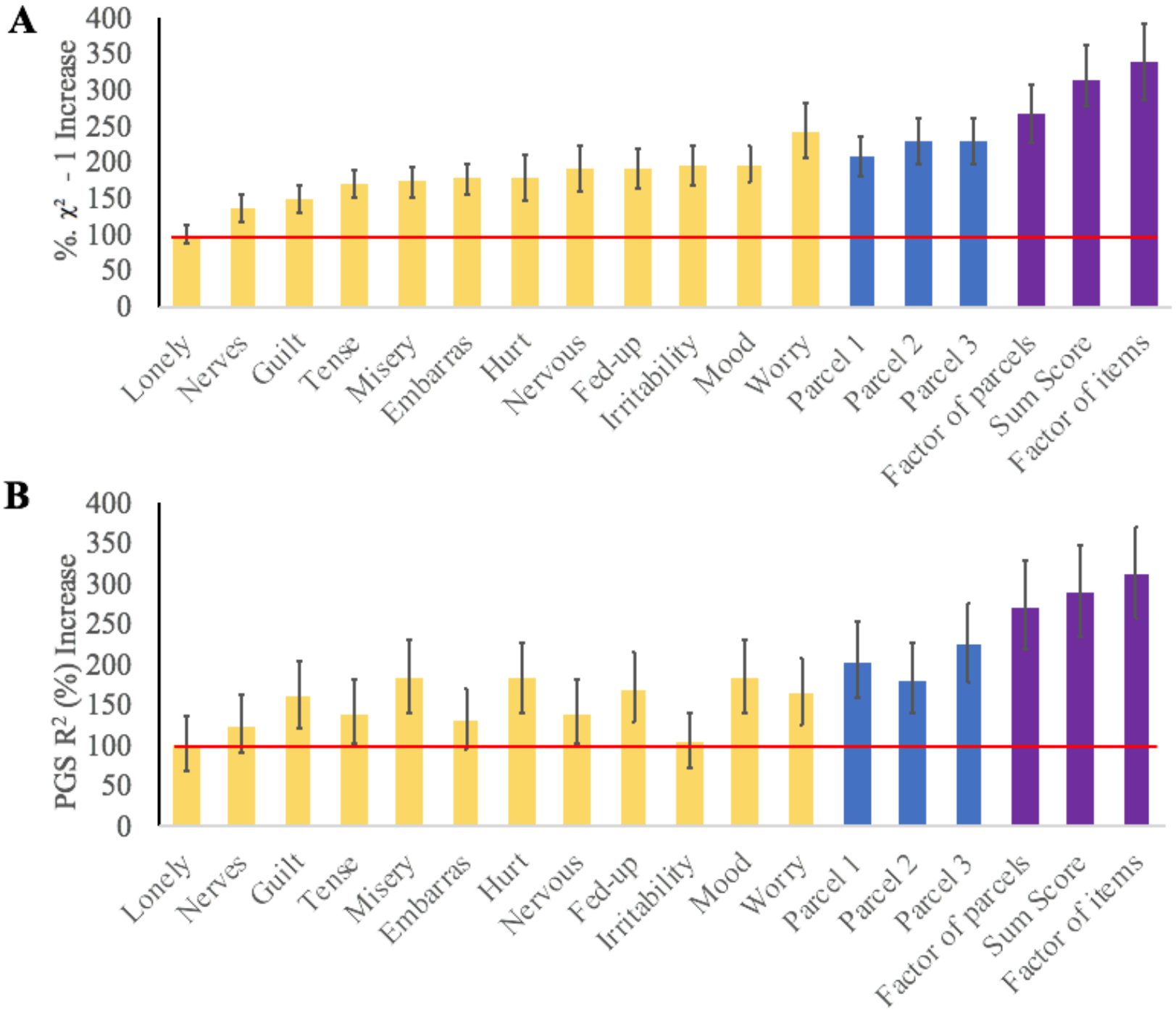


- 118 lead SNPs
- 38 unique loci not previously identified in any of the 12 univariate sum stats (▲)
- 60 previously significant in univariate sum stats, but not for neuroticism (◆)
- 69 significant Q_{SNP} estimates (✳)




Neuroticism
−log10 (p−value)
8
6
4
2
0
Cardiovascular
Cells
Digestive
Endocrine
Hemic and Immune
Integumentary
Musculoskeletal
Nervous
Respiratory
Sense Organs
Stomatognathic
Tissues
Urogenital

Relative Power



Genomic SEM is a broad framework
not just one model

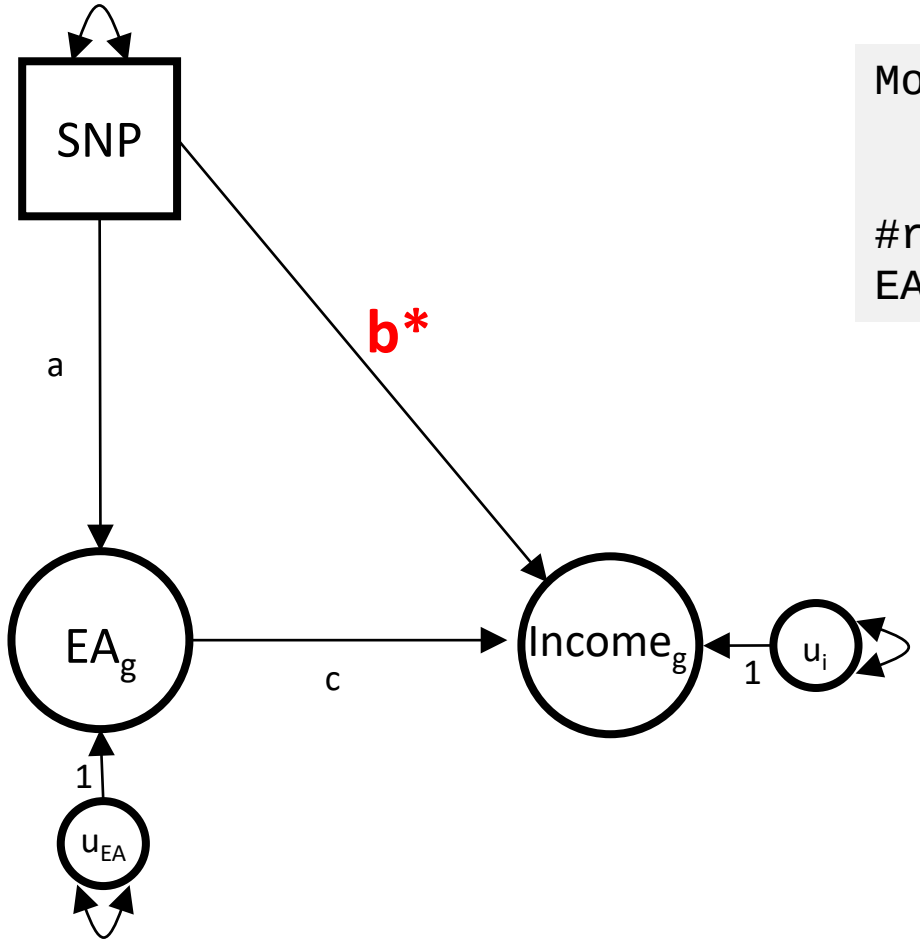
- Genomic SEM is a statistical framework (and freely available standalone software package) for estimating a nearly limitless number of ***user specified*** models to multivariate GWAS summary statistics
- Lots of other possibilities, e.g.:
  - Deriving Polygenic Scores for “Residual” Phenotypes
  - Mendelian-Randomization within Multivariate Networks

Empirical example

- Are the socioeconomic sequelae of ADHD mediated by educational attainment?

