

Classical segregation analysis

28.10.2005
GE02 day 4 part 3

Yurii Auchenko
Erasmus MC Rotterdam

Segregation analysis: dominant disease

- Segregation analysis assuming rare dominant disease:
 - Ascertainment: collect all families with one parent affected
 - Check if the segregation of the phenotype in offspring is 1:1, using either exact binomial test or Normal (chi-squared) approximation

Segregation analysis: recessive disease

- Segregation analysis assuming rare recessive disease:
 - Ascertainment: collect all families with at least one affected offspring
 - Check if the segregation of the phenotype in offspring is 3:1
- Is this OK?

Segregation analysis: recessive disease

- ...not really:
 - There are families with both parents heterozygous, but none of the offspring is because of random segregation. We will definitely miss such families in our analysis
- Ascertainment becomes important issue

Recessive disease: complete selection

- Let us assume that we know every case of the disease, and only families with no affected are missing
- This is called “complete selection” schema
 - Consider all possible families with heterozygous (MN) parents and 2 offspring. Offspring could be:
 - D,D – this one we ascertain
 - D,U – this one we ascertain
 - U,D – this one we ascertain
 - U,U – this one we DO NOT!

Recessive disease: complete selection

- What is the expected number of affected in families with 2 offspring, when at least one of the is affected?
- Let us consider probabilities of these families:
 - $P(D,D) = 1/16$
 - $P(D,U) = 3/16$
 - $P(U,D) = 3/16$
 - $P(U,U) = 9/16$

Recessive disease: complete selection

- Thus posterior probability
 - $P(D,D|\text{at least one is affected}) = 1/7$
 - $P(D,U|\text{at least one is affected}) = 3/7$
 - $P(U,D|\text{at least one is affected}) = 3/7$
- And expected number of affected is
 - $1/7 \cdot 2 + 6/7 = 8/7 = 1.14$
- What will be expected number of affected in a sibship of size s ?

Recessive disease: complete selection

$$E[s] = \sum_{i=0}^s i \cdot P(s \geq 1 \text{ affected}) = \sum_{i=0}^s i \cdot \frac{P(s)}{1 - P(0)} =$$

$$\frac{1}{1 - P(0)} \sum_{i=0}^s i \cdot P(s) = \frac{1}{1 - (3/4)^s} \sum_{i=0}^s i \cdot P(s) = \frac{(s/4)}{1 - (3/4)^s}$$

- say, for $s = 2$, $(2/4) / (1 - 3/4^2) = 7/8 = 1.14$
- say, for $s = 3$, $(3/4) / (1 - 3/4^3) = 7/8 = 1.30$
- say, for $s = 4$, $(4/4) / (1 - 3/4^4) = 7/8 = 1.46$
- Using this formula we can compute the expected number of affected and unaffected in total sample. Then we can use chi-squared test to check if this is in agreement with the observed data

Single selection

- In some countries, male (only) have to do compulsory military service
- Sex distribution in families of soldiers:
 - 129 soldiers approached with the question “how many boys and girls have been born in your family?”
 - Results: 228 boys and 95 girls
 - Sex ratio of 2.4 : 1 in favour of male
 - Test versus null of 1:1 gives $\chi^2 = 54.76$ at 1 d.f.!

Single selection

- Of course only the families with at least one boy were assessed!
- The chance that two brothers are at military service at the same time is negligible
- Correction: exclude probands. Then,
- $228 - 129 = 99$ boys and 97 girls.
 - This is fine 1 to 1
 - $\chi^2 = 0.08$ at 1 d.f. (not significant)

Recessive disease: single selection

- Compute expected number of affected as
 - $\frac{1}{4} (\text{total_affected} - \text{no_families}) + \text{no_families}$
- Use chi2 – test to test deviation of expected vs. observed

Ascertainment

- Complete selection
 - families without affected members have 0 chance to be in the sample. Others (with at least one affected member) has the probability of getting into the sample, which is proportional to the population frequency of these families
- Single selection
 - The pedigrees are sampled via one affected member only. Every affected person may become a proband with some (small probability). Thus, the probability that a family is sampled is proportional to the number of affected

Binomial schema

- Every affected person has some probability, π , to become a proband. Thus, generally, for $0 < \pi < 1$ the more is the number of affected members, the higher is the chance for this pedigree to be ascertained.
 - When $\pi \rightarrow 1$, this is complete selection
 - When $\pi \rightarrow 0$, this is single selection

Complex segregation analysis

- To correct for selection bias, compute
 - $P(\text{data} \mid \text{probands}) = P(\text{data})/P(\text{probands})$