Phenotypes Endophenotypes and Comorbidity

Michael C Neale
Virginia Institute for Psychiatric and Behavioral Genetics
Virginia Commonwealth University

Boulder Workshop 2011

Table of Contents

Measurement

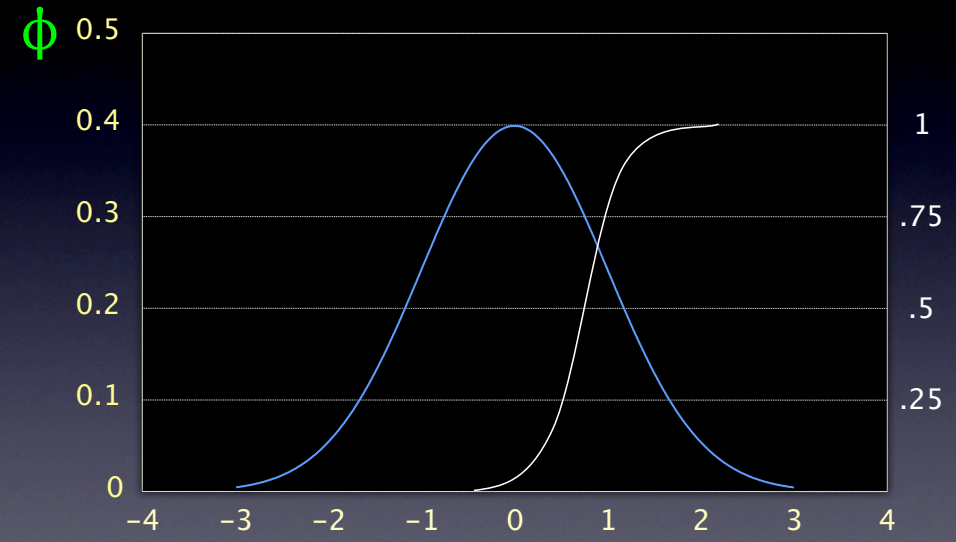
- Reduction of error variance
  - Increase external validity
    - Longitudinal (test-retest & prognostic)
    - Risk factors
    - Familial resemblance
    - Ability to find genes
- Efficiency of measurement
  - Fewer items to achieve same end
  - Equal assessment across range of (li)ability

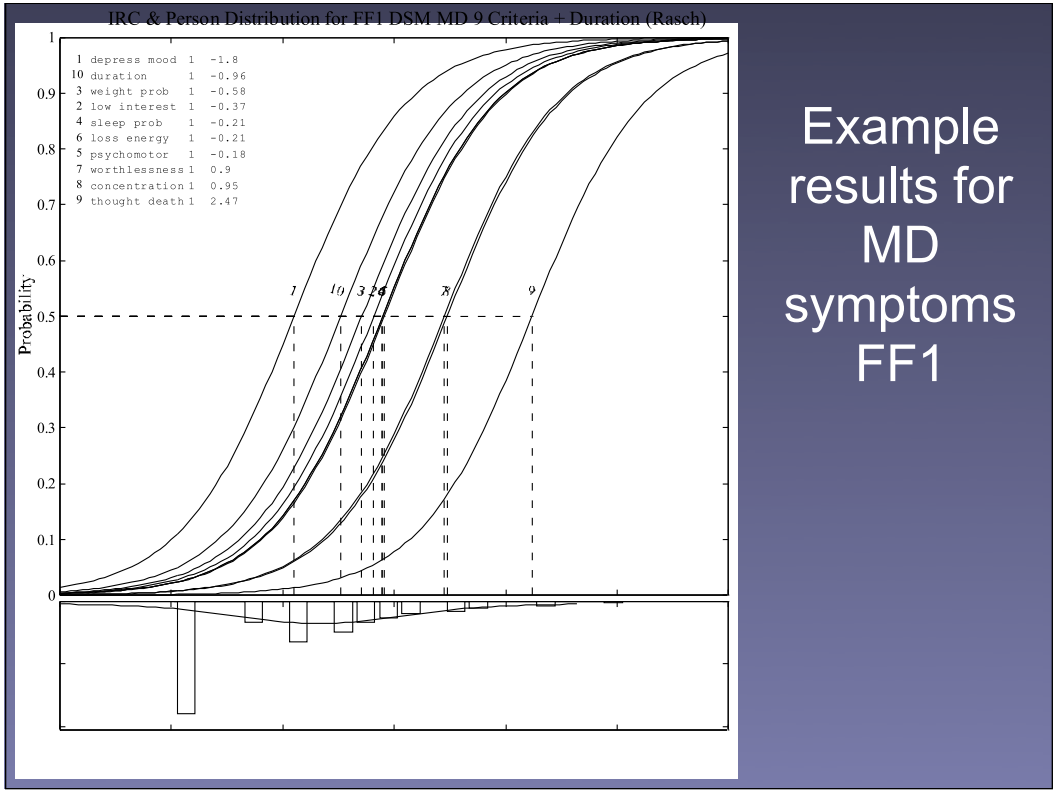
Mathematical models for measurement

- Item response theory
  - Sigmoidal function describing p(response=Y)
    - Logistic
    - Cumulative normal
  - Steep is better than flat
    - More precise delineation of where subject is in the distribution
  - Equality of slopes important
    - Order of item response probabilities is same at all places
  - Position of mid-point of slopes = 'difficulty'
    - Ideal scale should have range of difficulties