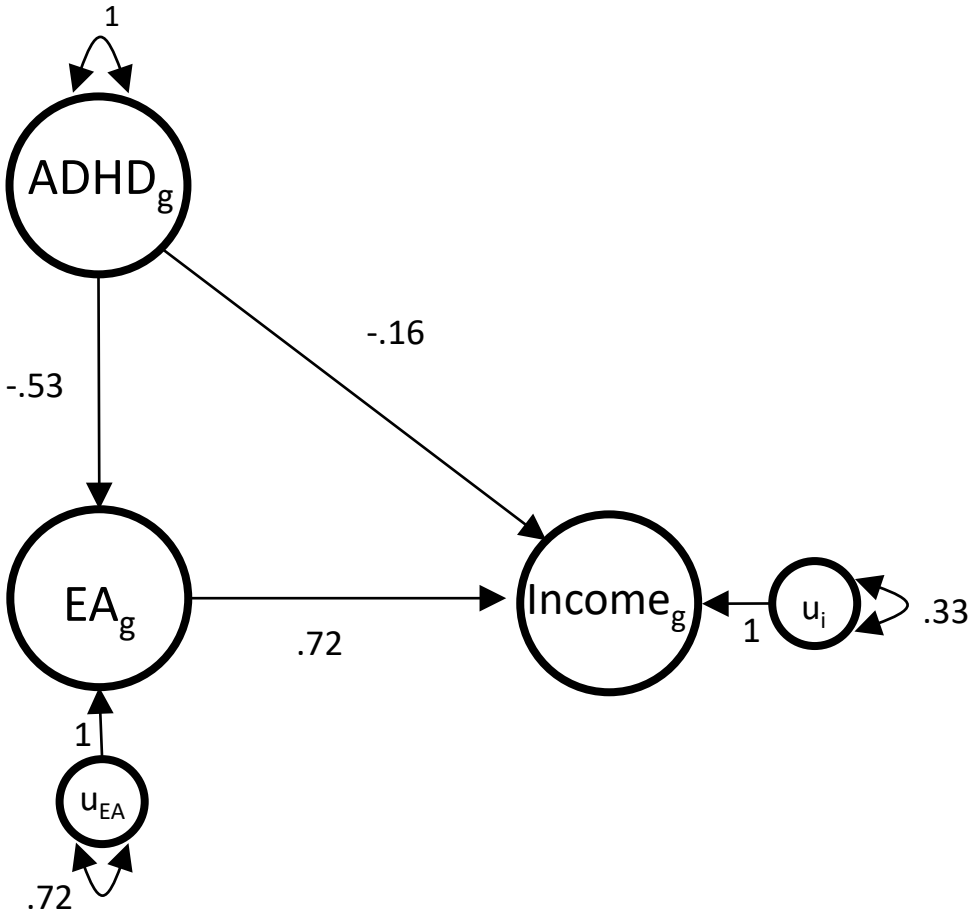
- Relevant because if true, staying in school may become a treatment goal for ADHD.

Creating sumstats (and computing polygenic scores) for a derived phenotype, e.g. a residual




```
Model1 <- 'EA ~ SNP
           Income ~ EA SNP'

#run the model
EA_Inc<-userGWAS(SNPcov, model = Model1)
```

Genetic Mediation in Latent Genetic Space

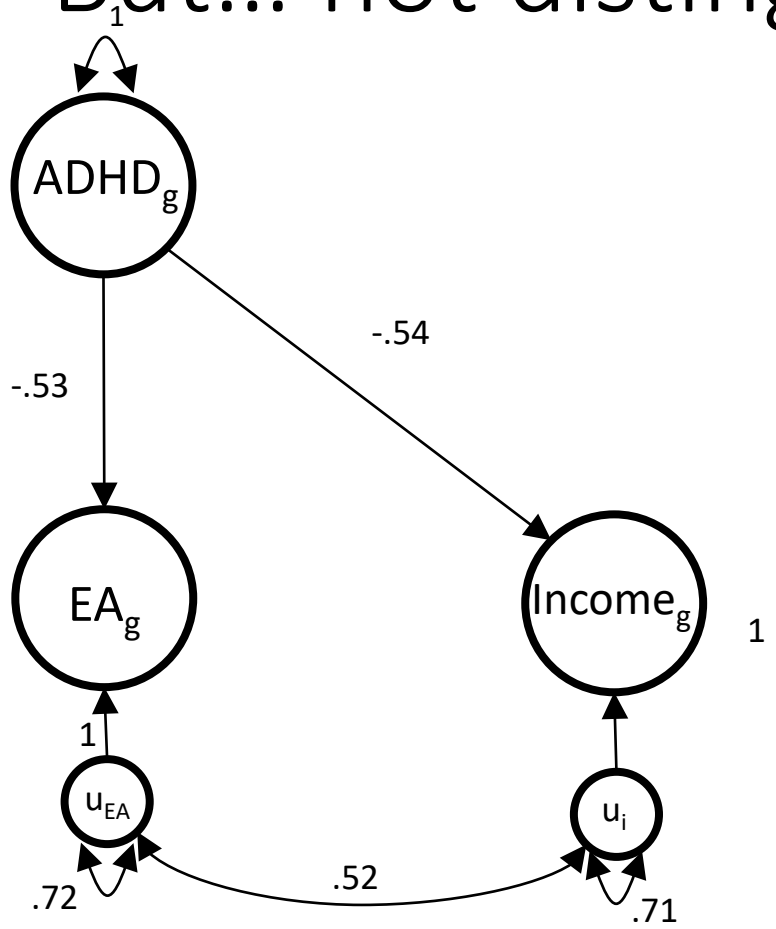



```
Model2 <- 'EA ~ ADHD
           Income ~ EA ADHD'

#run the model
ADHD_EA_Inc<-usermodel(LDSCoutput, model =
Model2)
```

Summary Statistics:

- ADHD (Demontis et al., 2017)
- Educational Attainment (Okbay et al. 2016)
- Income (Hill et al., 2016)

But… not distinguishable from other models




```
Model3 <- 'EA ~ ADHD
           Income ~ ADHD
           EA ~~ Income'

#run the model
ADHD_EA_Inc<-usermodel(LDSCoutput, model =
Model2)
```

Summary Statistics:

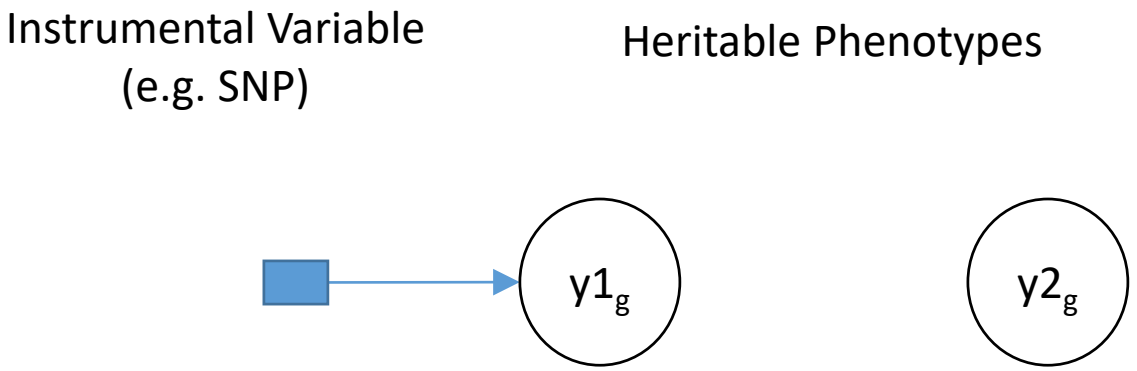
- ADHD (Demontis et al., 2017)
- Educational Attainment (Okbay et al. 2016)
- Income (Hill et al., 2016)

Identifying Plausible Causal Pathways:
Mendelian Randomization in Multivariate Networks

- Genomic SEM models genetic covariance structure
- Genomic SEM allows for SNPs in the model
- These can be combined to perform Mendelian Randomization (MR)

