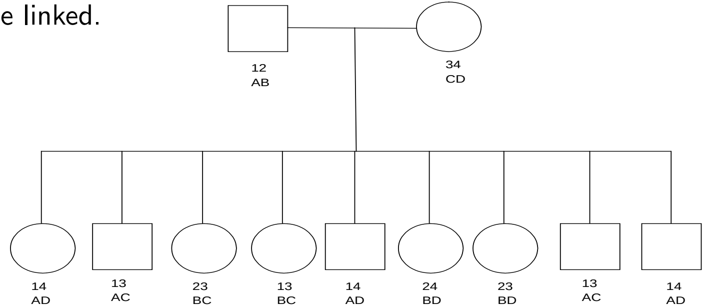
Parametric Linkage Analysis

Linkage Analysis

- Once aggregation and/or segregation studies established a genetic component for a phenotype of interest, parametric linkage analysis was the traditional approach used for Mendelian disease gene mapping since the 1970's
- Linkage analysis requires genetic marker data on pedigree.
- To illustrate linkage analysis, we will consider examples given by Suarez, B.K. and Cox, N.J. (1985)

Nuclear Family Example

- The figure below shows a large nuclear family segregating alleles from two loci: alleles at one of the loci are denoted by numbers while the alleles of the other are denoted by letters.
- Both of the parents are heterozygous at each locus and share no alleles in common, so the co-segregation of the alleles at the two loci can be unambiguously followed.
- We are interested in determining whether or not the two loci are linked.



Lod Scores

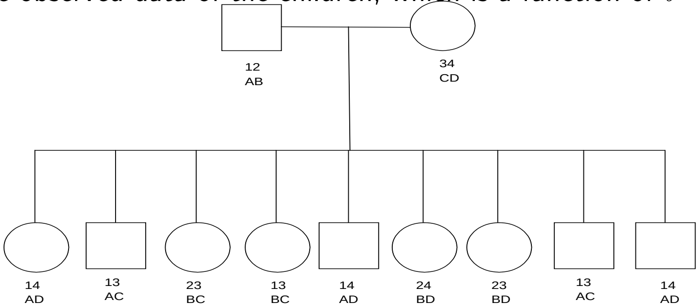
- **LOD scores** are calculated for recombination fraction θ values to determine if there is significant evidence for linkage
- For a given value of θ, the lod score is

$$log_{10} \frac{P(\text{observed data assuming recombination fraction is } \theta)}{P(\text{observed data assuming recombination fraction is } .5)}$$

- **LOD** stands for **L**og of **OD**ds
- Find the the value of θ that gives the maximum lod score
- Lod scores greater than 3 give evidence of linkage, and the null hypothesis of no linkage is rejected.
- How do you interpret a lod score equal to 3?
- Lod scores less than -2 give evidence that the loci are unlinked.

Nuclear Family Linkage Example

- Can calculate a lod score for the large nuclear family. We only observe the genotypes at the two loci so the phase is unknown. Possible phase for the parents:
    - 1A 2B 3C 4D
    - 1A 2B 3D 4C
    - 1B 2A 3C 4D
    - 1B 2A 3D 4C
- Given each parental phase type, can obtain the probability of the observed data of the children, which is a function of θ



Nuclear Family Linkage Example

Phase	1A,2B,3C,4D	1A,2B,3D,4C	1B,2A,3C,4D	1B,2A,3D,4C
Phase Probability	.25	.25	.25	.25
Offspring Probability	$\left(\frac{1}{2}\right)^{18}(1-\theta)^{16}\theta^2$	$\left(\frac{1}{2}\right)^{18}(1-\theta)^{9}\theta^9$	$\left(\frac{1}{2}\right)^{18}(1-\theta)^{9}\theta^9$	$\left(\frac{1}{2}\right)^{18}(1-\theta)^{2}\theta^{16}$

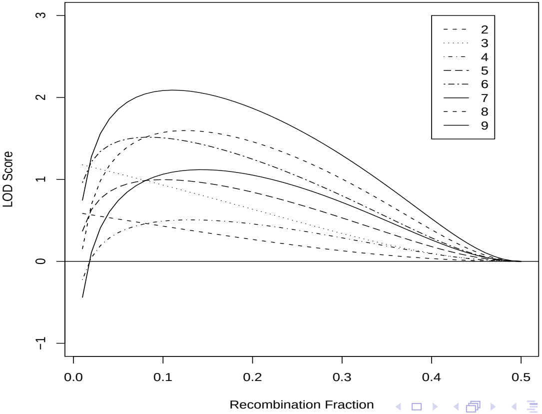
So, for $\theta = .1$ the lod score is

$$log_{10}\left(\frac{.25(.9)^{16}(.1)^2 + .5(.9)^9(.1)^9 + .25(.9)^2(.1)^{16}}{.25(.5)^{18} + (.5)^{19} + .25(.5)^{18}}\right)$$

$$= 2.08$$

Nuclear Family Linkage Example LOD Score Graph

- For linkage analysis with nuclear families, data must be available on at least 2 offspring
- The figure below gives the lod score curves obtained for the large nuclear family according to the number of children included in the calculation.



