Genetic Studies of Comorbid Traits

Heping Zhang
Yale University
Comorbidity

The simultaneous presence of 2 or more morbid conditions or diseases in the same patient.
\[ \{M_k[y_i]\} = \frac{1}{P\{y_i\}} \prod_j [P\{y_i|c_j = 0\} P\{c_j = 0\} + P\{c_j = 0\} P\{c_j = 1\}] \]

\[ = \frac{P\{M_k\}}{P\{y_i\}} \prod_j [\pi(\beta; y_i, 0) P\{c\} + \pi(\beta; y_i, 0) P\{c\}] \]

\[ = P\{M_k\} \prod_j [\pi(\beta; y_i, 0) P\{c\}] \]

It is possible to see that \((\partial/\partial \beta) \pi(\beta; k, \epsilon) = \epsilon\)

\[ \log(P\{M_k[y_i]\}) = \frac{\partial}{\partial \beta} \log(P\{y_i\}) + \sum_j \frac{\partial}{\partial \beta} \log[\pi(\beta; y_j)] \]

The null hypothesis that \(\beta = 0\), we have

\[ \frac{\partial}{\partial \beta} \log[\pi(\beta; y_j, 0) P\{dd|M_k\}] = 1 - \gamma(0; y_j, 0) \gamma(0; y_j, 0) \]

\[ \frac{\partial}{\partial \beta} \log P\{y_i\}|_{\beta = 0} = \sum_j [1 - \gamma(0; y_j, 0)] \gamma(0; y_j, 0) \]

Convenience, we drop the two irrelevant

\[ \log(P\{M_k[y_i]\})|_{\beta = 0} = \sum_j [1 - \gamma(0; y_j)] - \gamma(0; y_j, 0) \gamma(0; y_j, 0) \]

The coefficient of linkage disequilibrium:

\[ \{dd, AA\} - P\{dd, AA\} - P\{AA\}|P\{DL\} \]
Genetic Studies

\[
\{M_i|\gamma_i\} = \frac{1}{P(\gamma_i)} \prod_i \left[ P(\gamma_i|\epsilon_i = 0) \right] \prod_i P(\gamma_i|\gamma_i - 1, \epsilon_i = 0, \gamma_i = 0), \text{ and } \gamma_i|\gamma_i, K.
\]

The null hypothesis is

\[\frac{\partial}{\partial \beta} \log P(\gamma_i|\beta = 0) = \sum_j [1 - \gamma_j (1 - \gamma_j)] - \gamma_i (1 - \gamma_j)
\]

For convenience, we drop the two irrelavant terms

\[\log(P(M_i|\gamma_i))|\beta = 0 = \sum_j [1 - \gamma_j (1 - \gamma_j)] - \gamma_i (1 - \gamma_j) - \sum_j \gamma_j (1 - \gamma_j) - \gamma_i (1 - \gamma_j)
\]
Outline

Comorbidity of Psychiatric Disorders
- A Century Ago
- As We Are Speaking
- Association Analysis of Multivariate Traits
- Data Analysis of Alcoholism
- Closing Comments and Acknowledgements
PRELIMINARY REPORT OF A STUDY OF HEREDITY IN INSANITY IN THE LIGHT OF THE MENDELIAN LAWS

BY GERTRUDE L. CANNON, A.M., AND A. J. ROSANOFF, M.D.
KINGS PARK STATE HOSPITAL, NEW YORK

Insane hospital statistics show plainly that heredity has much to do with the causation of certain forms of nervous and mental disease. Yet we know but little of the exact conditions under which such disease is transmitted from parent to offspring. The object of the present research has been to accumulate and examine such data as may serve to throw some light upon this obscure problem.

It has been shown that the laws governing the transmission of traits by heredity, as established by Mendel, hold good not only for plants and the lower animals, but also for man, at least as regards certain characters, such as color of hair and color of eyes. In view of this fact our problem has assumed for us a more definite form. It is simply: Are any of the forms of nervous and mental disease transmitted from generation to generation in accordance with the Mendelian laws?

§ 1. The Mendelian Laws.—Perhaps a brief review of the essential facts of the Mendelian laws will not be superfluous.

The total inheritance of an individual from his parents is divisible into unit characters, each of which is inherited independently of all the rest and may therefore be studied without reference to other characters.

The inheritance of any such character is believed to be dependent upon the presence in the germ plasm of a unit of substance called a determiner.