Normal liability distribution ϕ

Example item response probability shown in white





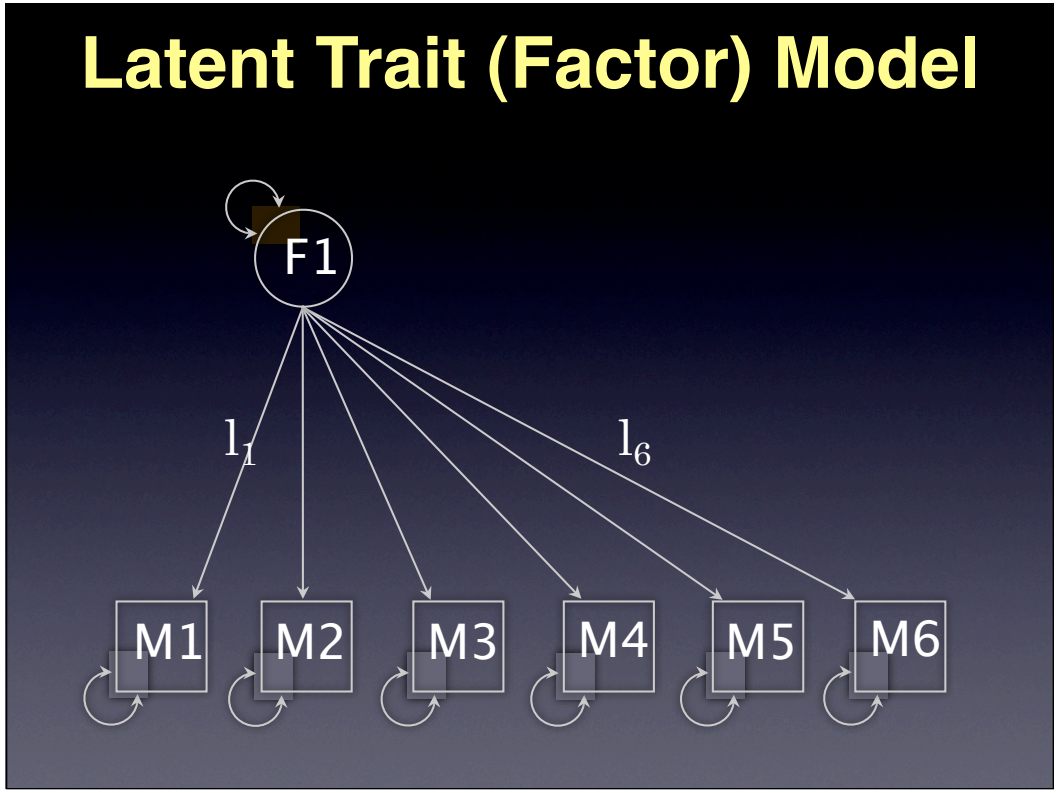
Example results for MD symptoms FF1

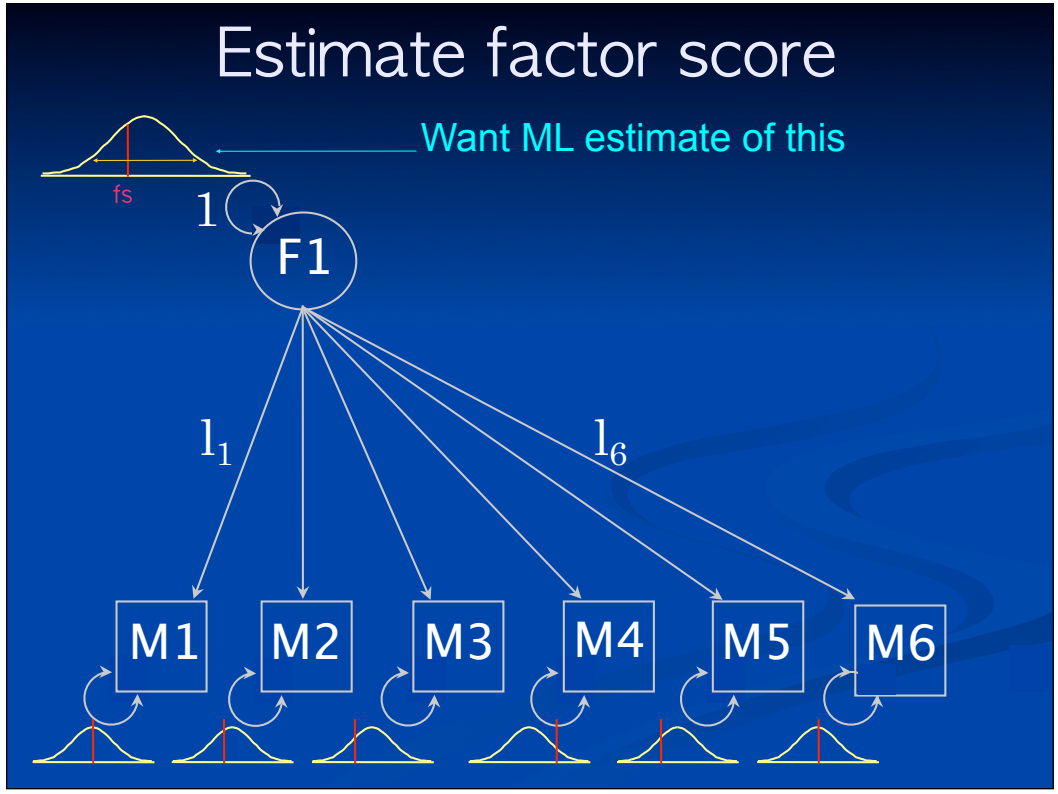
Structural Equation Model

- Two kinds of relationships
  - Linear regression X -> Y single-headed
  - Unspecified Covariance X<->Y double-headed

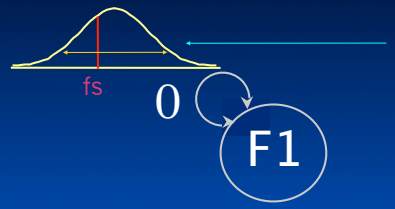
- Four kinds of variable
  - Squares – observed variables
  - Circles – latent, not observed variables
  - Triangles – constant (zero variance) for specifying means
  - Diamonds -- observed variables used as moderators (on paths)

Latent Trait (Factor) Model



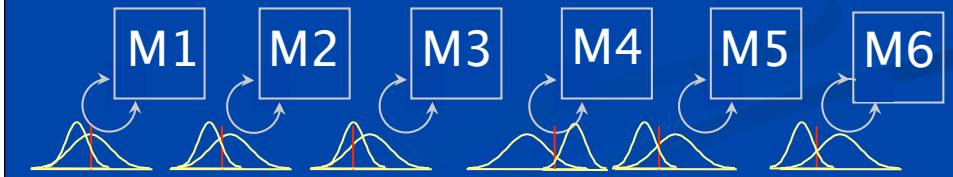

Estimate factor score
Want ML estimate of this
fs
1
F1
l_1
l_6
M1
M2
M3
M4
M5
M6

Estimate factor score

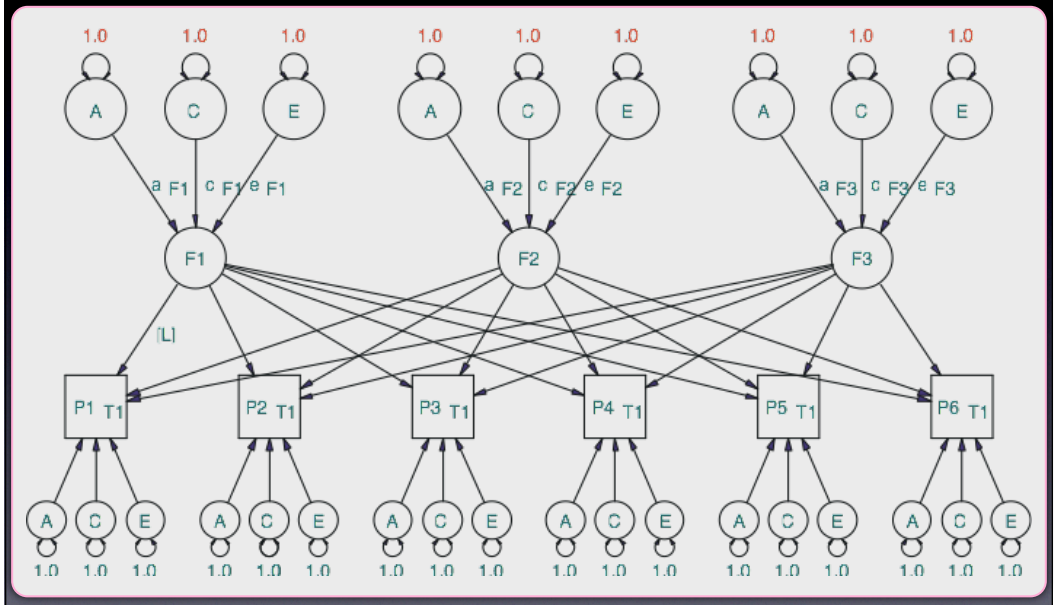


Conditional on this factor score
Factor mean = fs
Factor variance = 0

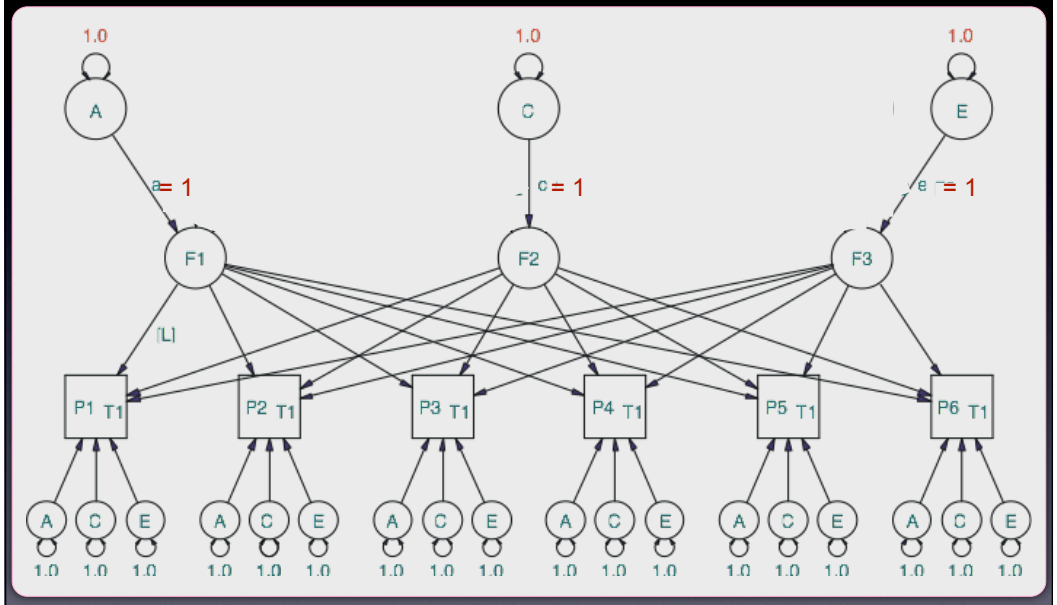
Item means move
Item variances shrink



Common Pathway Model



Independent Pathway Model



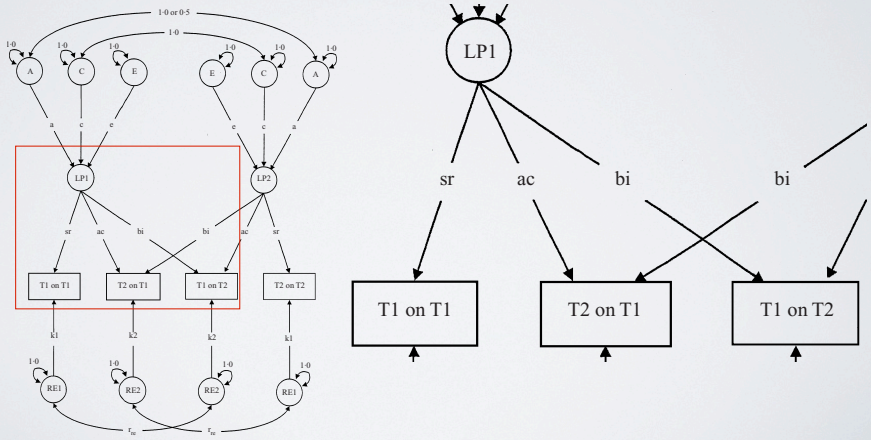
Independent pathway model is submodel of 3 factor common pathway model

Psychological Medicine, 2002, **32**, 829–842. © 2002 Cambridge University Press
DOI: 10.1017/S0033291702005858 Printed in the United Kingdom

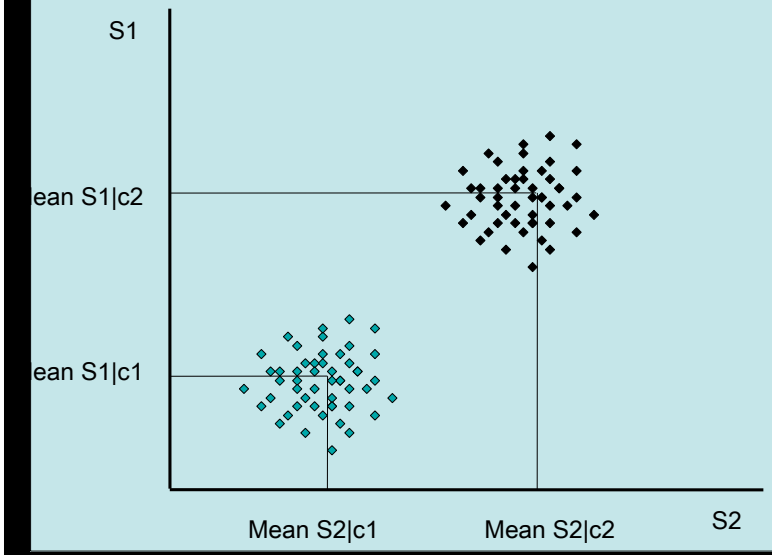
The joint analysis of personal interview and family history diagnoses: evidence for validity of diagnosis and increased heritability estimates