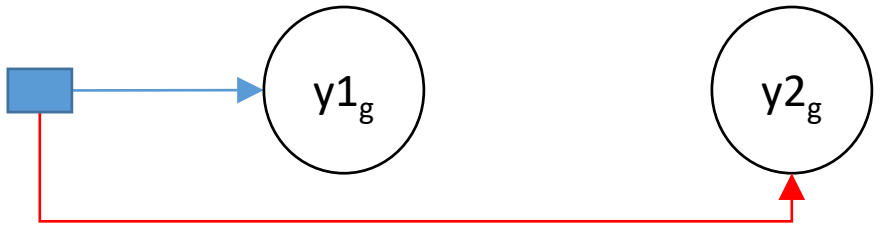
MR in Genomic SEM

- Mendelian randomization using GWAS summary data



MR in Genomic SEM

- Mendelian randomization using GWAS summary data

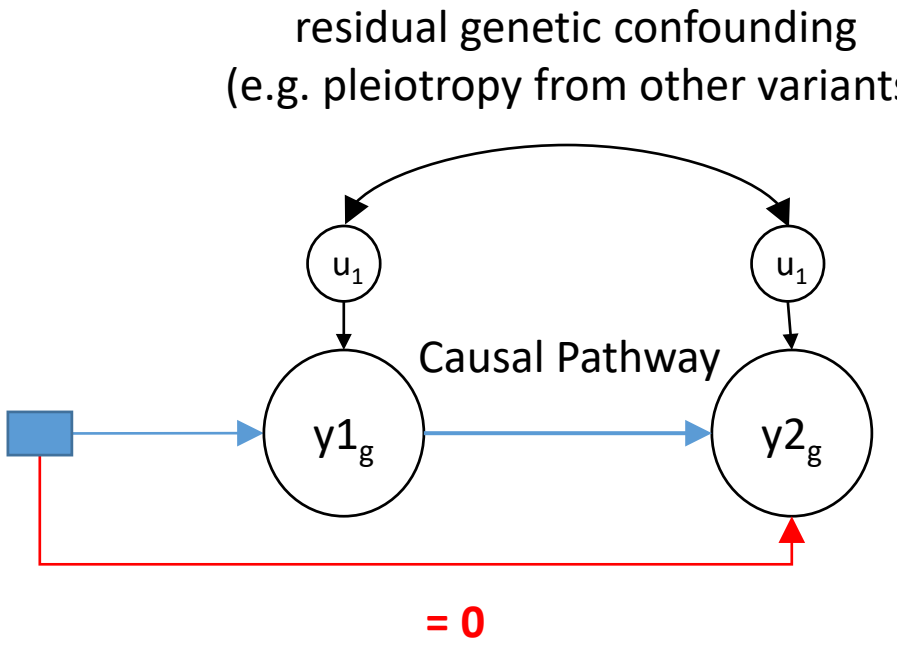


= 0

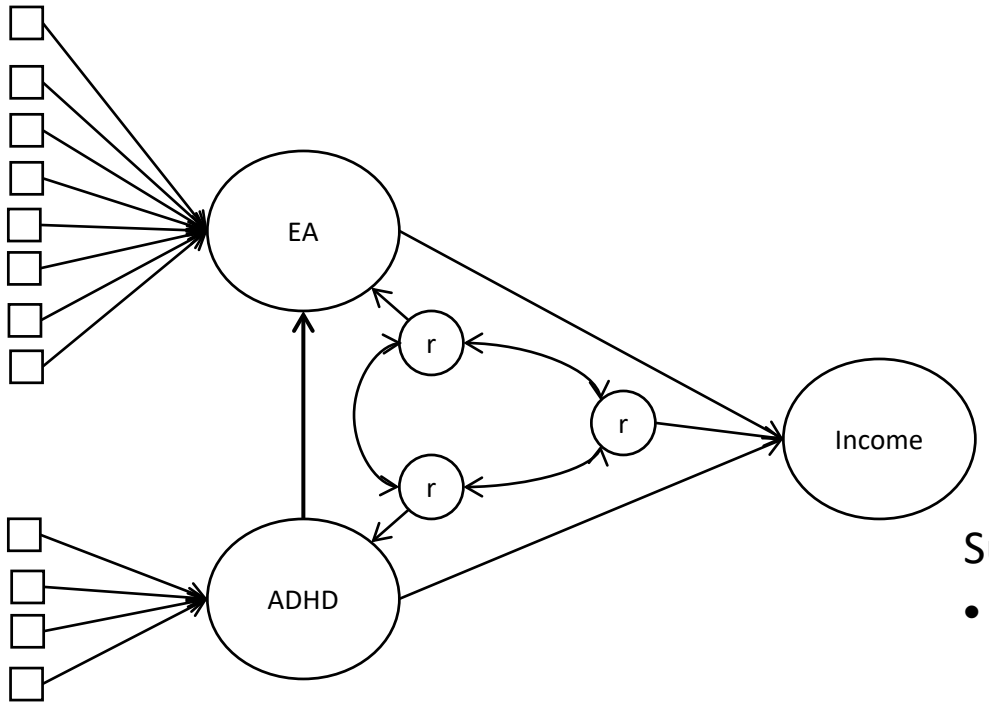
the “Exclusion Restriction”

MR in Genomic SEM

- Mendelian randomization using GWAS summary data



MR in Genomic SEM Networks

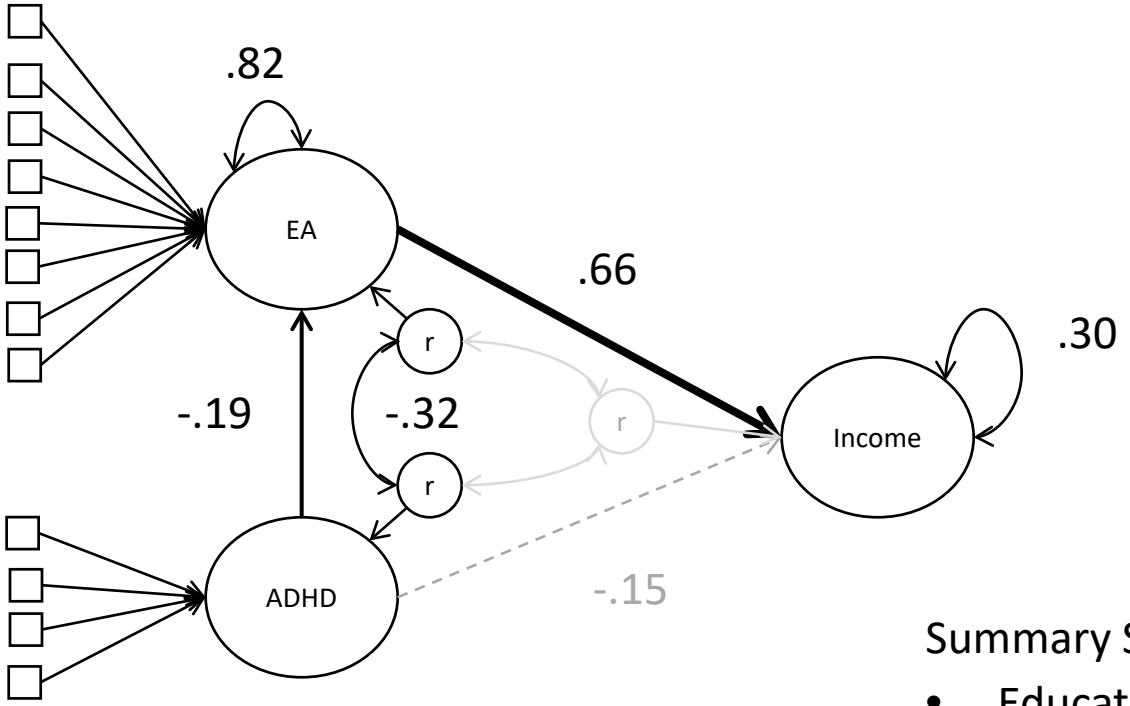


Summary Statistics:

- Educational Attainment (Okbay et al. 2016)
  - 160 hits (Sample 8 hits for this example)
- ADHD (Demontis et al., 2017)
  - 11 hits, 4 present in al
- Income (Hill et al., 2016)
  - Used as outcome in this example

See also: Burgess & Thompson (2015)