With reference to any given character the condition in an individual may be dominant or recessive: the character is dominant when, depending upon the presence of its determiner in the germ plasm, it is plainly manifest; and it is recessive when, owing to the lack of its determiner in the germ plasm, it is not present in the individual under consideration.
Correlated Phenotypes

\[
\{M_i[y_i]\} = \frac{P\{y_i\}}{P\{y_i\}} \prod_i [P\{y_i; \beta_i = 0\} P\{\beta_i\}] = \prod_i [P\{y_i; \beta_i = 0\} P\{\beta_i\}]
\]

\[
P\{y_i\} = \prod_i [P\{y_i; \beta_i = 0\} P\{\beta_i\}]
\]

\[
= \prod_i [P\{y_i; \beta_i = 0\} P\{\beta_i\}]
\]

\[
\frac{\partial}{\partial \beta} \log(P\{M_i[y_i]\}) = \frac{\partial}{\partial \beta} \log(P\{y_i\}) + \sum_j \frac{\partial}{\partial \beta} \log(P\{x_j\})
\]

The null hypothesis that \(\beta = 0\), we have

\[
\frac{\partial}{\partial \beta} \log(P\{M_i[y_i]\})|_{\beta = 0} = \sum_j [1 - \gamma(0; y_j; 0) - 1 - \gamma(0; y_j; 0)]
\]

The pedigree charts contain a number of instances of neuropathic children born of normal parents, but not a single instance of a normal child born of parents both of whom are neuropathic. This proves that the neuropathic make-up cannot be dominant over normal; but that if its transmission occurs at all in a manner corresponding to the Mendelian laws, it must be recessive to normal.

In preparing the pedigree charts we have made use of the following symbols and abbreviations.

\(\square\) = male individual. \(\circ\) = female individual. A square or a circle unmarked = normal individual. \(P\) = normal individual with neuropathic offspring. \(I\) = insanity. \(Cv\) = convulsions. \(E\) = epilepsy. \(N\) = feeble-mindedness, hysteria, or other pronounced neuropathic manifestation. \(o\) within a square = normal individual without offspring. \(\dagger\) = died in childhood. \(?\) = data unascertained.

Number above each mating indicates type of combination.
Fagerstrom Test for Nicotine Dependence (FTND)

1. How many cigarettes a day do you usually smoke?
   - 1 to 10: 0 point
   - 11 to 20: 1 point
   - 21 to 30: 2 points
   - 30 or more: 3 points

2. How soon after you wake up do you smoke your first cigarette?
   - After 60 minutes: 0 point
   - 31-60 minutes: 1 point
   - 6-30 minutes: 2 points
   - <5 minutes: 3 points

3. Do you smoke more during the first two hours of the day than during the rest of the day?
   - No: 0 point
   - Yes: 1 point

4. Which cigarette would you most hate to give up?
   - Any other cigarette than the first one: 0 point
   - The first cigarette in the morning: 1 point

5. Do you find it difficult to refrain from smoking in places where it is forbidden, such as public buildings, on airplanes or at work?
   - No: 0 point
   - Yes: 1 point

6. Do you still smoke even when you are so ill that you are in bed most of the day?
   - No: 0 point
   - Yes: 1 point

Total points
Subclinical syndromes (e.g., minor depression and heavy drinking) probably influence smoking initiation and cessation more because they are so much more prevalent. In prospective studies, comorbidity predicts smoking and smoking predicts comorbidity (Hughes J R 1999)
Comorbid psychiatric disorders are common and their determinants are multi-factorial. Comorbidity
Nuclear Families

\[
\begin{align*}
\{M_t[y_t]\} &= \frac{1}{P[y_t]} \prod_i P[y_t]_i e_i = 0 \} P[M_t] \\
&= \frac{P[M_t]}{P[y_t]} \prod_i \{\pi(\beta, y_i, 0) P[y_i] \}
\end{align*}
\]

\[
\beta \text{; } k \text{; } e = P[y_k = k; e] = \gamma(\beta; k) \text{; } K - 1, \gamma(\beta; 0, e) = 0, \text{ and } \gamma(\beta; K, e).
\]

\[
P[y_i] = \prod_i \{P[y_i]_i e_i = 0 \} P[y_i] \\
&= \prod_i \{\pi(\beta, y_i, 0) P[y_i] \},
\]

possible to see that \((\partial/\partial \beta) \pi(\beta; k, e) = e \)

\[
\log(P[M_t[y_t]]) = \frac{\partial}{\partial \beta} \log(P[y_t]) \\
+ \sum_i \frac{\partial}{\partial \beta} \log(\pi(\beta; y_i))
\]

The null hypothesis that \(\beta = 0\), we have

\[
\frac{\partial}{\partial \beta} \log(\pi(\beta; y_i, 0)) P[y_i]_i M_k = [1 - \gamma(0; y_i, 0)] \gamma(0; y_i)
\]

\[
\frac{\partial}{\partial \beta} \log P[y_i]|_{\beta = 0} = \sum_i \frac{1}{P[Y_i]_i M_k}
\]

Convenience, we drop the two irrelevant

\[
\log(P[M_t[y_t]])|_{\beta = 0} = \sum_i [1 - \gamma(0; y_i)] - \gamma(0; y_i)
\]

\[
= \sum_i \frac{1 - \gamma(0; y_i) - \gamma(0; y_i)}{P[Y_i]_i M_k}
\]

The coefficient of linkage disequilibrium:

\[
\{M_t[y_t]\} - P[dd, AA] - P[AA]|P[DL]
\]
Although we do not observe the causal relationship between the genotypes and traits or among the traits, we generate the data from 40 directed acyclic graphs (DAGs).

An arrow between any two elements points to a causal relationship.
DAGs 1-20
DAGs 21-40
Power: Quantitative Traits (Alpha=0.01)

FBAT: dots and FBAT-GEE: triangles.

Black: \( \rho_{kj} = -1 \), Red: \( \rho_{kj} = 0.2 \), Green: \( \rho_{kj} = -0.2 \).

\( r^2 = 0.35 \)

\( r^2 = 0.15 \)

\( r^2 = 0.05 \)
Kendall’s Tau

Kendall’s Tau: a non-parametric statistic measuring the strength of the relationship between two variables.

Let \((X_i, Y_i)\) and \((X_j, Y_j)\) be a pair of observations. If \(X_j - X_i\) and \(Y_j - Y_i\) have the same sign, we say that the pair is concordant. If they have different sign, we say that the pair is discordant.

For a sample size \(n\). The Kendall Tau is defined as

\[
\tau = \frac{2(C - D)}{n(n-1)}
\]

where \(C\) and \(D\) are the number of concordant and discordant pairs.
A vector of traits $T = (T^{(1)}, \ldots, T^{(p)})'$ and a vector of markers $M = (M^{(1)}, \ldots, M^{(G)})'$.
Let $u_{ij} = (f_1(T_i^{(1)} - T_j^{(1)}), ..., f_p(T_i^{(p)} - T_j^{(p)}))'$

where can be the identity function for a quantitative or binary trait, or the sign function for an ordinal trait (or any trait).

Let $\nu_{ij} = (C_i(1) - C_j(1), ..., C_i(G) - C_j(G))'$

$C$ is a function of marker $M$ such as the count of any chosen allele of genotype.
Let \( U = \left( \begin{array}{c} n \\ 2 \end{array} \right)^{-1} \sum_{i<j} u_{ij} \otimes v_{ij} \)

Kendall’s tau:

\[
W = U' \text{Cov}_0^{-1} (U \mid T) U \sim \chi^2_{\text{rank}(\text{Cov}_0(U|T))} - \text{distributed}
\]
Association Test

For a family study

\[ \text{Cov}_0(U \mid T) = \]

The \( i \)th member in the \( k \)th sibship is the \( d_k(i) \)th subject in the entire study cohort.

\[ \bar{u}_i = \frac{1}{n} \sum_{j=1}^{n} u_{ij} \]
Nominal type I error comparison

The coefficient of linkage disequilibrium $\delta$ takes value of 0

Power evaluation

The coefficient of linkage disequilibrium $\delta$ takes value of 0.11

Given the genotype at the trait locus, a non-proportional odds model is used to generate ordinal phenotype data and a Gaussian distributed model is used for quantitative phenotype.
## Type I error comparison

<table>
<thead>
<tr>
<th>#(family)</th>
<th>K</th>
<th>O-FBAT</th>
<th>FBAT</th>
<th>O-FBAT</th>
<th>FBAT</th>
<th>O-FBAT</th>
<th>FBAT</th>
</tr>
</thead>
<tbody>
<tr>
<td>200</td>
<td>3</td>
<td>0.043</td>
<td>0.044</td>
<td>0.009</td>
<td>0.009</td>
<td>0.001</td>
<td>0.001</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>0.049</td>
<td>0.051</td>
<td>0.008</td>
<td>0.007</td>
<td>0.001</td>
<td>0.001</td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>0.059</td>
<td>0.062</td>
<td>0.013</td>
<td>0.01</td>
<td>&lt;0.001</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td></td>
<td>6</td>
<td>0.047</td>
<td>0.043</td>
<td>0.005</td>
<td>0.005</td>
<td>&lt;0.001</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>400</td>
<td>3</td>
<td>0.049</td>
<td>0.051</td>
<td>0.012</td>
<td>0.009</td>
<td>0.002</td>
<td>0.002</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>0.055</td>
<td>0.054</td>
<td>0.009</td>
<td>0.011</td>
<td>0.001</td>
<td>0.001</td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>0.042</td>
<td>0.041</td>
<td>0.006</td>
<td>0.006</td>
<td>0.001</td>
<td>0.002</td>
</tr>
<tr>
<td></td>
<td>6</td>
<td>0.045</td>
<td>0.045</td>
<td>0.006</td>
<td>0.008</td>
<td>0.001</td>
<td>0.001</td>
</tr>
<tr>
<td>600</td>
<td>3</td>
<td>0.036</td>
<td>0.038</td>
<td>0.006</td>
<td>0.006</td>
<td>&lt;0.001</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>0.054</td>
<td>0.055</td>
<td>0.013</td>
<td>0.010</td>
<td>0.001</td>
<td>0.001</td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>0.061</td>
<td>0.055</td>
<td>0.005</td>
<td>0.009</td>
<td>0.001</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td></td>
<td>6</td>
<td>0.038</td>
<td>0.038</td>
<td>0.006</td>
<td>0.007</td>
<td>&lt;0.001</td>
<td>&lt;0.001</td>
</tr>
</tbody>
</table>
## Power Comparison

<table>
<thead>
<tr>
<th>#(family)</th>
<th>K</th>
<th>alpha = 0.05</th>
<th>alpha = 0.01</th>
<th>alpha = 0.001</th>
</tr>
</thead>
<tbody>
<tr>
<td>200</td>
<td>3</td>
<td>0.783</td>
<td>0.553</td>
<td>0.261</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.778</td>
<td>0.541</td>
<td>0.249</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>0.732</td>
<td>0.492</td>
<td>0.213</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.702</td>
<td>0.456</td>
<td>0.184</td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>0.760</td>
<td>0.541</td>
<td>0.277</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.672</td>
<td>0.429</td>
<td>0.193</td>
</tr>
<tr>
<td></td>
<td>6</td>
<td>0.504</td>
<td>0.266</td>
<td>0.076</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.403</td>
<td>0.184</td>
<td>0.042</td>
</tr>
<tr>
<td>400</td>
<td>3</td>
<td>0.980</td>
<td>0.922</td>
<td>0.757</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.982</td>
<td>0.916</td>
<td>0.752</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>0.961</td>
<td>0.882</td>
<td>0.664</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.946</td>
<td>0.857</td>
<td>0.627</td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>0.978</td>
<td>0.914</td>
<td>0.757</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.949</td>
<td>0.839</td>
<td>0.604</td>
</tr>
<tr>
<td></td>
<td>6</td>
<td>0.792</td>
<td>0.584</td>
<td>0.328</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.664</td>
<td>0.437</td>
<td>0.203</td>
</tr>
<tr>
<td>600</td>
<td>3</td>
<td>0.999</td>
<td>0.989</td>
<td>0.958</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.999</td>
<td>0.991</td>
<td>0.954</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>0.996</td>
<td>0.978</td>
<td>0.920</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.988</td>
<td>0.970</td>
<td>0.885</td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>0.999</td>
<td>0.987</td>
<td>0.935</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.990</td>
<td>0.957</td>
<td>0.837</td>
</tr>
<tr>
<td></td>
<td>6</td>
<td>0.947</td>
<td>0.826</td>
<td>0.582</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.859</td>
<td>0.658</td>
<td>0.379</td>
</tr>
</tbody>
</table>
Collaborative Studies on Genetics of Alcoholism (COGA)

- In United States, 12.5% of Adults has ever had alcohol dependence problem in their life time (Hasin, et al, 2007)
- A large scale, multi-center study to map alcohol dependence susceptible genes.
- 143 families with 1614 individuals. 4720 SNPs from Illumina genotype data set.
- One ordinal trait with 4 levels was recorded (pure unaffected, never drank, unaffected with some symptoms, and affected).
- FBAT was also used for comparison
Application for COGA Data

• Phenotypes:
  — Alcohol DX-DSM3R+Feighner (ALDX1)
  • 4 categories
  — Maximum number of drinks in a 24 hour period (MaxDrink)
  • 4 categories
  — Spent so much time drinking, had little time for anything else (TimeDrink)
  • 3 categories
Single trait analysis

D7S679 with p-value 0.002879 for ALDX1 > 0.000538 = 0.05/(3*31)
Multiple traits analysis

P-value is 0.000553 < 0.0016129 = 0.05/31 at marker D7S679, which is around 1 cM away from D7S1793 that has been reported to have linkage evidence.
• Genetic studies of mental diseases involve many challenges: some are clinical, some are statistical, and some are scientific.

• We attempt to deal with an important issue on comorbidity and demonstrate the benefit to analyze comorbidity in genetic study.
Acknowledgements

Ching-Ti Liu  Xueqin Wang  Wensheng Zhu
Thank You!