K. S. KENDLER,[1] C. A. PRESCOTT, K. JACOBSON, J. MYERS AND M. C. NEALE

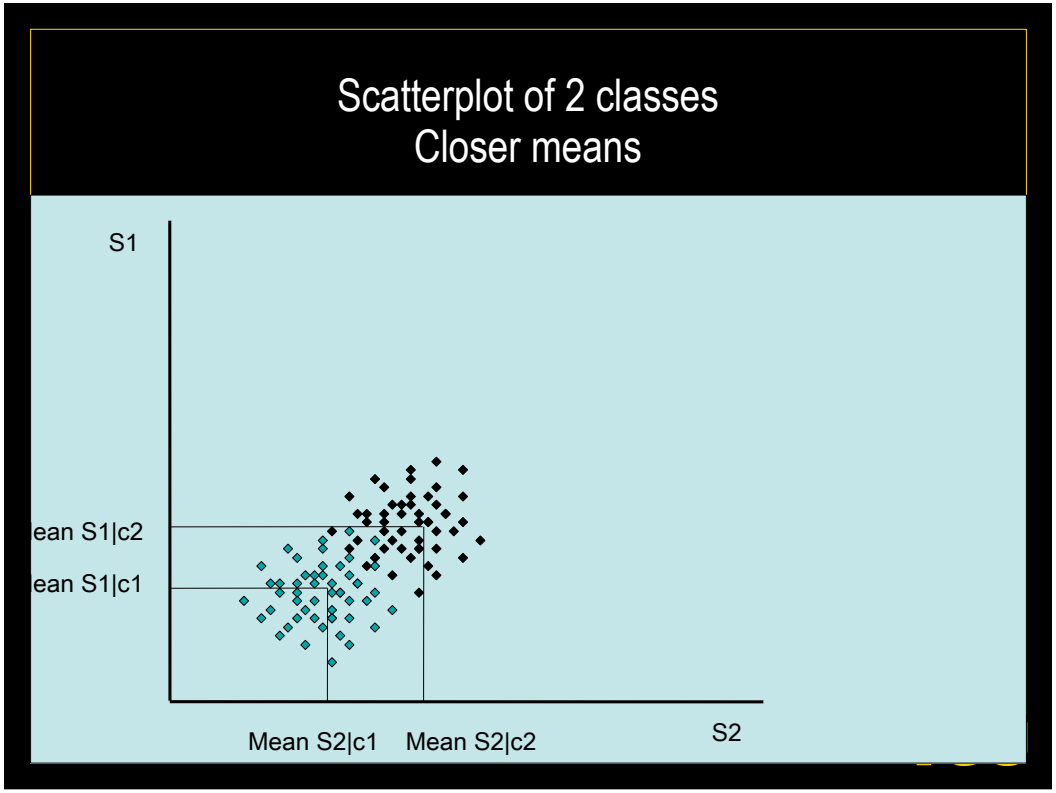
From the Virginia Institute for Psychiatric and Behavioral Genetics and Departments of Psychiatry and Human Genetics, Medical College of Virginia of Virginia Commonwealth University, Richmond, VA, USA



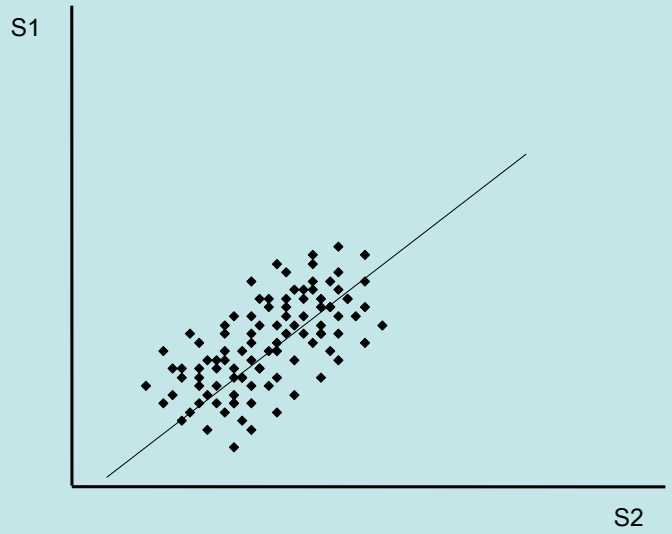
Latent Class Analysis (& FMM)



Scatterplot of 2 classes
Closer means



Scatterplot of 2 classes
Latent heterogeneity: Factors or classes?



Assessment of Psychiatric Disorders

- Psychiatrists can agree on symptoms better than on diagnoses (Kendell et al 1971)
- Diagnostic and Statistical Manual of Mental Disorders (DSM-III 1980; DSM-IIIR 1987; DSM-IV 1994; DSM-IV 2012). Widespread use
- Little empirical basis for classification
- "If you believe…"

Endophenotype Definitions

Gottesman and Gould (AJP 2003)

- Associated with illness in the population
- Heritable
- Primarily state independent (manifests in an individual whether or not illness is active)
- Co-segregates with illness within families
- Found in unaffected family members at a higher rate than in the general population.

Endophenotype Concept

- "Intermediate phenotypes that form causal links between genes and overt expression of disorders" (Cannon and Keller, 2005)

- "Intermediate trait that sits closer to the genotype in the developmental scheme" (Gottesman and Hanson, 2005)

Endophenotype Defined

- Moderate heritability
- Endophenotype and illness co-segregate within families
- Found in affected family members at higher rate than in population
- Association with illness in the population
- State-independent

Endophenotype Defined

- Association with causes rather than effects of disorders
- Endophenotype should affect a disorder
- Should have continuous variation in a population
- Should be measured across several levels of analysis

Endophenotypes Take 2

Preston & Weinberger (2005)

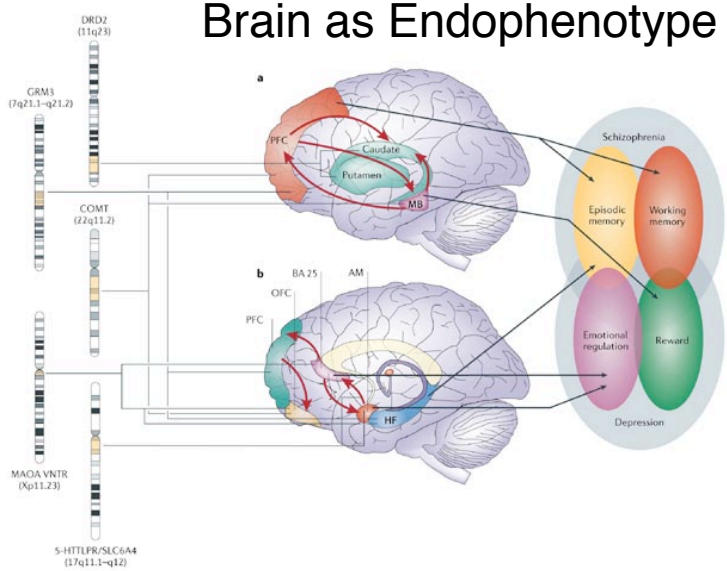
- A quantitative biological trait that is reliable and reasonably heritable, i.e., shows greater prevalence in unaffected relatives of patients than in the general population
- Should be associated with variant alleles that distinguish patients and their unaffected siblings from healthy controls on quantitative measures
- Based in part on (the) … assumption that intermediate phenotypes in schizophrenia (reflect) … a less complex genetic architecture than the disorder as a whole.