MR in Genomic SEM Networks

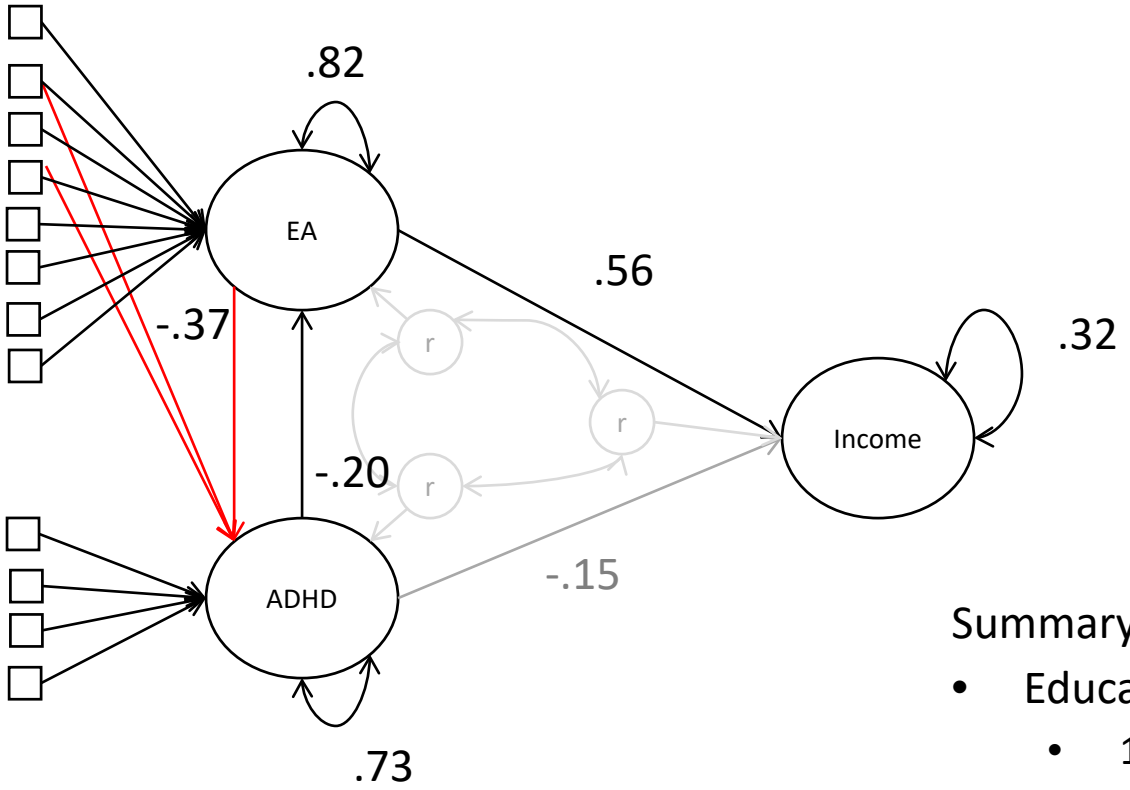


Summary Statistics:

- Educational Attainment (Okbay et al. 2016)
  - 160 hits (Sample 8 hits for this example)
- ADHD (Demontis et al., 2017)
  - 11 hits, 4 present in al
- Income (Hill et al., 2016)
  - Used as outcome in this example

MR in Genomic SEM Networks

Allow for reciprocal causation
and
pleiotropic SNPs
(that violate exclusion
restriction)

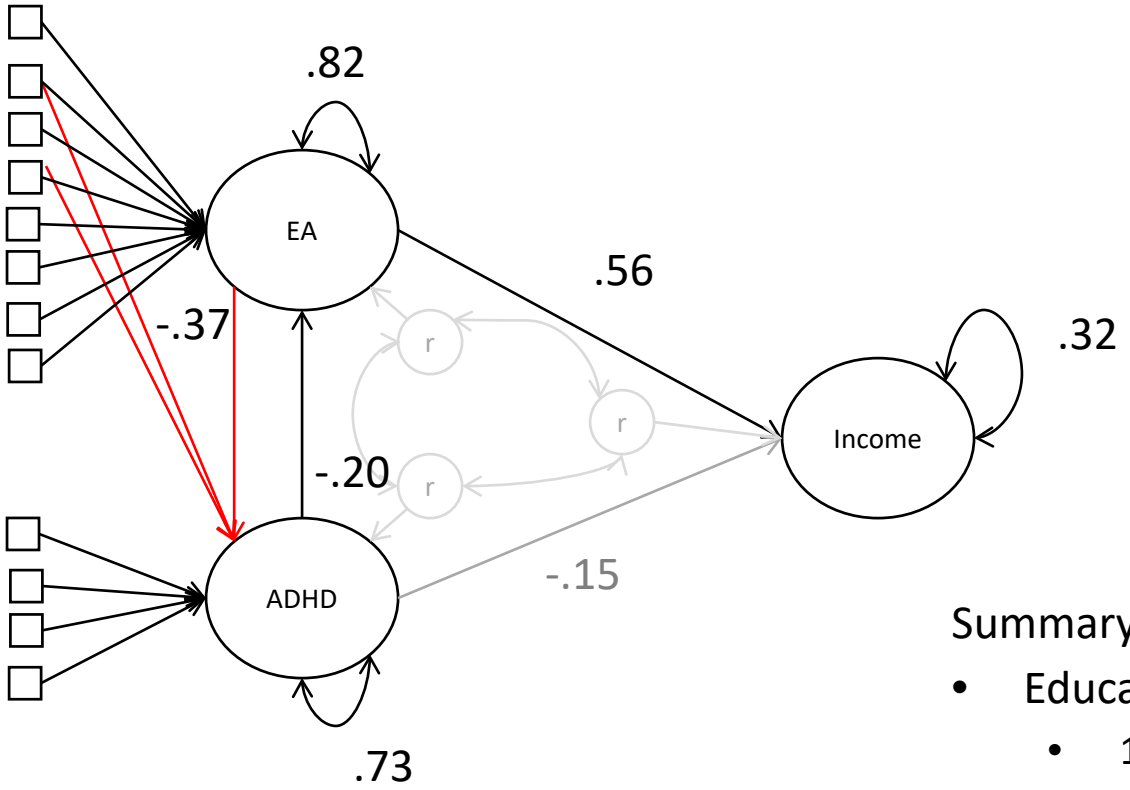


Summary Statistics:

- Educational Attainment (Okbay et al. 2016)
  - 160 hits (Sample 8 hits for this example)
- ADHD (Demontis et al., 2017)
  - 11 hits, 4 present in al
- Income (Hill et al., 2016)
  - Used as outcome in this example

MR in Genomic SEM Networks

Allow for reciprocal causation
and
pleiotropic SNPs
(that violate exclusion
restriction)



Summary Statistics:

- Educational Attainment (Okbay et al. 2016)
  - 160 hits (Sample 8 hits for this example)
- ADHD (Demontis et al., 2017)
  - 11 hits, 4 present in al
- Income (Hill et al., 2016)
  - Used as outcome in this example

Overview

- Genomic SEM is ready for use today!
  - Work through examples and tutorials on our wiki (https://github.com/MichelNivard/GenomicSEM/wiki)
  - Ask questions on our google forum
- Lots can be done using existing, openly available GWAS summary statistics
- Models are flexible and up to the user
- Modeling language is very straightforward
  - Regression: y ~ x
  - Covariance: x1 ~~ x2
- Use Genomic SEM to derive sumstats for novel phenotypes for use in PGS analyses

Acknowledgements

- NIH grants R01HD083613, R01AG054628, R21HD081437, R24HD042849
- Jacobs Foundation
- Royal Netherlands Academy of Science Professor Award PAH/6635
- ZonMw grants 531003014, 849200011
- European Union Seventh Framework Program (FP7/2007-2013) ACTION Project
- MRC grant MR/K026992/1
- AgeUK Disconnected Mind Project

extras

Stage 2 Estimation

We specify a Structural Equation Model that implies a genetic covariance matrix Σ(θ) as a function of a set of model parameters θ.

