

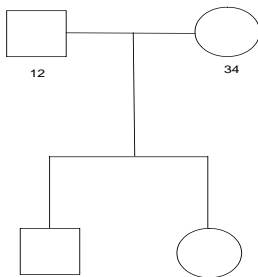
Nonparametric Linkage Analysis

Limitations of Parametric Linkage Analysis

- We previously discussed parametric linkage analysis
- Genetic model for the disease must be specified: allele frequency parameters and penetrance parameters
- Lod scores results are highly sensitive to the assumed mode of transmission of the disease, which will generally be unknown
- Nonparametric linkage analysis methods does not make any assumptions about the disease model

Sib Pair IBD Sharing Distribution

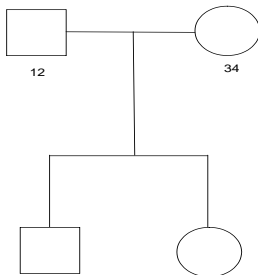
- Consider the nuclear family below with 2 siblings segregating alleles for a locus
- What is the probability of the siblings sharing 2, 1, or 0 alleles identical by descent (IBD)?



Sib Pair IBD Sharing Distribution

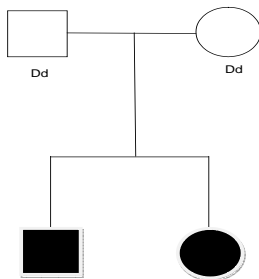
- Expected IBD Sharing

2	:	1	:	0
0.25	:	0.5	:	0.25



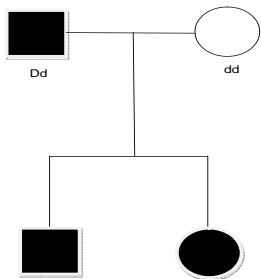
Affected Sib Pair Example

- Now consider a disease that is caused by a single locus.
- What would the allele sharing probabilities be for a sib pair at the disease locus?
- This depends on the mode of transmission of the disease. Assume for now that disease is caused by the D allele and D is recessive.



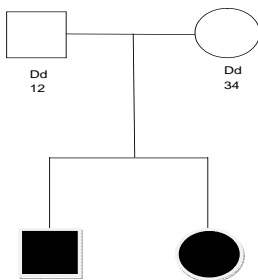
Affected Sib Pair Example

- Now assume that disease is caused by the D allele, and D is dominant.
- What would the allele sharing probabilities be for a sib pair at the disease locus?

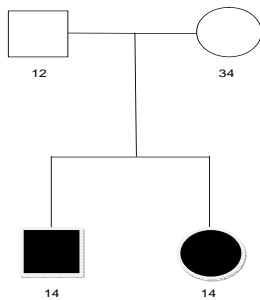
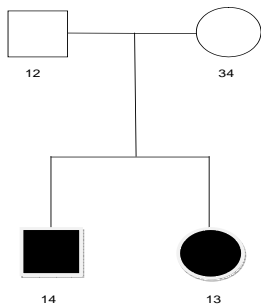


Affected Sib Pair Example

- The location of the disease gene is unknown and we would like to determine if the locus is linked to the disease gene.
- If the locus is linked to the disease gene, then the expected IBD probabilities of sharing 2, 1, and 0 alleles IBD for sibs at the disease gene will not be .25, .5, and .25, respectively, regardless of the mode of inheritance of the disease.



Affected Sib Pair Example



Affected Sib Pair Example

- The null hypothesis: locus is transmitted independently of the disease locus D/d.
- Under the null, the expected IBD sharing for sibs is

$$\begin{array}{rcl} 2 & : & 1 & : & 0 \\ 0.25 & : & 0.5 & : & 0.25 \end{array}$$

- Under the alternative, the locus is linked to the disease locus, and as a result, the IBD sharing probabilities do not follow the distribution specified under the null hypothesis.
- If the null is false, then you should see an increase in affected sibs sharing either 1 or 2 alleles IBD.
- For example if disease is caused by a rare dominant allele and the locus is tightly linked to the disease gene, then expected IBD sharing for sibs might be around

$$\begin{array}{rcl} 2 & : & 1 & : & 0 \\ 0.5 & : & 0.5 & : & 0 \end{array}$$

Affected Sib Pair Example

- More realistic scenario: marker is very close to locus which influences risk of disease in a more subtle manner (heterogeneity, epistasis, gene-environment interaction)

2	:	1	:	0
0.35	:	0.45	:	0.2

Model-Free Linkage Test

- The Pearson chi-squared goodness of fit test is a simple way of comparing the observed counts of sib pairs sharing 0, 1 and 2 alleles IBD with that expected under the null of no linkage.
- Let N be the number of affected sib pairs.
- Let n_i be the number of sib pairs that share i alleles IBD, where $i = 0, 1, \text{ or } 2$.
- Under the null, what is the expected value of n_i for each i ?
- Let the expected value of n_i under the null be E_{n_i} . The test statistic is:

$$\chi^2 = \sum_{i=0}^2 \frac{(n_i - E_{n_i})^2}{E_{n_i}}$$

- Under H_0 , the χ^2 test statistic has an approximate χ^2 distribution with 2 degrees of freedom

Extended Pedigrees

- Nonparametric linkage analysis can also be used for extended pedigrees, not just nuclear families with affected sib pairs
- Can calculate the expected IBD sharing for more distant relatives
- What is the expected IBD sharing probabilities for first cousins under the null?

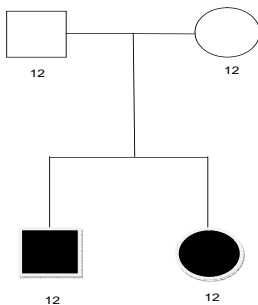
2	:	1	:	0
0	:	0.25	:	0.75

- What is the expected IBD sharing probabilities for second cousins under the null?

2	:	1	:	0
0	:	0.0625	:	0.9375

IBD Allele Sharing Uncertainty

- It may not be possible to determine exactly how many alleles a pair share IBD.
- In the example below, the affected sib pair could be sharing 2 or 0 alleles IBD, with each possibility having a probability of .5?



IBD Allele Sharing Uncertainty

- Methods to allow for this uncertainty developed, e.g., Kruglyak et al. (1996), Kong and Cox (1997).
- Multi-point method that incorporates the genotypes of nearby loci
- Obtain a probability distribution of IBD sharing at the locus being tested for linkage

Allele Sharing Statistics

- Allele sharing statistics S are often used for nonparametric linkage analysis. The general form of the statistics are

$$Z = \frac{S - \mu_0}{\sigma_0}$$

where μ_0 and σ_0 are the expected value and variance of S , respectively, calculated under the null hypothesis. If a locus is not linked to a disease, Z will follow a standard Normal distribution.

- There are various types of allele sharing statistics
- S_{pairs} counts, for each pair of affected relatives, the number of alleles shared IBD, and then sums that counts over all pairs of affected relatives.
- If all affected individuals in a pedigree have a common ancestor in the pedigree, S_{all} is the number of alleles shared IBD by all affected relatives.

Allele Sharing Statistics

- S_{max} is the size of the largest group of related cases who all inherit the same allele IBD (high power for dominant disease alleles)
- McPeck (1999) showed that the optimal sharing statistic depends on the disease model