Endophenotypes Take 3

Cannon & Keller (2006)

- Endophenotypes should be:
  - Heritable
  - Associated with causes rather than effects of disorders
  - Affect a given complex disorder
  - Vary continuously in the general population
  - (optimally) measured across several levels of analysis
  - Found for genetically related disorders if they affect multiple disorders

Brain as Endophenotype



3,300,000,000 40,000 - 300,000 Copy Several

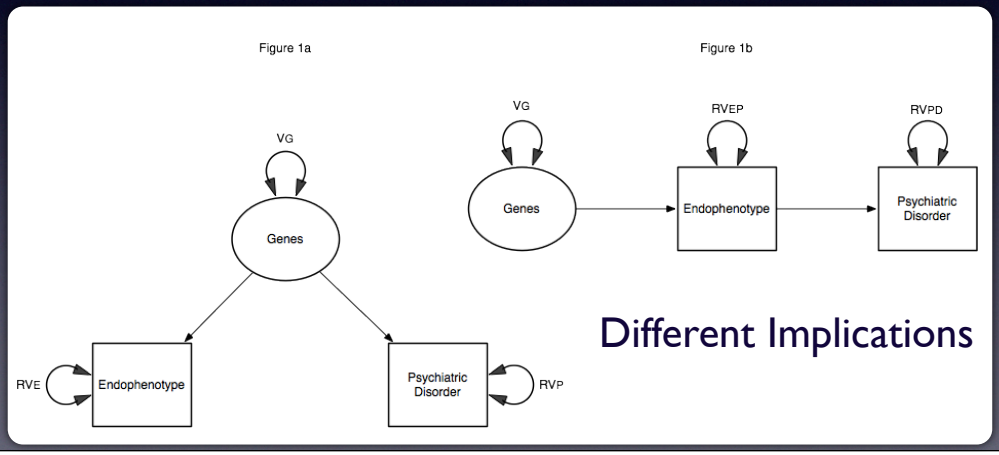
Meyer-Lindenberg and Weinberger *Nature Reviews Neuroscience* **7**, 818–827 (October 2006) | doi:10.1038/nrn1993

nature REVIEWS NEUROSCIENCE

EP Conceptual Analysis

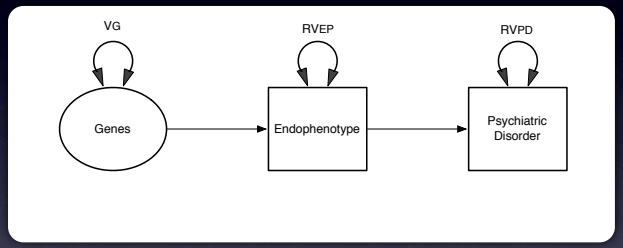
Kendler & Neale (Mol Psych 2010)

- Walters & Owen (2007): Distinction between 'Liability Index' and 'Mediational' models not explicit



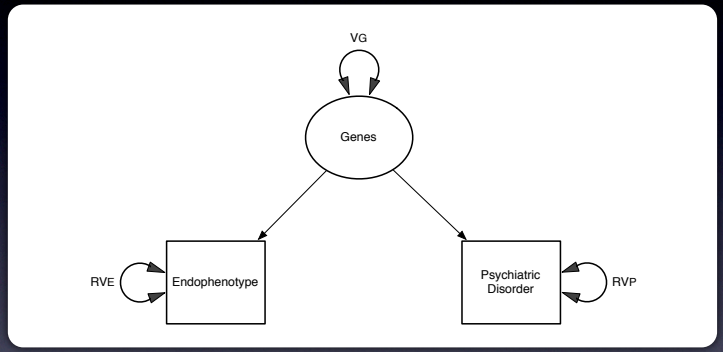
Different Implications

Traditional Endophenotype Model



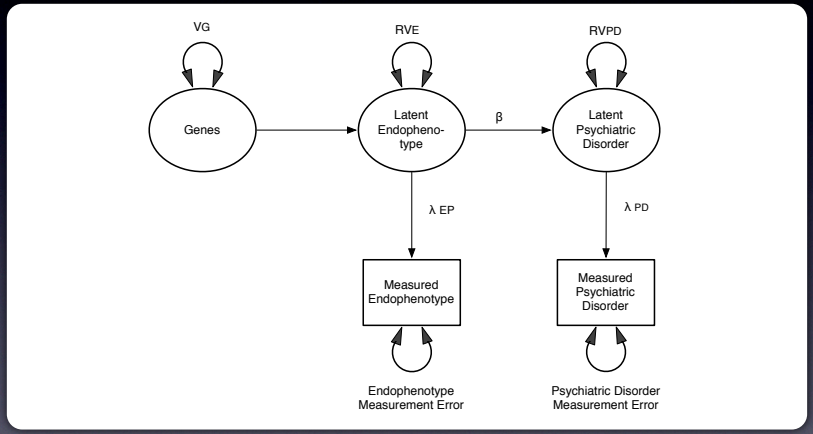
- Endophenotype is intermediate
- Ideally more heritable than PD

Common Factor Model



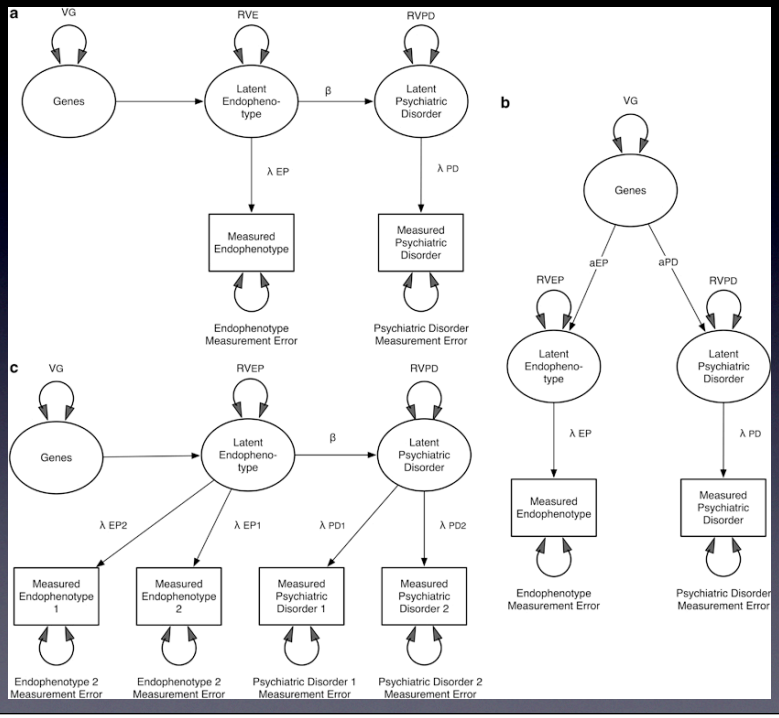
- Endophenotype is intermediate??
- Ideally more heritable than PD?

Endophenotype Error Model



- Error variances λ can change relative h^2
- Does it matter if more heritable than PD?

Include Measurement Error

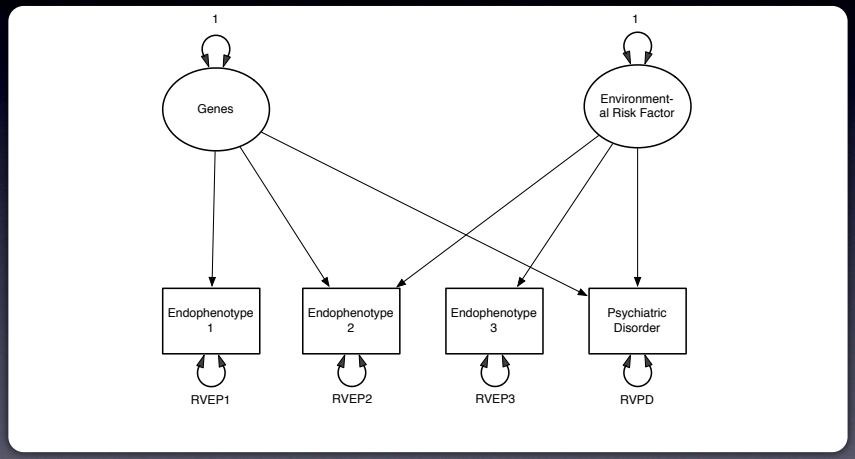


a) If $\beta < 1.0$
$h^2_{EP} < h^2_{PD}$

b) If $a_{EP}\lambda_{EP} < a_{PD}\lambda_{PD}$
$h^2_{EP} < h^2_{PD}$

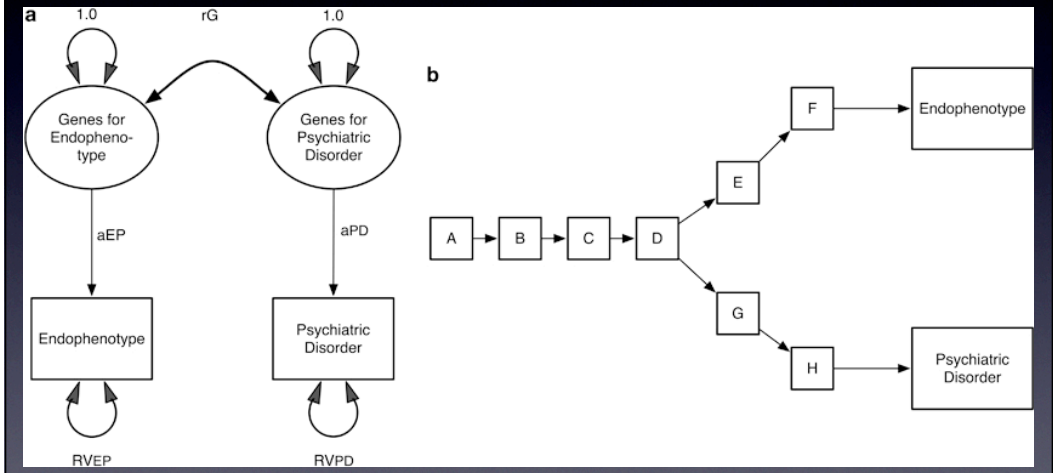
c) Useful if retest or twin data on EP & PD

Environmental Endophenotypes



- Establishing an endophenotype is a job for comorbidity modeling

Bivariate Model



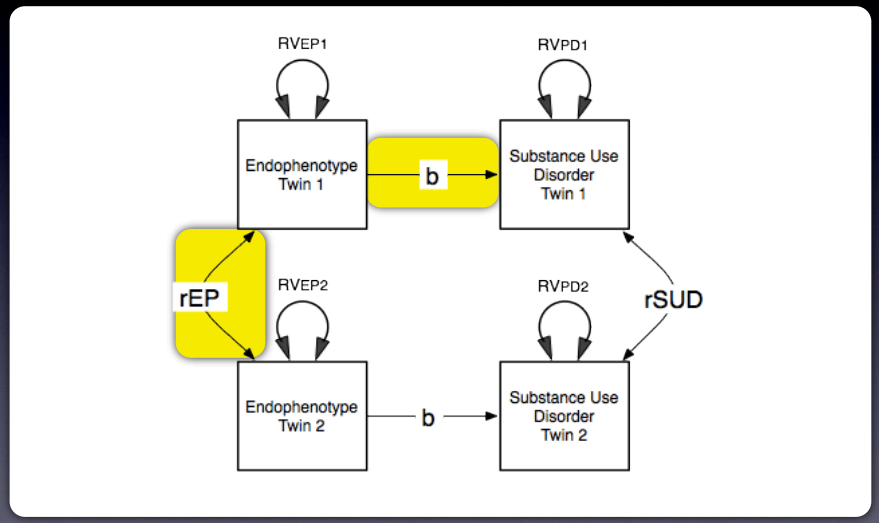
Gottesman Response

- Thank you for your in depth exegesis of EP. The nomological network you construct for the construct is in the finest tradition of Meehlian "straight thinking".
- A few scholars and deep thinkers (rare mutations in our field) may actually take the time to digest your ideas and make our field better.