Parameters are estimated such that they minimize the discrepancy between the model implied genetic covariance matrix Σ(θ) and the S genetic covariance matrix estimated in Stage 1, weighted by the inverse of diagonal elements of the V matrix.

$$F_{WLS}(\theta) = \left(S - \Sigma(\theta)\right)' \, diag\left(V_S\right)^{-1} \left(S - \Sigma(\theta)\right)$$

"Asymptotic Distribution Free" (Brown, 1984; Muthen, 1993)

Stage 2 Estimation

Standard errors are obtained with a sandwich correction using the full V_s matrix

$$V_\theta = \left(\hat{\Delta}'\Gamma^{-1}\hat{\Delta}\right)^{-1}\hat{\Delta}'\Gamma^{-1}V_S\Gamma^{-1}\hat{\Delta}\left(\hat{\Delta}'\Gamma^{-1}\hat{\Delta}\right)^{-1}$$

where $\hat{\Delta}$ is the matrix of model derivatives evaluated at the parameter estimates, Γ is the naïve weight matrix, diag(V_s), used in paramemeter estimation, and V_s is the full sampling covariance matrix of the genetic variances and covariances.

Model Fit Statistics (model χ^2, AIC, CFI) are derived using S and V matrices, rather than the usual formulas that only apply to raw data-based estimates of covariance matrices

MTAG builds off the LDSC framework

$$\varphi_k = X\,\beta_k + \epsilon_k$$

- φ_k is an $N\times 1$ vector of scores on phenotype k
- X is an $N\times M$ matrix of standardized genotypes
- β_k is an $M\times 1$ vector of genotype effect sizes for phenotype k
- ϵ_k is an $N\times 1$ vector of residuals for phenotype k

β_k are random effects

- $E(\beta_k)= 0$ and $\text{cov}(\beta_k)= \Omega$
- Σ is the sampling covariance matrix of GWAS estimates of β_k
- In other words:

$$\Omega_{\text{MTAG}} = \frac{1}{M} S_{\text{GSEM}} \text{ and } \Sigma_{\text{MTAG}} \approx V_{\text{SNP GSEM}}$$

How Does Genomic SEM Relate to Other Multivariate Methods for GWAS Discovery?

e.g. MTAG (Turley et al., 2018)

MTAG is a Specific Model in Genomic SEM

MTAG Moment Condition

$$\mathrm{E}\left(\widehat{\boldsymbol{\beta}}_j - \frac{\boldsymbol{\omega}_t}{\boldsymbol{\omega}_{tt}}\boldsymbol{\beta}_{j,t}\right) = \mathbf{0}$$

$$\beta_{GWAS\ j,s} = \frac{cov(t,s)_{LDSC}}{var(t)_{LDSC}}\,\beta_{MTAG\ j,t}$$

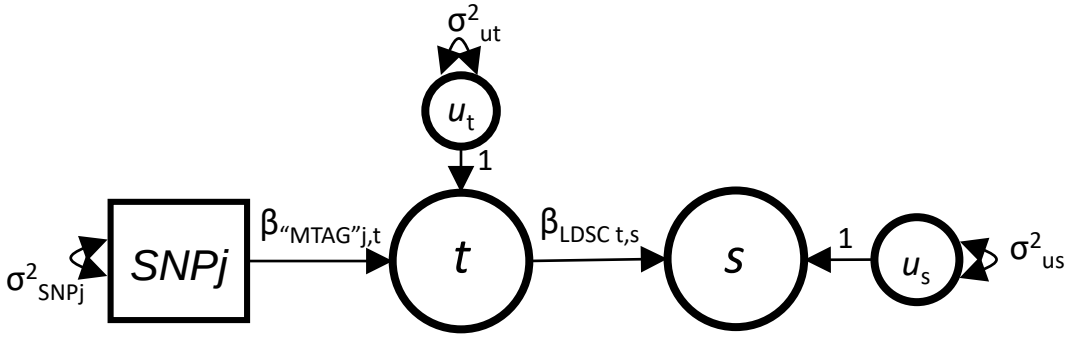
$$i.e.,\ \beta_{MTAG\ j,t} = \frac{\beta_{GWAS\ j,s}}{\beta_{LDSC\ t,s}}$$

and

$$\beta_{MTAG\ j,t} = \beta_{GWAS\ j,t}$$

$$(\ \Omega_{\mathrm{MTAG}} = \frac{1}{M}\mathrm{S}_{\mathrm{GSEM}}\ \text{ and }\ \Sigma_{\mathrm{MTAG}} \approx \mathrm{V}_{\mathrm{SNP\,GSEM}}\ )$$

MTAG as a Genomic SEM



$$\frac{\sigma_{GWAS\ j,s}}{\sigma^2_{SNPj}} = \beta_{\text{“MTAG”}j,t}\,\beta_{LDSC\ t,s}$$

$$\text{i.e.,}\ \beta_{\text{“MTAG”}j,t} = \frac{\beta_{GWAS\ j,s}}{\beta_{LDSC\ t,s}}$$