Nuclear Family Linkage Example

- The previous figure illustrates how the lod score curve changes as more information becomes available
- The lod score is always 0 at $\theta = \frac{1}{2}$ since the odds ratio is 1
- The lod score calculated using the first 2 children and using the first 3 children steadily increases as $\theta \longrightarrow 0$
- With the addition of the fourth child, the lod score curve changes from its monotonically increasing from as $\theta \longrightarrow 0$ to one that increases as θ moves away from $\frac{1}{2}$
- Evidence for linkage becomes a little stronger with the addition of the fifth and sixth children, and decreases with the seventh child (due to an apparent maternal recombinant), and then increases with the remaining 2 children

Nuclear Family Linkage Example

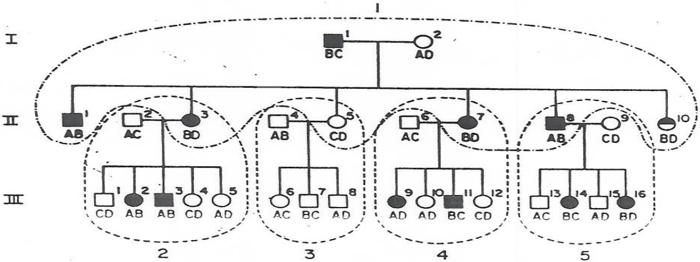
- This nuclear family provides moderate evidence that the 2 loci are linked. At $\hat{\theta} = 0.11$, the lod score curve reaches its maximum value of 2.09, indicating that the hypothesis of linkage with 11% recombination is about 123 times more likely than the hypothesis of no linkage
- Since the maximum lod score is in the range of -2 to 3, more families need to be sampled before a decision of $\theta = \frac{1}{2}$ or $\theta < \frac{1}{2}$ can be accepted or rejected.

Extended Family Linkage Example

- Linkage analysis for co-dominant loci is straightforward and a decision in favor or against the hypothesis of linkage can usually be reached with a few informative families.
- For linkage analysis with a disease, the true mode of inheritance of the disease will generally be unknown. Parametric linkage analysis requires a disease model to be specified, i.e., need to specify the allele frequency and penetrance **parameters**.

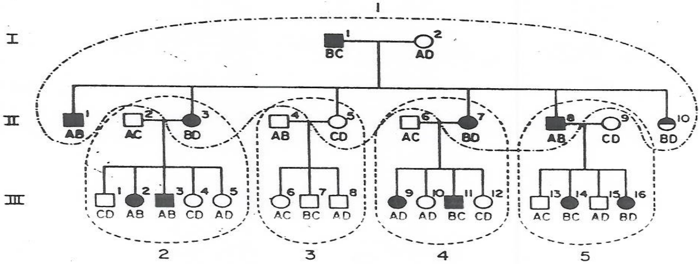
Extended Family Linkage Example

- In general, nuclear families are less efficient than extended 3-generation pedigrees because extended pedigrees provide more information regarding phase
- Consider the 28 member 3-generation pedigree below

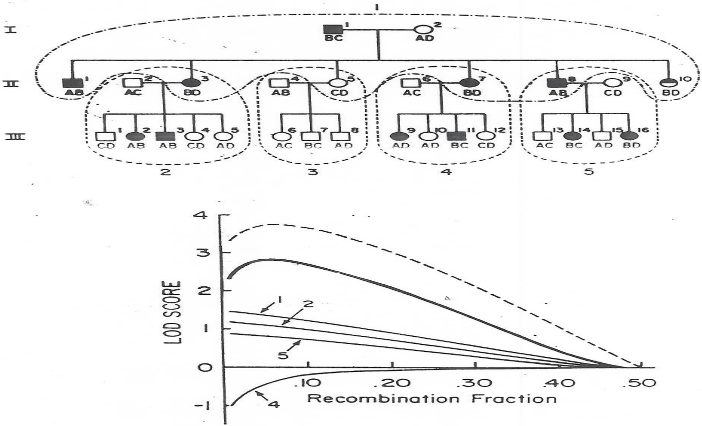


Extended Family Linkage Example

- We would like to determine if the locus with available genotype data is linked to a disease locus for which we do not know the location.
- What are the possible genotypes for the individuals in the pedigree if the disease is caused by a single autosomal locus that is fully penetrant and dominant ?



Extended Family Linkage Example



Extended Family Linkage Example

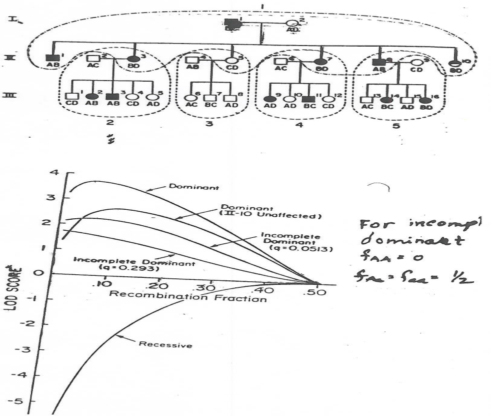
- Nuclear families 1, 2, and 5 provide evidence for linkage. The lod score curves monotonically increasing as $\theta \longrightarrow 0$ suggest that these families do not contain any recombinants. The different height of the lod score curves reflects that fact that larger nuclear families are more informative than smaller ones.
- Nuclear family 3 provides no information regarding linkage since neither partent is affected and at least one parent must be a double heterozygote to be informative.
- Nuclear family 4 provides slight evidence against the hypothesis of linkage.

Extended Family Linkage Example

- If the nuclear families were truly independent, then the lod scores could be summed, giving a maximum lod score of 2.81 at $\hat{\theta} = 0.05$.
- When analyzing the pedigree as a whole, the maximum is also at $\hat{\theta} = 0.05$ but with a lod score of 3.72.

Extended Family Linkage Example

- The plot below illustrates that misspecification of the mode of transmission of the disease affects the linkage analysis results.



For incompl dominant

$f_{AA} = 0$

$f_{Aa} = f_{aa} = 1/2$