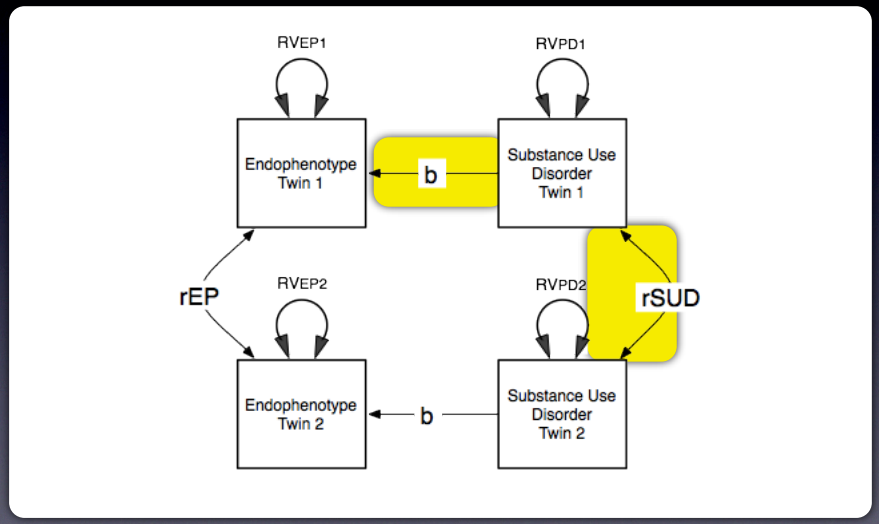
Direction of Causation

- Longitudinal Data
  - Reasonable assumption that Time 2 does not affect Time 1
  - Includes experimental designs
- Twin/Family Data
  - Contemporaneous assessment

EP causes SUD



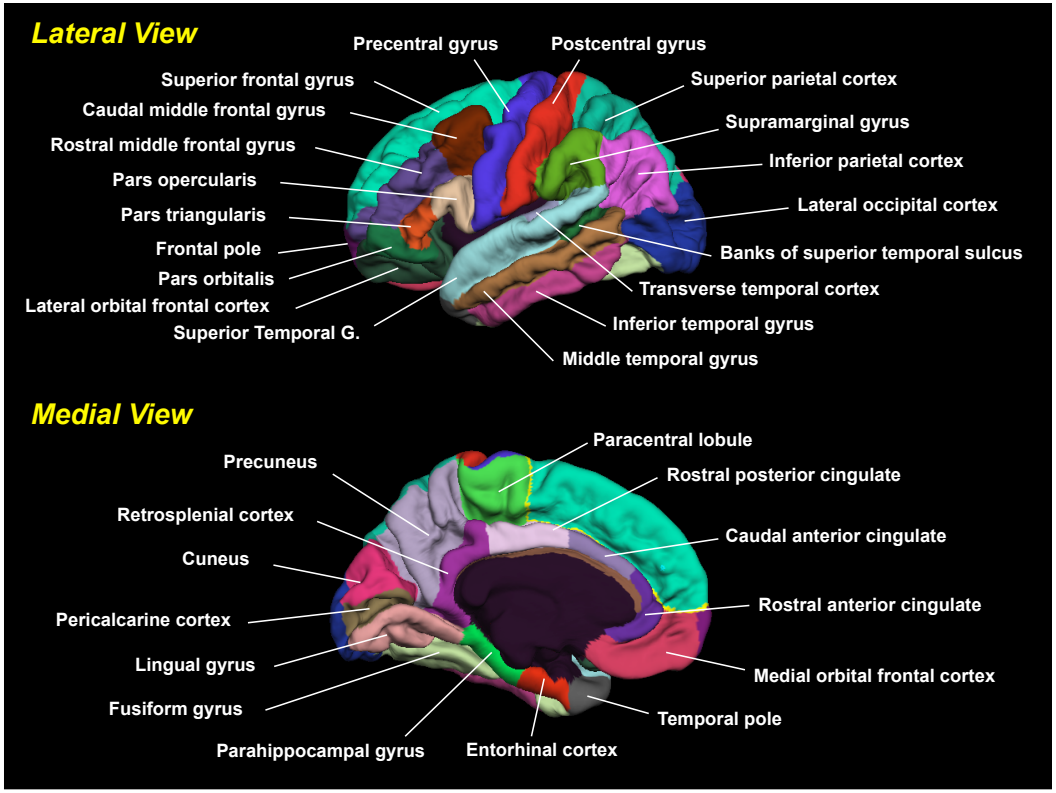
SUD causes EP



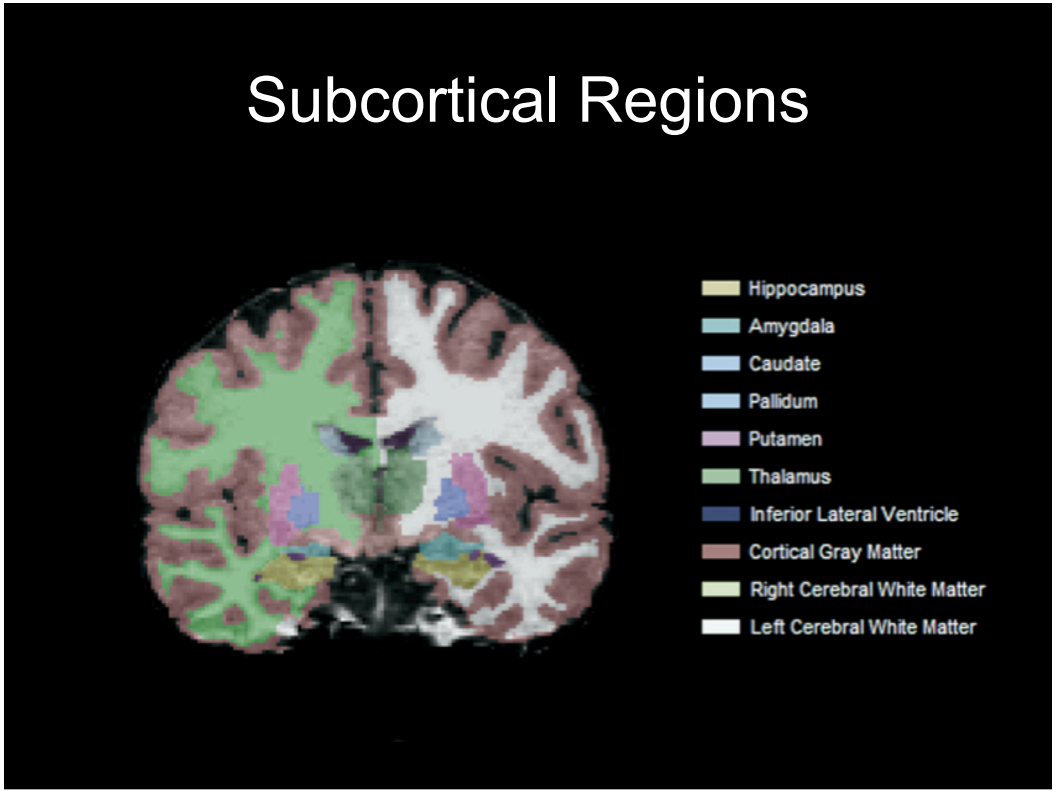
Makes different prediction about cov(EP1, SUAD2)

Rationale

- Brain: candidate endophenotype
  - Psychopathology
  - Substance abuse
  - Obesity
  - Personality
  - Cognition