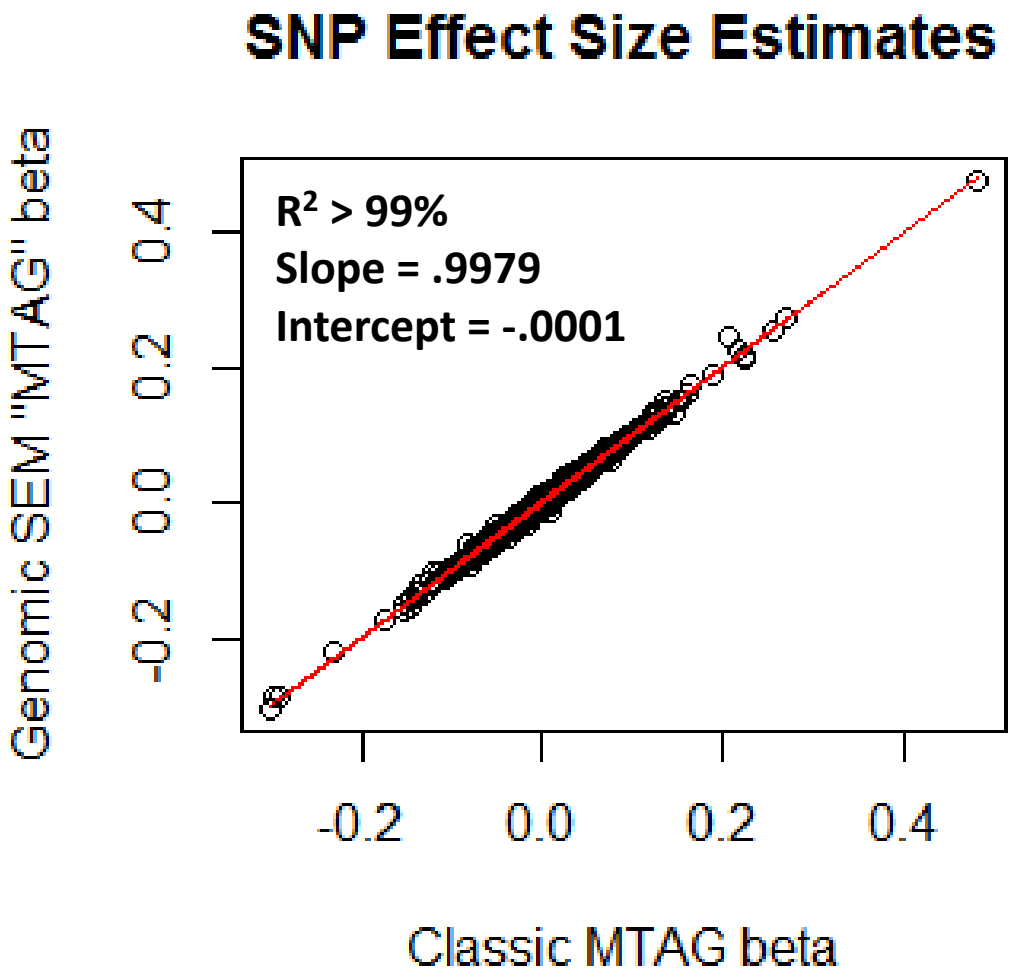
and

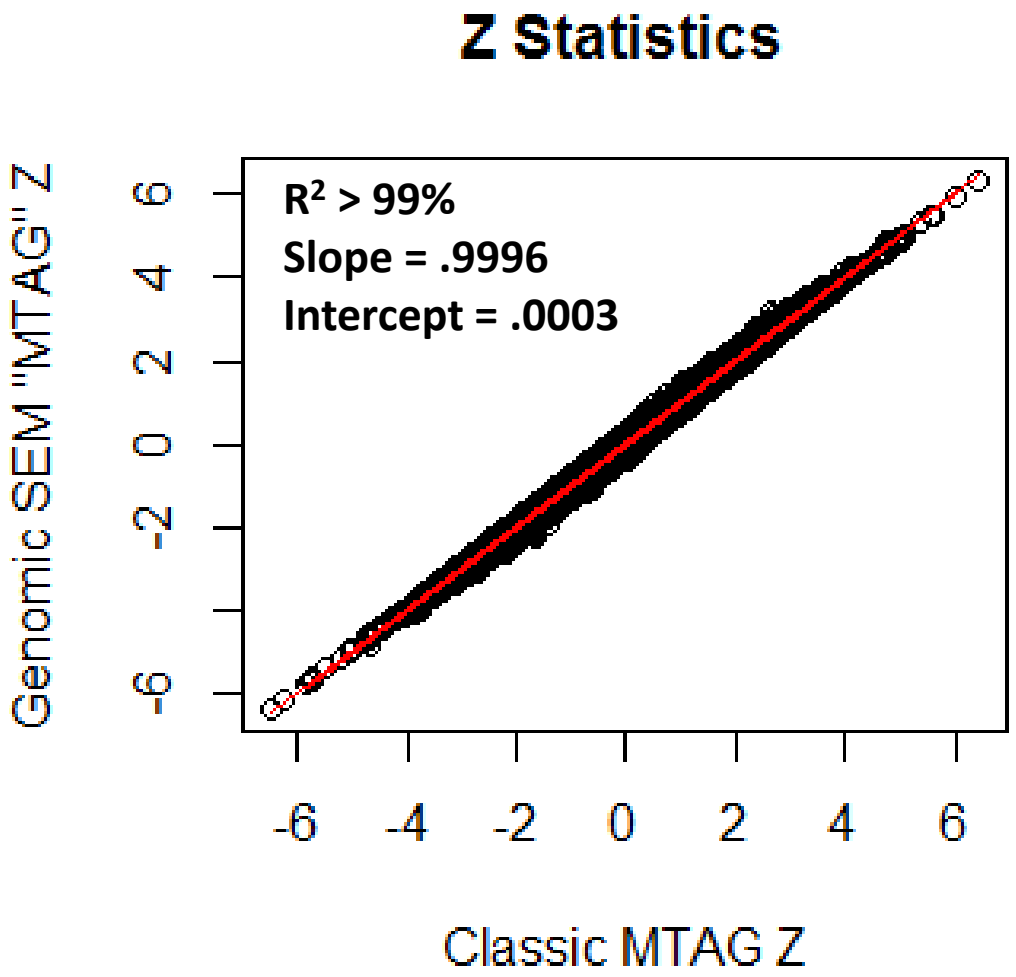
$$\sigma_{GWAS\ j,t} = \sigma^2_{SNPj} \times \beta_{\text{“MTAG”}j,t}$$

$$\text{i.e,}\ \beta_{\text{“MTAG”}j,t} = \beta_{GWAS\ j,t}$$

Classic MTAG vs. Genomic SEM "MTAG"

(Simulation Data: 2 phenotypes, 40% sample overlap)





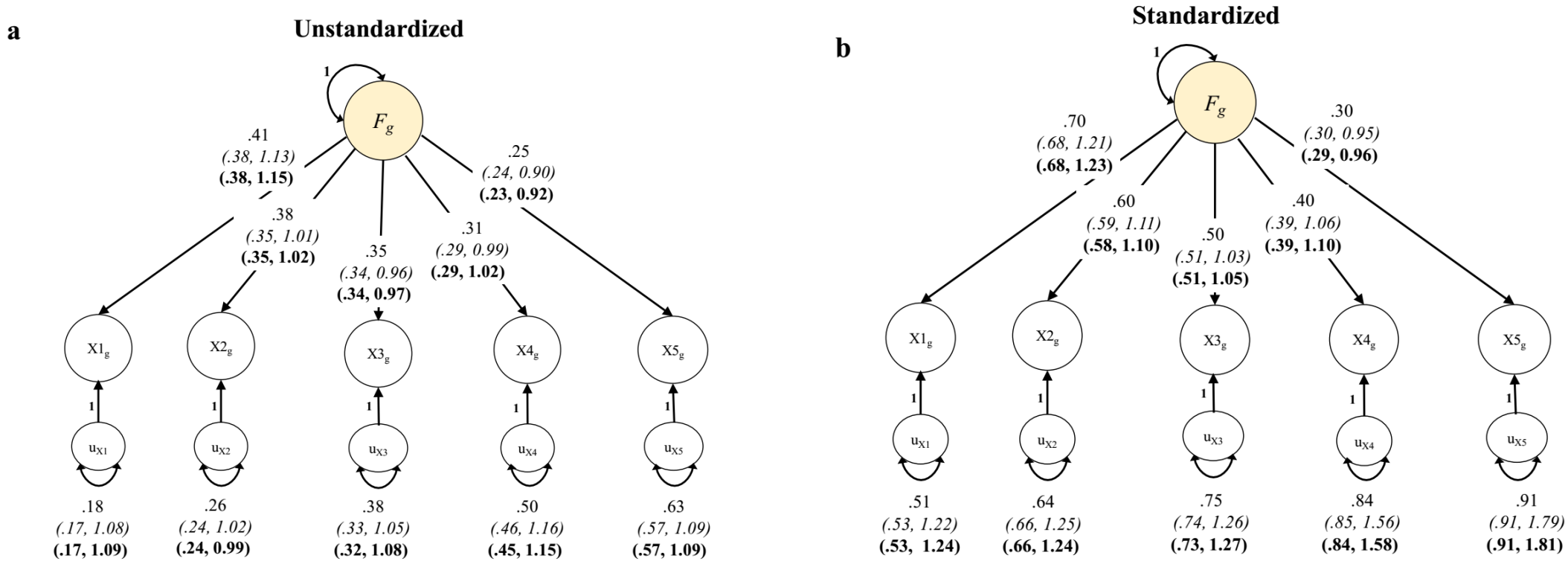


Fig. S1. Genomic SEM simulation results. Results from 100 runs of Genomic SEM using data simulated at the level of the SNPs. Results are presented for unstandardized (panel a) and standardized (panel b) estimates. Parameters outside of the parentheses indicate those provided in the generating population. In parentheses, we provide for WLS (*in italics*) and ML (**in bold**) estimation the average point estimate and the ratio of the mean *SE estimate* across the 100 runs over the empirical *SE* (calculated as the standard deviation of the parameter estimates across the 100 runs). The ratio of mean and empirical *SE*s was close to 1 in all cases, although slightly above 1 (i.e., conservative) for standardized estimates of residual variance. These *SE estimates* are expected to be upwardly biased in the standardized case due to heritability estimates being fixed to 100%.

