Phenotypes
Endophenotypes and Comorbidity

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

Boulder Workshop 2011
Table of Contents

- Phenotype Measurement
- Definition of Endophenotype
- Models for Endophenotypes
- Models for Comorbidity
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 $p(\text{response}=Y)$
    - 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 \( \phi \)

Example item response probability shown in white
Example results for MD symptoms FF1
Structural Equation Model

- Two kinds of relationships
  - Linear regression $X \rightarrow Y$ single-headed
  - Unspecified Covariance $X \leftrightarrow Y$ 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)
Estimate factor score

Want ML estimate of this
Conditional on this factor score
Factor mean = \( f_s \)
Factor variance = 0

Item means move
Item variances shrink

\[ M_1 \quad M_2 \quad M_3 \quad M_4 \quad M_5 \quad M_6 \]
Common Pathway Model
Independent pathway model is submodel of 3 factor common pathway model
The joint analysis of personal interview and family history diagnoses: evidence for validity of diagnosis and increased heritability estimates

K. S. Kendler,† C. A. Prescott, K. Jacobson, J. Myers and M. 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
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)
- 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  Several

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?
- Error variances $\lambda$ 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
Establishing an endophenotype is a job for comorbidity modeling
Bivariate Model

Diagram of genetic pathways involving genes for endophenotype and psychiatric disorder, illustrating genetic influences and shared environmental factors.
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
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

Twin 1
Phenotype 1

Tobacco or EtOH Use

ROI

Twin 2
Phenotype 1

Tobacco or EtOH Use

ROI

A1
C1
A2
C2
E1
E2

a_{11}
a_{21}
c_{11}
c_{21}
a_{22}
c_{22}
e_{11}
e_{21}
e_{22}
### Significant Bivariate Results

#### Nicotine Abuse/Dependence

<table>
<thead>
<tr>
<th>ROI</th>
<th>rA</th>
<th>rC</th>
<th>rE</th>
<th>covA</th>
<th>covC</th>
<th>covE</th>
<th>rP</th>
</tr>
</thead>
<tbody>
<tr>
<td>LCerebellumWM</td>
<td>-0.271</td>
<td>-0.397</td>
<td>-0.054</td>
<td>-0.208</td>
<td>0.000</td>
<td>-0.035</td>
<td>-0.197</td>
</tr>
<tr>
<td>RCerebellumWM</td>
<td>-0.245</td>
<td>0.107</td>
<td>0.027</td>
<td>-0.188</td>
<td>0.000</td>
<td>0.018</td>
<td>-0.159</td>
</tr>
<tr>
<td>RCTXbankssts1</td>
<td>1.000</td>
<td>-1.000</td>
<td>-0.238</td>
<td>0.727</td>
<td>0.277</td>
<td>-0.150</td>
<td>-0.029</td>
</tr>
<tr>
<td>RCTXprecuneus1</td>
<td>-0.127</td>
<td>0.747</td>
<td>0.147</td>
<td>-0.098</td>
<td>0.000</td>
<td>0.094</td>
<td>-0.026</td>
</tr>
<tr>
<td>RCTXSUPtemporal1</td>
<td>-0.033</td>
<td>0.474</td>
<td>0.315</td>
<td>-0.026</td>
<td>0.000</td>
<td>0.202</td>
<td>0.086</td>
</tr>
<tr>
<td>RCTXtransversetemporal1</td>
<td>0.424</td>
<td>-0.999</td>
<td>-0.262</td>
<td>0.326</td>
<td>0.000</td>
<td>-0.168</td>
<td>0.059</td>
</tr>
<tr>
<td>RCTXbankssts2</td>
<td>1.000</td>
<td>-1.000</td>
<td>-0.242</td>
<td>0.759</td>
<td>0.162</td>
<td>-0.152</td>
<td>-0.075</td>
</tr>
<tr>
<td>LCTXLAToccipital2</td>
<td>-0.740</td>
<td>1.000</td>
<td>0.265</td>
<td>-0.554</td>
<td>0.081</td>
<td>0.174</td>
<td>-0.016</td>
</tr>
<tr>
<td>RCTXrostralMidfrontal2</td>
<td>0.594</td>
<td>1.000</td>
<td>-0.369</td>
<td>0.449</td>
<td>0.013</td>
<td>-0.242</td>
<td>-0.040</td>
</tr>
<tr>
<td>RCTXSUPtemporal2</td>
<td>0.056</td>
<td>0.826</td>
<td>0.263</td>
<td>0.043</td>
<td>0.000</td>
<td>0.168</td>
<td>0.169</td>
</tr>
<tr>
<td>RCTXtransversetemporal2</td>
<td>0.699</td>
<td>-0.621</td>
<td>-0.220</td>
<td>0.541</td>
<td>0.000</td>
<td>-0.139</td>
<td>0.106</td>
</tr>
<tr>
<td>RCTXTotalSurfaceArea2</td>
<td>0.528</td>
<td>-0.228</td>
<td>-0.256</td>
<td>0.397</td>
<td>0.000</td>
<td>-0.169</td>
<td>0.103</td>
</tr>
<tr>
<td>LCTXcaudalANTCING3</td>
<td>1.000</td>
<td>1.000</td>
<td>-0.252</td>
<td>0.753</td>
<td>0.134</td>
<td>-0.162</td>
<td>0.029</td>
</tr>
<tr>
<td>RCTXcorpuscallosum3</td>
<td>-0.636</td>
<td>1.000</td>
<td>0.273</td>
<td>-0.477</td>
<td>0.159</td>
<td>0.176</td>
<td>0.011</td>
</tr>
<tr>
<td>LCTXNparietal3</td>
<td>-0.128</td>
<td>1.000</td>
<td>0.354</td>
<td>-0.100</td>
<td>0.029</td>
<td>0.222</td>
<td>0.121</td>
</tr>
<tr>
<td>LCTXpostcentral3</td>
<td>0.390</td>
<td>-0.842</td>
<td>-0.336</td>
<td>0.297</td>
<td>0.000</td>
<td>-0.218</td>
<td>-0.009</td>
</tr>
<tr>
<td>RCTXrostralANTCING3</td>
<td>-1.000</td>
<td>1.000</td>
<td>0.254</td>
<td>-0.744</td>
<td>0.202</td>
<td>0.162</td>
<td>0.100</td>
</tr>
</tbody>
</table>
Is Brain Morphometry an Appropriate Endophenotype of Alcohol/Tobacco Use?

<table>
<thead>
<tr>
<th>Satisfied?</th>
<th>Requirement</th>
</tr>
</thead>
<tbody>
<tr>
<td>Y</td>
<td>Moderate heritability</td>
</tr>
<tr>
<td></td>
<td>Endophenotype and illness co-segregate within families</td>
</tr>
<tr>
<td></td>
<td>Found in affected family members at higher rate than in population</td>
</tr>
<tr>
<td>Y (kinda)</td>
<td>Association with illness in the population</td>
</tr>
<tr>
<td></td>
<td>State-independent</td>
</tr>
<tr>
<td></td>
<td>Association with causes rather than effects of disorders</td>
</tr>
<tr>
<td></td>
<td>Endophenotype should affect a disorder</td>
</tr>
<tr>
<td>Y</td>
<td>Should have continuous variation in a population</td>
</tr>
<tr>
<td>Y</td>
<td>Should be measured across several levels of analysis</td>
</tr>
</tbody>
</table>
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 6. A comorbidity network for MDD and GAD. Larger nodes represent more frequent symptoms, closer connections, higher centrality; thicker edges, higher frequency of co-occurrence; darker edges, stronger associations. Only edges with log odds ratio higher than 1.96 are represented. Centrally positioned nodes indicate higher degree of co-association. Nodes labeled with overlapping symptoms. Non-overlapping MDD symptoms are displayed on the left, non-overlapping GAD symptoms on the right.
Modeling Comorbidity
Reciprocal Causation

Diagram showing the relationships between MDD and GAD with variables A, C, and E.
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: $r = 0.5$

Causal/Correlational Model

Jump Model

$p(\text{comorbid}) = \text{chance of getting second disorder}$

Liability to first disorder
Alternate forms: One underlying continuum

\[ \frac{1}{pr} \]

\[ (1-p)(1-r) \]

\[ p(1-r) \]

\[ (1-p) \]
Alternate forms: Detail of pairs

\[
P(A_1, B_1, A_2, B_2) = LL + 2(1 - \rho)(1 - \rho)UU \\
+ (1 - \rho)^2(1 - \rho)^2UU
\]  
(30)

\[
P(A_1, B_1, \bar{A}_2, B_2) = r(1 - \rho)LU \\
+ (1 - \rho)^2(1 - \rho)^2UU
\]  
(31)

\[
P(A_1, B_1, A_2, B_2) = p(1 - \rho)LU \\
+ (1 - \rho)^2(1 - \rho)^2UU
\]  
(32)

\[
P(A_1, B_1, A_2, B_2) = \rho^2LU \\
+ (1 - \rho)^2(1 - \rho)^2UU
\]  
(33)

\[
P(A_1, B_1, \bar{A}_2, B_2) = (1 - \rho)^2UU
\]  
(34)

\[
P(A_1, B_1, A_2, B_2) = p(1 - \rho)LU \\
+ (1 - \rho)^2(1 - \rho)^2UU
\]  
(35)

\[
P(A_1, B_1, A_2, B_2) = p(1 - \rho)^2UU
\]  
(36)

\[
P(A_1, B_1, A_2, B_2) = \rho^2(1 - \rho)^2UU
\]  
(37)
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.
Unified Genetic Comorbidity Model?

DNA -> DG

DG -> d

Dep <- d

Anx <- b

Dep <- jumpa

Anx <- meandep

D1 <- meananx

D2 <- jumpd

Anx <- Dep

Dep <- D1

D2 <- Anx
Conclusion

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