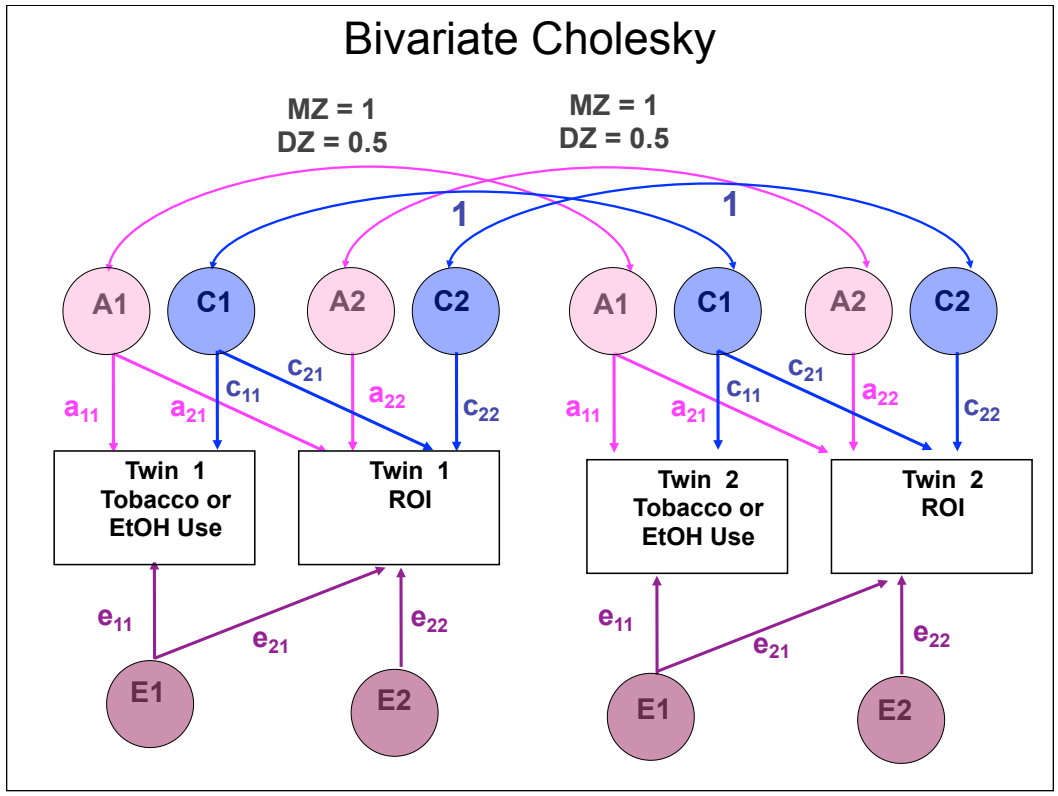

Lateral View
Precentral gyrus
Postcentral gyrus
Superior frontal gyrus
Caudal middle frontal gyrus
Rostral middle frontal gyrus
Pars opercularis
Pars triangularis
Frontal pole
Pars orbitalis
Lateral orbital frontal cortex
Superior Temporal G.
Superior parietal cortex
Supramarginal gyrus
Inferior parietal cortex
Lateral occipital cortex
Banks of superior temporal sulcus
Transverse temporal cortex
Inferior temporal gyrus
Middle temporal gyrus
Medial View
Paracentral lobule
Precuneus
Rostral posterior cingulate
Retrosplenial cortex
Caudal anterior cingulate
Cuneus
Rostral anterior cingulate
Pericalcarine cortex
Lingual gyrus
Medial orbital frontal cortex
Fusiform gyrus
Temporal pole
Parahippocampal gyrus
Entorhinal cortex

Subcortical Regions



(Possible) Functional Significance of Morphometric Measures

- Volume (cortical and subcortical)
  - Overall size, implied neuronal connectivity and function
  - Associations between poor outcomes and decreased volume or increased ventricles
- Cortical structure
  - Neurons are organized into columns perpendicular to brain surface
- Radial Unit Hypothesis of Cortical Development (Rakic)
  - Cells within a column share a common origin and migrate to their location within the cortex during development
  - Cortical surface area driven by the number of columns
  - Cortical thickness is influenced by the number of cells within a column
  - Different genetic architecture for cortical thickness and surface area


Bivariate Cholesky
MZ = 1
DZ = 0.5
MZ = 1
DZ = 0.5
1
1
A1
C1
A2
C2
A1
C1
A2
C2
a11
a21
c11
c21
a22
c22
a11
a21
c11
c21
a22
c22
Twin 1
Tobacco or
EtOH Use
Twin 1
ROI
Twin 2
Tobacco or
EtOH Use
Twin 2
ROI
e11
e21
e22
e11
e21
e22
E1
E2
E1
E2

Significant Bivariate Results
Nicotine Abuse/Dependence

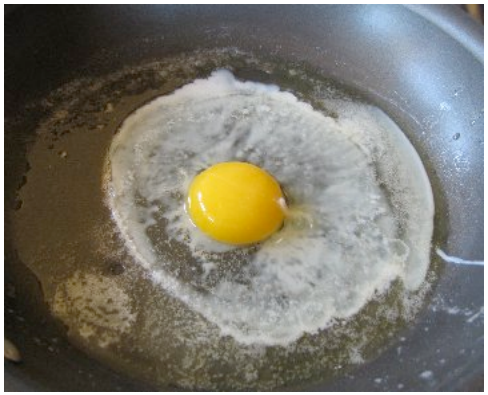
ROI	rA	rC	rE	covA	covC	covE	rP
LCerebellumWM	-0.271	-0.397	-0.054	**-0.208**	0.000	-0.035	**-0.197**
RCerebellumWM	-0.245	0.107	0.027	**-0.188**	0.000	0.018	-0.159
RCTXbankssts1	1.000	-1.000	-0.238	0.727	0.277	-0.150	**-0.029**
RCTXprecuneus1	-0.127	0.747	0.147	**-0.098**	0.000	0.094	**-0.026**
RCTXSUPtemporal1	-0.033	0.474	0.315	-0.026	0.000	0.202	**0.086**
RCTXtransversetemporal1	0.424	-0.999	-0.262	**0.326**	0.000	-0.168	**0.059**
RCTXbankssts2	1.000	-1.000	-0.242	0.759	0.162	-0.152	**-0.075**
LCTXLAToccipital2	-0.740	1.000	0.265	-0.554	0.081	0.174	**-0.016**
RCTXrostralMIDfrontal2	0.594	1.000	-0.369	0.449	0.013	-0.242	**-0.040**
RCTXSUPtemporal2	0.056	0.826	0.263	0.043	0.000	0.168	**0.169**
RCTXtransversetemporal2	0.699	-0.621	-0.220	**0.541**	0.000	-0.139	**0.106**
RCTXTotalSurfaceArea2	0.528	-0.228	-0.256	**0.397**	0.000	-0.169	**0.103**
LCTXcaudalANTCING3	1.000	1.000	-0.252	0.753	0.134	-0.162	**0.029**
RCTXcorpuscallosum3	-0.636	1.000	0.273	-0.477	0.159	0.176	**0.011**
LCTXINFparietal3	-0.128	1.000	0.354	-0.100	0.029	0.222	**0.121**
LCTXpostcentral3	0.390	-0.842	-0.336	**0.297**	0.000	-0.218	**-0.009**
RCTXrostralANTCING3	-1.000	1.000	0.254	-0.744	0.202	0.162	**0.100**

Is Brain Morphometry an Appropriate Endophenotype of Alcohol/Tobacco Use?

Satisfied?	Requirement
Y	Moderate heritability
	Endophenotype and illness co-segregate within families
	Found in affected family members at higher rate than in population
Y (kinda)	Association with illness in the population
	State-independent
	Association with causes rather than effects of disorders
	Endophenotype should affect a disorder
Y	Should have continuous variation in a population
Y	Should be measured across several levels of analysis

This is your brain…

This is your brain on drugs…



This is your brain…

This is your brain on drugs…

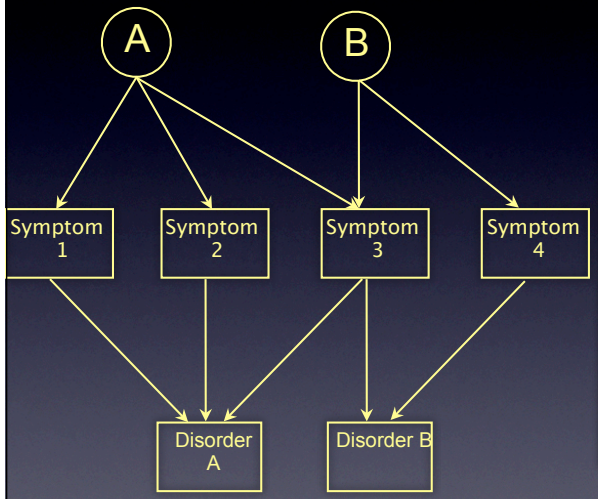
Modeling Comorbidity

- Psychiatric Disorders: binary phenotypes
  - –Lots of comorbidity
  - –Substance abuse similar
- ACE model is but one of many
- Two twins, two binary variables
  - –16 outcome combinations
- Fit models by maximum likelihood
  - –(alternatives exist)

Comorbidity is High

- High for Psychiatric Disorders
  - Anxiety
  - Depression
  - Phobias
  - Panic
  - Alcohol Abuse
- 70% of those with history of one dx have history of at least one other (Kessler 1993; N=18,000)
- Similar rates in 10,000+ Virginia twins

Comorbidity due to symptom sharing



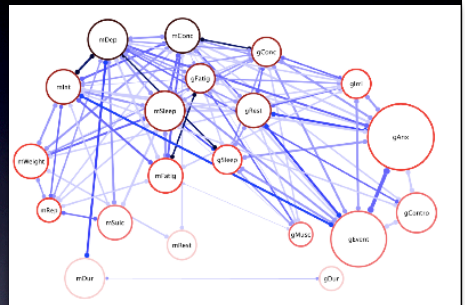
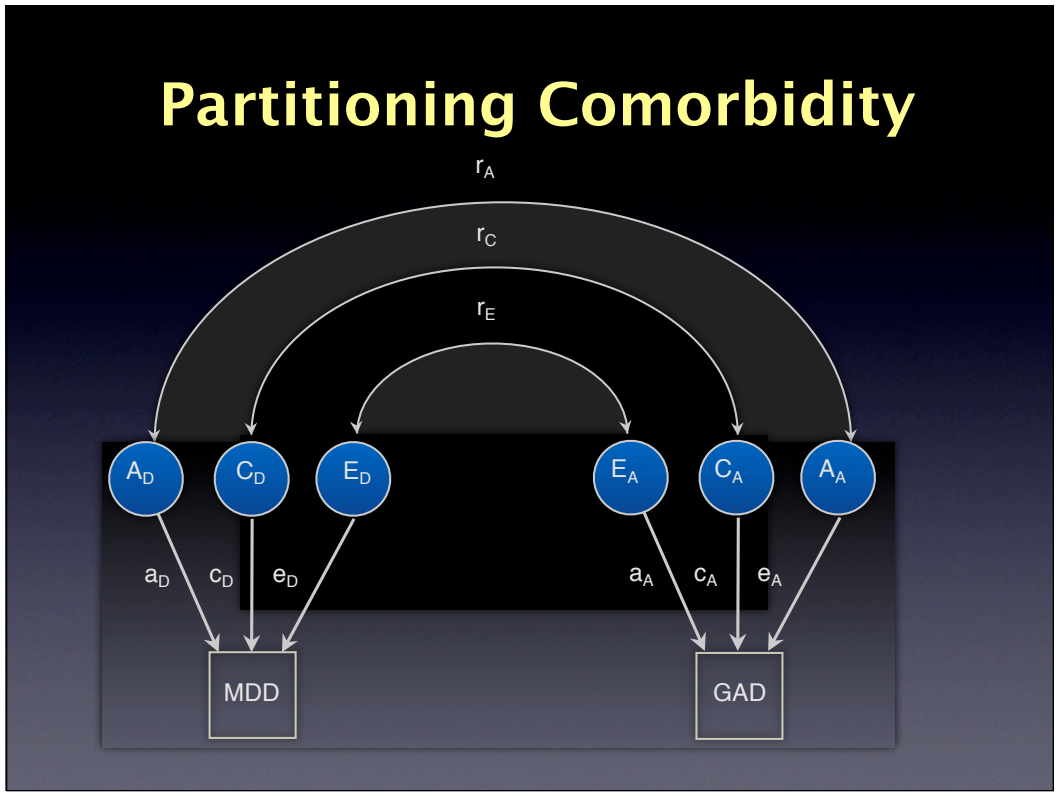


Figure 4. A comorbidity network for MDD and GAD. Larger nodes represent more frequent symptoms; darker circumference, higher centrality; thicker edges, higher frequency of co-occurrence; darker edges, stronger associations. Only edges with log odds ratio higher than (-) 0.60 are represented. Centrally positioned nodes (*mConc*, *gConc*, *mSleep*, *gSleep*, *mFatig*, *gFatig*, *mRest* and *gRest*) represent overlapping symptoms. Non-overlapping MDD symptoms are displayed on the left of the figure, non-overlapping GAD symptoms on the right.

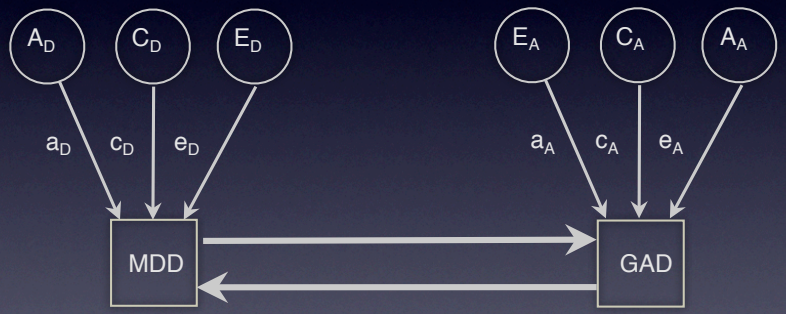
Not today!

Partitioning Comorbidity



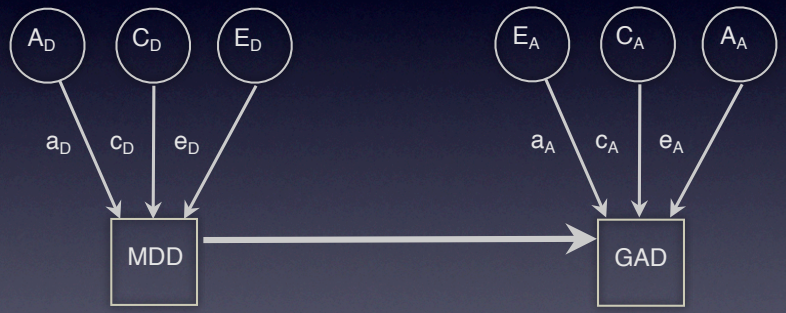
Modeling Comorbidity

Reciprocal Causation



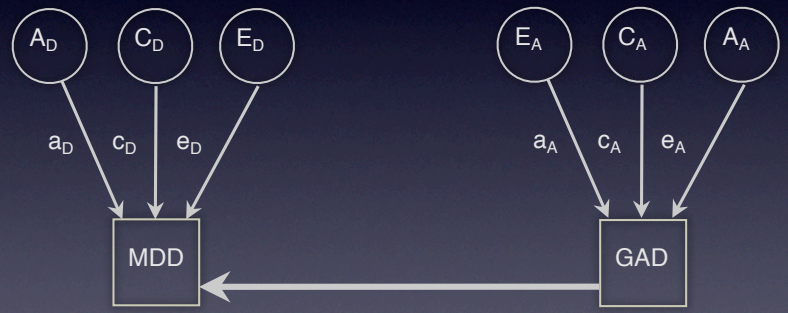
Modeling Comorbidity

Major Depression Causes Generalized Anxiety Disorder

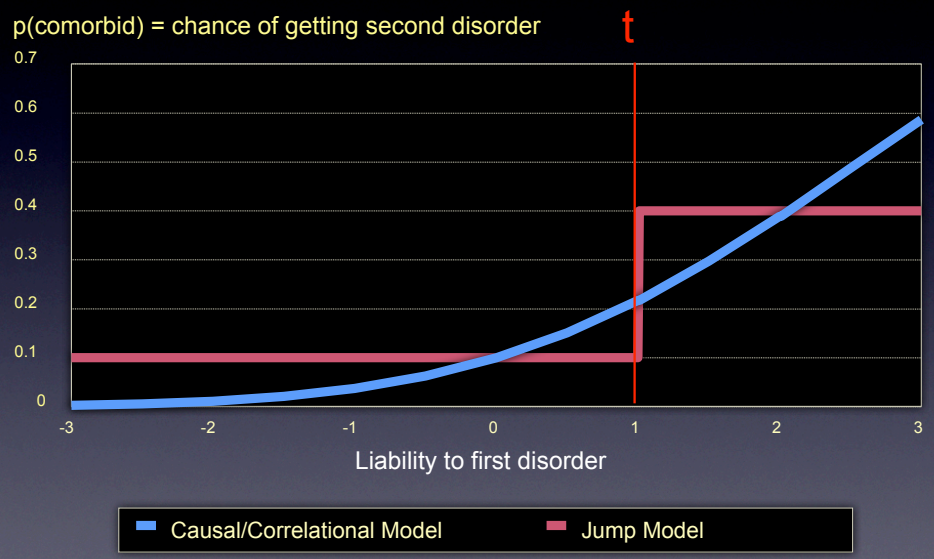


Modeling Comorbidity

Generalized Anxiety Disorder causes Major Depression

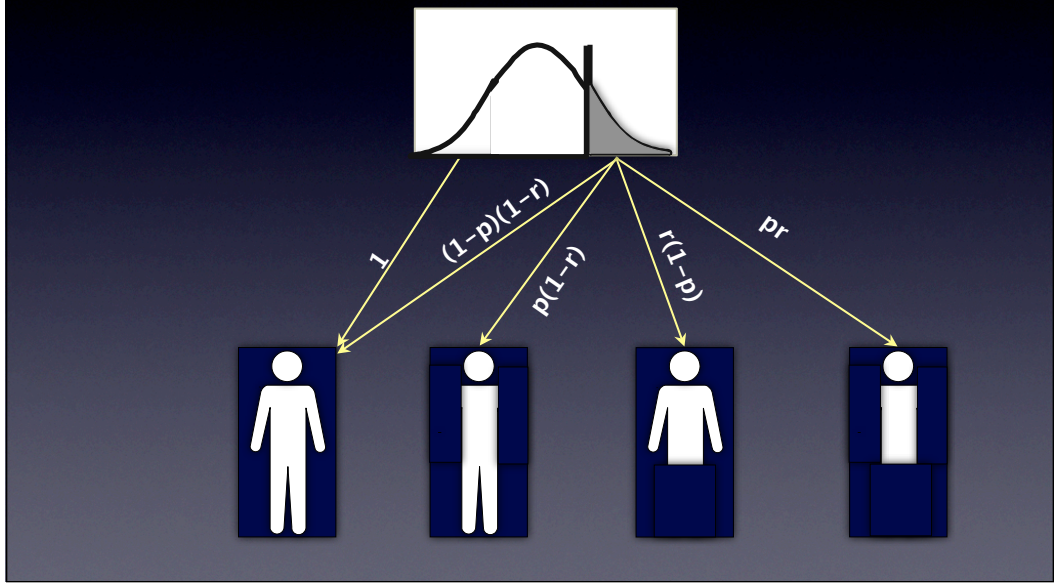


Alternative models of increasing risk to a second disorder



Threshold Model r=.5

Alternate forms: One underlying continuum



Alternate forms: Detail of pairs

$$P(\bar{A}1, \bar{B}1, \bar{A}2, \bar{B}2) = LL + 2(1-p)(1-r)UL + (1-p)^2(1-r)^2UU \quad (30)$$