Chi Square Statistic Null Distribution

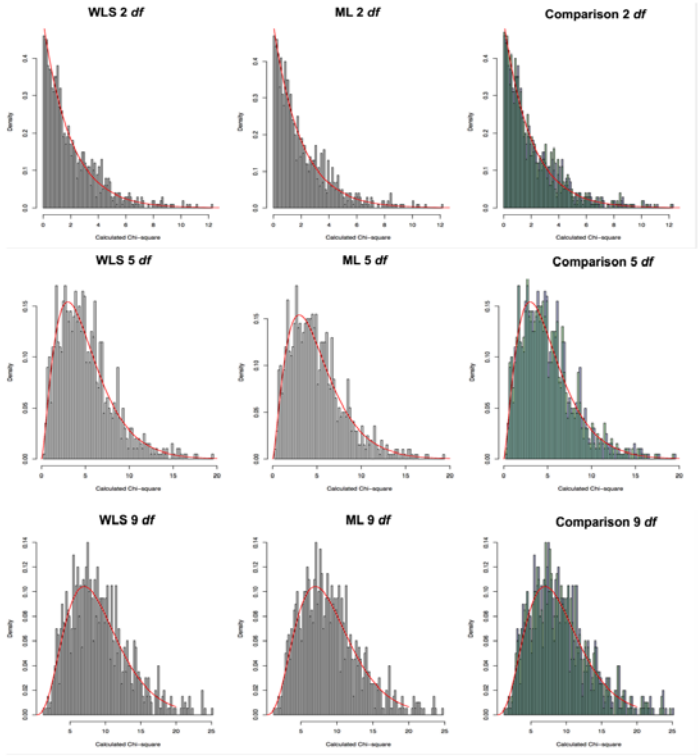


Fig. S25. Distributions of calculated and theoretical χ^2 statistics. Comparison between distribution of χ^2 values for model estimated from S and V matrices using WLS (left column) and ML (middle column) against a theoretical χ^2 distribution. The right column compares the distributions of WLS (blue bars) and ML (green bars).

Chi Square (SumStat) vs. Chi Square (Raw)

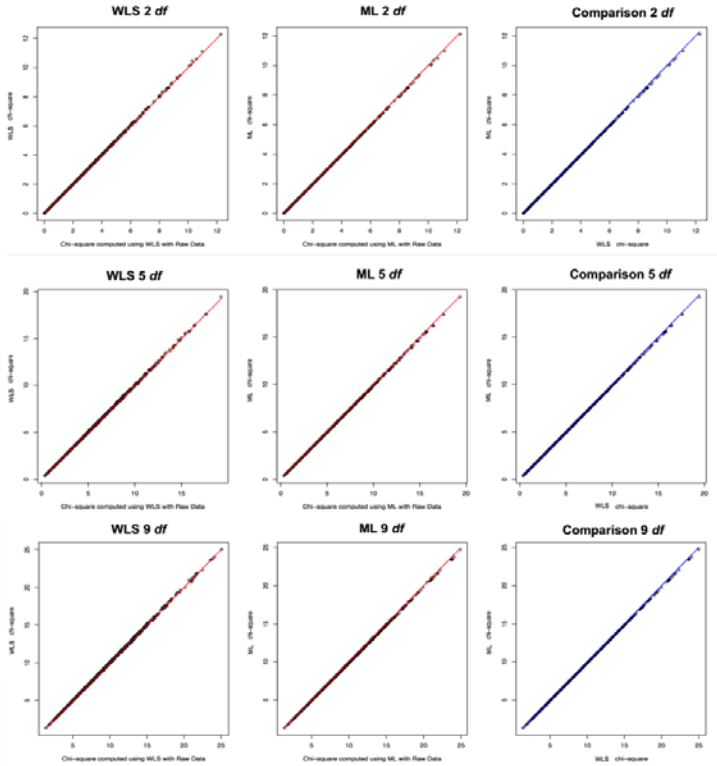


Fig. S24. Associations between model χ^2 values computed from summary data and model χ^2 values computed from raw data. Raw data-based estimates of model χ^2 were computed directly from the data using lavaan. Summary data-based estimates of model χ^2 were computed using the S and V matrices with WLS (left) and ML (right) estimation. The red line in the middle and left panel reflects the regression line for the raw data-based model χ^2 predicting itself. The blue line in the right panel reflects the regression line for the WLS χ^2 predicting itself.

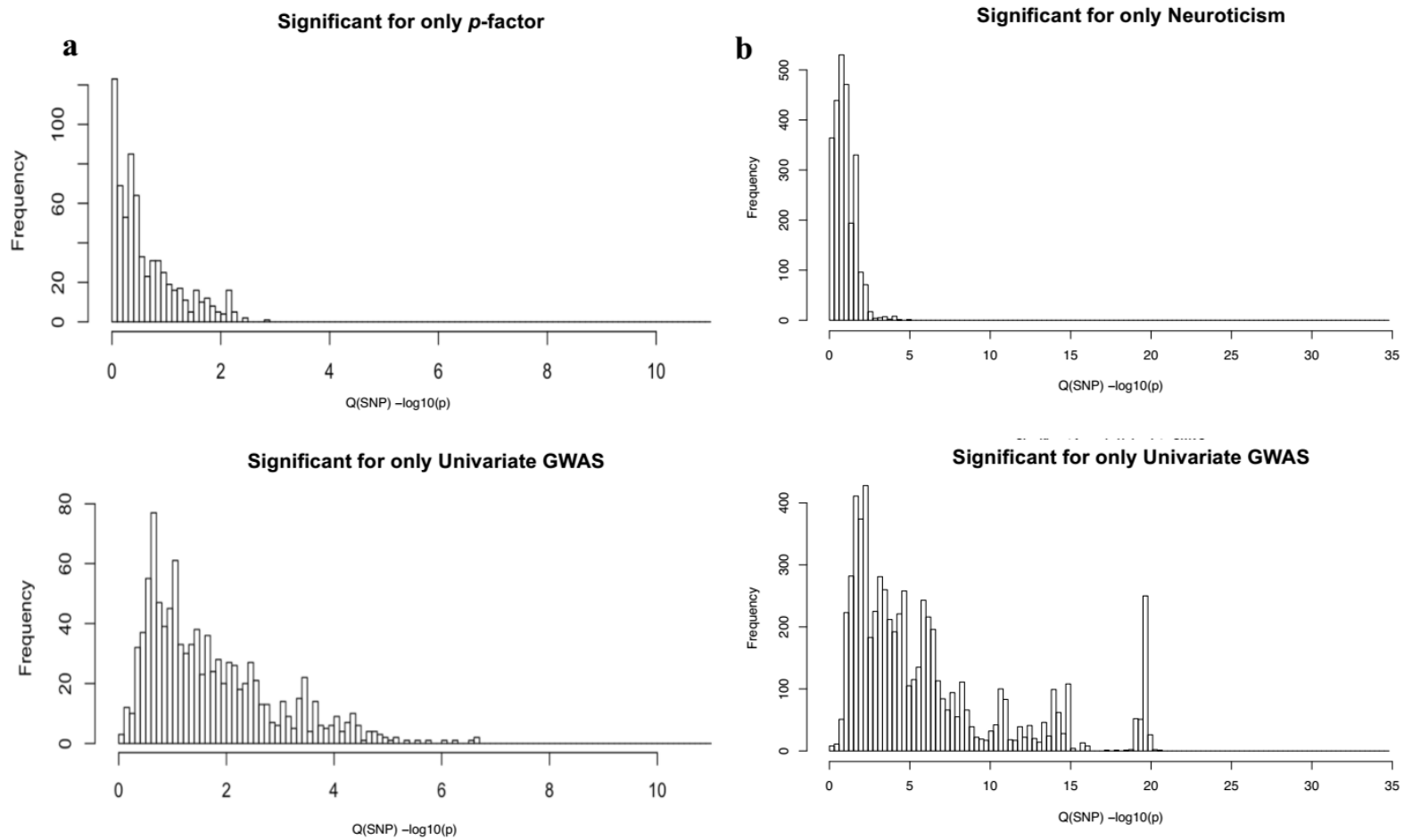


Fig. S15. Q_{SNP} –log10 *p*-values for common-factor and indicator-specific hits. Results are depicted for WLS estimation of the *p*-factor (panel a) and neuroticism (panel b). There were 684 non-independent SNPs identified as genome-wide significant for *p*-factor, but not the univariate GWAS, and 1,022 indicator-specific SNPs. For neuroticism, there were 2,540 non-independent hits specific to the common factor and 6,523 hits specific to the indicators. The average –log10 Q_{SNP} *p*-value was 0.61 for hits only on the *p*-factor and 1.81 for hits specific to the univariate indicators. For neuroticism, the average –log10 Q_{SNP} *p*-value was 0.95 for hits unique to the common factor and 5.95 for hits unique to the indicators. Thus, Q_{SNP} values were generally more significant for those SNPs not identified as significant for the common factor.

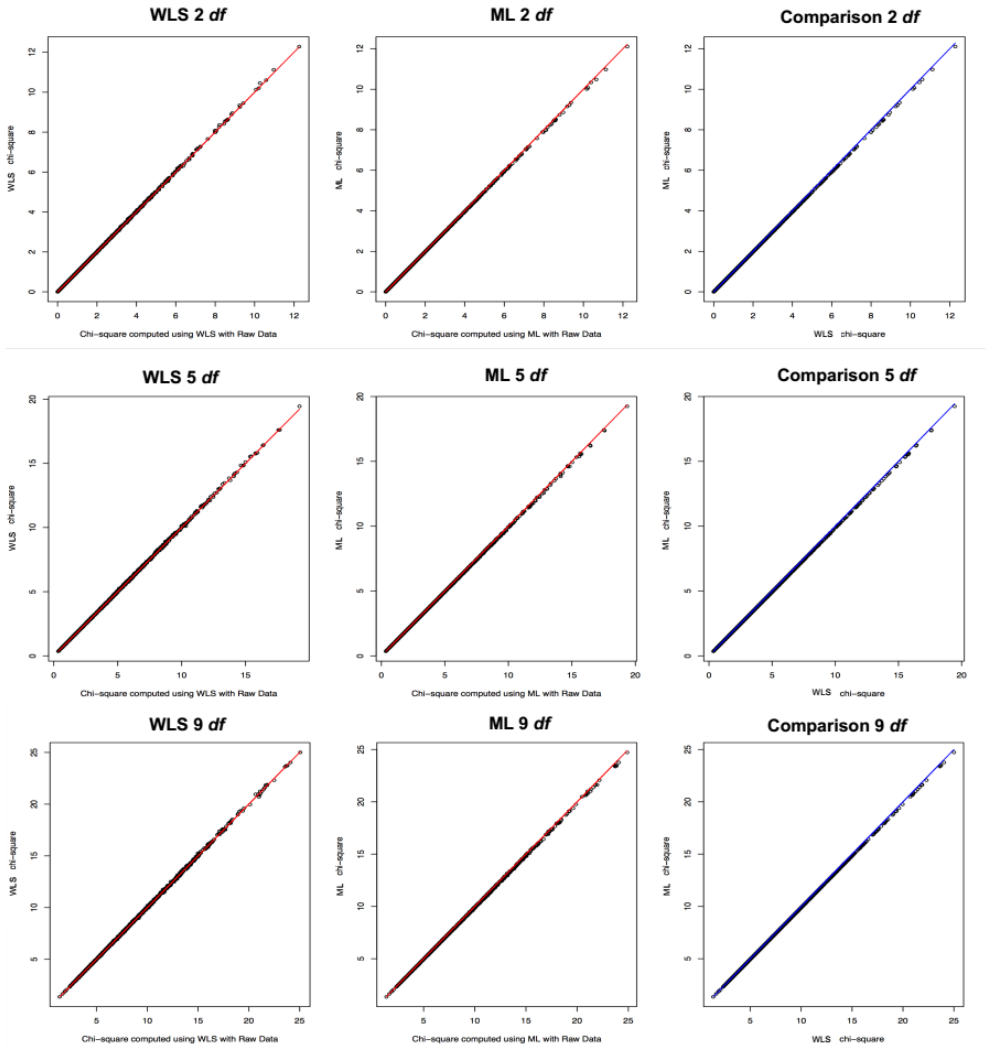


Fig. S24. Associations between model χ^2 values computed from summary data and model χ^2 values computed from raw data. Raw data-based estimates of model χ^2 were computed directly from the data using lavaan. Summary data-based estimates of model χ^2 were computed using the *S* and *V* matrices with WLS (left) and ML (right) estimation. The red line in the middle and left panel reflects the regression line for the raw data-based model χ^2 predicting itself. The blue line in the right panel reflects the regression line for the WLS χ^2 predicting itself.

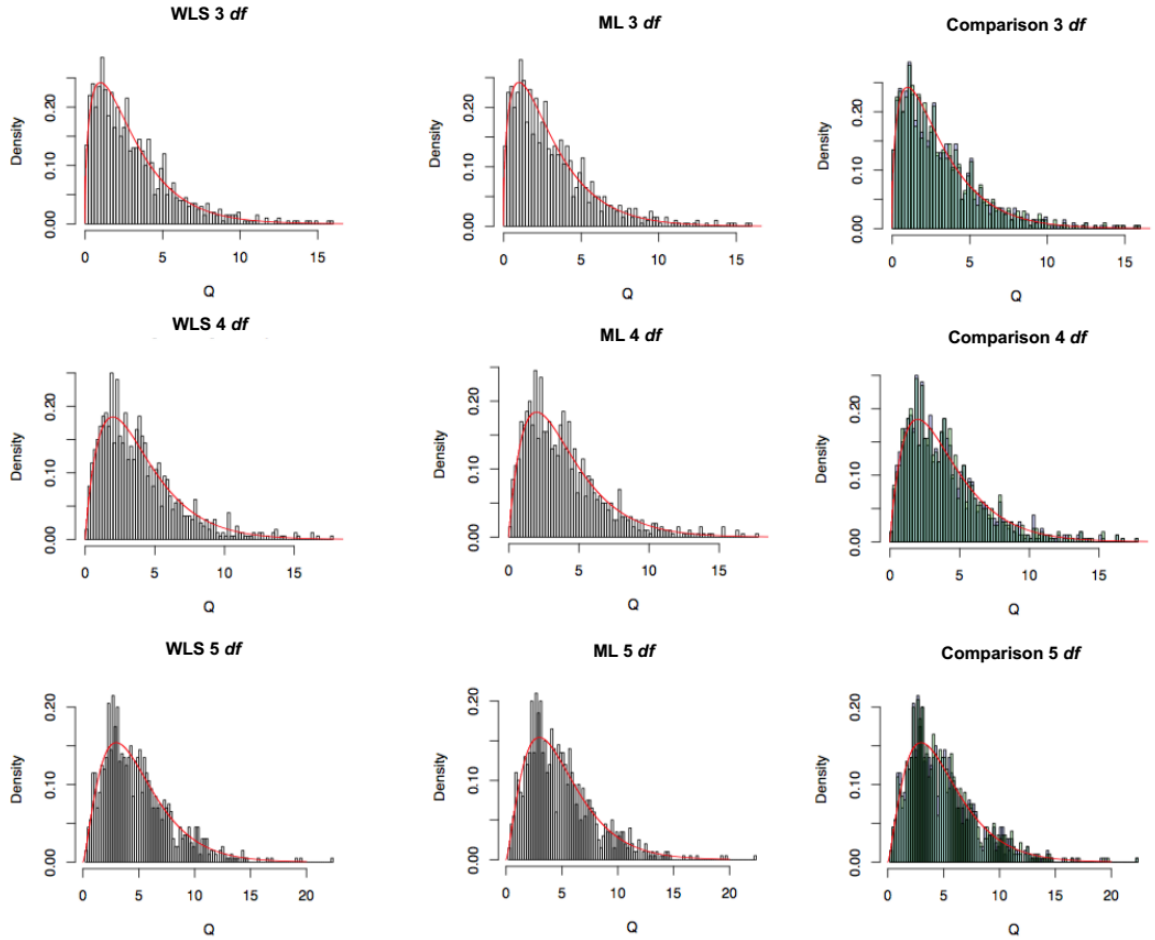


Fig. S28. Null distributions of Q_{SNP} for 1,000 simulations per model. Red lines for all panels depict the chi-square distribution with the relevant *df*. The top, middle, and bottom panels depict the sampling distributions for 3, 4, and 5 *df,* respectively. The left-most column shows estimates for WLS, the middle column estimates for ML and the right-most column overlays the WLS (depicted in light blue) and ML (light green) Q_{SNP} estimates.