Modeling IBD for Pairs of Relatives

Biostatistics 666

Previously ...

- Linkage Analysis of Relative Pairs
- IBS Methods
  - Compare observed and expected sharing
- IBD Methods
  - Account for frequency of shared alleles
  - