Pitfalls in GxE Research and How to Avoid Them

Michael C Neale PhD

PGC Presentation

25 April 2016
Overview

- What is GxE? Why should we care?
- Latent G & E
- Measured G & E
- Modeling GxE at the locus level
- Conclusions
GxE Definition

- Genotype differentially affects responses to environmental events
- Environment differentially affects expression of genotype
- Statistical GxE, not Tabery’s developmental GxE
- Can conceive of different treatments according to Genotype or Exposure (Precision Medicine)
Mx
MCNeale et al

Variance Components Models for Gene–Environment Interaction in Twin Analysis

Shaun Purcell
Social, Genetic and Developmental Psychiatry Research Centre, Institute of Psychiatry, King's College, London, UK

A Note on False Positives and Power in G × E Modelling of Twin Data

Sophie van der Sluis · Danielle Posthuma · Conor V. Dolan
Pitfall 1: Measurement

- Binary: Ordinal: Continuous: Interval
- Aggregated: multiple binary items or symptoms
  - Sum Scores, Factor Scores, Weighted Sums
- Item Information Curves
- Test Information Curves
Van der Sluis & Dolan Model
Example: Brain & Cognition

* Significant moderation effect of ICV
Item Response Probability

Example item response probability shown in white
Possible population distribution in green

$$\phi(x)$$
normal pdf

Response Probability
Cumulative Normal
AFQT
100 Items

Subscales
1 Arithmetic Reasoning
2 Mathematics Knowledge
3 Word Knowledge
4 Paragraph Comprehension
AFQT: Overall Test Information Curve

More information at left

By design

Consequences for GxE?
Simulate to Estimate Effects of Uneven TIC on GxE Analyses

1. Generate Factor Scores using `mvrnorm()` in R

2. Generate 100 Item Responses from Factor Scores Consistent with Observed Factor Loading Estimates, then Compute Sum Score

3. Fit Models with vs. without Moderation Parameters $a_{21}$ $a_{22}$ $e_{21}$ $e_{22}$ to:
   A. True Factor Scores
   B. Sum Scores

4. Compare Type 1 Error Rates from Factor Scores vs. Sum Scores

5. Rinse & Repeat (10,000 times)
Marginal Distributions: Simulation

**Factor Score Distribution**

**Sum Score Distribution**
Does Moderator Affect Phenotype’s Parameters?
## Comparison

### Factor Scores

<table>
<thead>
<tr>
<th></th>
<th>All 4</th>
<th>A21</th>
<th>A22</th>
<th>E21</th>
<th>E22</th>
</tr>
</thead>
<tbody>
<tr>
<td>.01</td>
<td>0.01</td>
<td>0.01</td>
<td>0.01</td>
<td>0.01</td>
<td>0.01</td>
</tr>
<tr>
<td>.05</td>
<td>0.05</td>
<td>0.05</td>
<td>0.05</td>
<td>0.05</td>
<td>0.06</td>
</tr>
<tr>
<td>.10</td>
<td>0.11</td>
<td>0.10</td>
<td>0.10</td>
<td>0.10</td>
<td>0.11</td>
</tr>
</tbody>
</table>

### Sum Scores

<table>
<thead>
<tr>
<th></th>
<th>All 4</th>
<th>A21</th>
<th>A22</th>
<th>E21</th>
<th>E22</th>
</tr>
</thead>
<tbody>
<tr>
<td>.01</td>
<td>0.26</td>
<td>0.02</td>
<td>0.31</td>
<td>0.02</td>
<td>0.25</td>
</tr>
<tr>
<td>.05</td>
<td>0.45</td>
<td>0.07</td>
<td>0.51</td>
<td>0.06</td>
<td>0.44</td>
</tr>
<tr>
<td>.10</td>
<td>0.56</td>
<td>0.12</td>
<td>0.61</td>
<td>0.12</td>
<td>0.54</td>
</tr>
</tbody>
</table>

**Conclusion:** GxE Interaction Tests are Highly Sensitive to Inconsistently Informative Measurement
Estimated FTND Item Response Probability Curves
20-Year-Old Females
DSM Criteria: Substance Abuse & Dependence (All Drugs)

A1 Hazardous use
A2 Consequences: Legal
A3 Consequences: Social
A4 Consequences: Physical and psychological
A5 Used often when doing something ... important
A6 Stay away from school/ miss appointments

D1 Used more or longer than thought/planned
D2 Loss of control: Unable to stop/desire to stop, tried to cut down or stop using
D3 Spend time taking/using it, recovering from it, or doing whatever
D4 Used instead of work/hobbies
D5 Need for larger amounts/doses (tolerance)
D6 Withdrawal symptoms: Feeling sick when cutting down/stoppping
D7 Withdrawal symptoms: After not using ... use to prevent sickness
Comparison of Substances: Symptom Characteristic Curves

Used more or longer than thought/planned (D1)

GxE with Measured G & E

100% of the [logistic regression] tests yield significant G × E when none was included at the outset. That is, treating selected samples as if they were random is grossly misleading with respect to the detection of G × E.

Eaves LJ Twin Research and Human Genetics Volume 9 Number 1 pp. 1–8
What Does GxE Do to a Twin Model?

- Depends on Type of Environment:
  - Shared (C) vs Specific (E)
- GxC Inflates Estimates of G
- GxE Inflates Estimates of E
- Based on Statistical Model for Interactions
Basic Model for Genetic Variation

The diagram illustrates the basic model for genetic variation, showing the relationship between genotype (Aa) and phenotype effects. The model includes:

- **Aa**: Heterozygote
- **m**: Mid-homozygote
- **aa**: Decreasing
- **AA**: Increasing
- **h**: Dominance deviation
- **d**: Homozygous effect

The model also indicates the direction of genetic variation and the influence of dominant and recessive alleles.
Basic Model for Genotype x Environment Interaction

- $p$
- $1-p$

Diagram:
- Mid-homozygote
- Decreasing $aa$
- Increasing $AA$
- Dominance deviation
- Homozygous effect
Basic Model for Genotype x Environment Interaction

\[ f \]  
\[ 1-f \]  

\[ -a \]  
\[ a \]  

Decreasing \[ aa \]  
Increasing \[ AA \]  

\[ -b \]  
\[ b \]  

Decreasing \[ aa \]  
Increasing \[ AA \]  

Mid-homozygote

Homozygous effect
GxC at locus level inflates C as well as G

Sibling Genetic Covariance:
\[ f: \text{frequency of group 1} \]
\[ p: \text{allele frequency} \]
\[ a: \text{allelic effect at locus} \]

…..without exposure

\[ b: \text{allelic effect at locus} \]
…..with exposure

\[ a = 1 \]
\[ 1 < b < 2 \]

\[
\frac{(f - 1)^2(1 - 2p)^2(a - b)^2 - a^2(p^2 - p)}{(f - 1)^2(1 - 2p)^2(a - b)^2 - 2a^2(p^2 - p)}
\]

http://tinyurl.com/jt9qh3m.pdf
Other Gotchas in GxE Modeling

- Model Changes Effect Sizes of **Same** Genes
- Reparameterize to Estimate **New** Genetic Factors Expressed at other Ages
- At the Locus Level, Note 3 Genotypes, GxE may be **Inconsistent**
- Other **Non-Linear** Effects may seem like GxE
Single Locus Interaction

Three Genotypes

- **Usual Approach:** Regression on E, on G (scored 0/1/2) and on E x G
- If additivity assumption is incorrect - bias ensues
- **Solution 1:** 6 parameter model
  - Aliev et al (2014) Behav Genet 44:165-181
- **Solution 2:** Moderate variance components’ responses to environments
Other Issues

- Non-linear effect of genotype on phenotype
- \( P = aA + bA^2 + cA^3 + \ldots + eE \)
- Latent E influences genetic variation (Bayes)
- G-E covariance in presence of GxE in non-experimental situations
- Be careful out there!
Thank You

- VIPBG & VU Colleagues
- Leuven
- Egmond-aan-Zee
- Boulder
- Alabama
- APS
- Penn State
- U Minnesota
- King’s College London