$$P(\bar{A}1, \bar{B}1, \bar{A}2, B2) = r(1-p)LU + (1-p)^2r(1-r)^2UU \quad (31)$$

$$P(\bar{A}1, \bar{B}1, A2, \bar{B}2) = p(1-r)LU + p(1-p)(1-r)^2UU \quad (32)$$

$$P(\bar{A}1, \bar{B}1, A2, B2) = prLU + p(1-p)r(1-r)UU \quad (33)$$

$$P(\bar{A}1, B1, \bar{A}2, B2) = (1-p)^2r^2UU \quad (34)$$

$$P(\bar{A}1, B1, A2, \bar{B}2) = p(1-p)r(1-r)UU \quad (35)$$

$$P(\bar{A}1, B1, A2, B2) = p(1-p)r^2UU \quad (36)$$

$$P(A1, \bar{B}1, A2, \bar{B}2) = p^2(1-r)^2UU \quad (37)$$

$$LL_A = \int_{-\infty}^{t1_A}\int_{-\infty}^{t1_A} \phi(R_{A1}, R_{A2})dR_{A2}dR_{A1} \quad (24)$$

$$LM_A = \int_{-\infty}^{t1_A}\int_{t1_A}^{t2_A} \phi(R_{A1}, R_{A2})dR_{A2}dR_{A1} \quad (25)$$

$$LU_A = \int_{-\infty}^{t1_A}\int_{\Omega_A}^{\infty} \phi(R_{A1}, R_{A2})dR_{A2}dR_{A1} \quad (26)$$

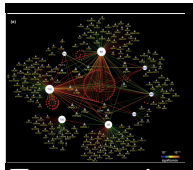
$$MM_A = \int_{t1_A}^{t2_A}\int_{t1_A}^{t2_A} \phi(R_{A1}, R_{A2})dR_{A2}dR_{A1} \quad (27)$$

$$MU_A = \int_{t1_A}^{t2_A}\int_{\Omega_A}^{\infty} \phi(R_{A1}, R_{A2})dR_{A2}dR_{A1} \quad (28)$$

$$UU_A = \int_{\Omega_A}^{\infty}\int_{\Omega_A}^{\infty} \phi(R_{A1}, R_{A2})dR_{A2}dR_{A1}\,, \quad (29)$$

$$P(A1, \bar{B}1, A2, B2) = p^2r(1-r)UU \quad (38)$$

$$P(A1, B1, A2, B2) = p^2r^2UU\,. \quad (39)$$



Large Database Studies

Top genetic associations in seven autoimmune diseases and T2D.

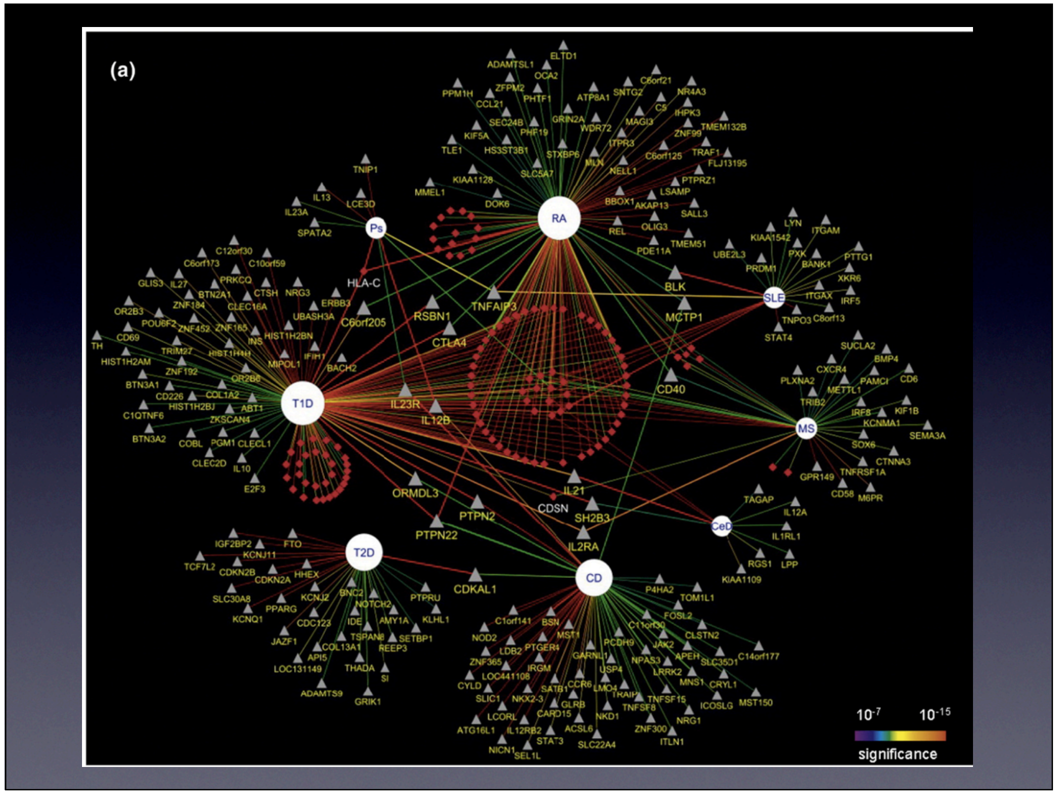
Most significant SNP per gene

Only associations with the significance of at least $P < 10^{-7}$ are visualized.

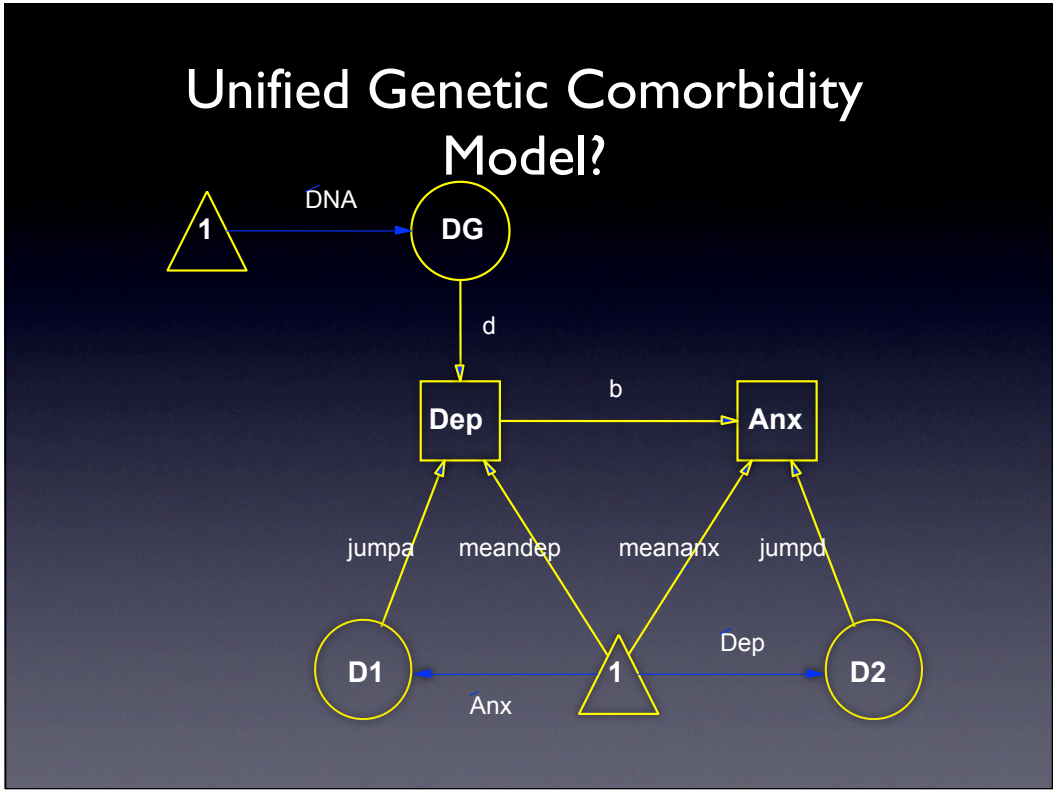
If a given gene was identified in more than one disease, multiple lines connecting it with each disease were drawn.

Lines are colored using a “heat” scheme according to the evidence for association. Thus “hot” edges (e.g. red, orange) represent more significant associations than “cold” edges (e.g. purple, blue).

Diseases are depicted by circles of size proportional to the number of associated genes, non-MHC genes by grey triangles, and genes in the MHC region are shown as red diamonds.


(a)
RA
Ps
SLE
MS
T1D
T2D
CD
CeD
HLA-C
C6orf205
RSBN1
TNFAIP3
BLK
MCTP1
CTLA4
CD40
IL23R
IL12B
ORMDL3
IL21
CDSN
PTPN2
SH2B3
PTPN22
IL2RA
CDKAL1
10-7
10-15
significance

Unified Genetic Comorbidity Model?



Conclusion

- Endophenotypes have potential
  - Gene-finding
  - Environment-finding
  - Understanding etiology
  - Are intrinsically comorbid with outcomes
  - Have Prevention & Treatment implications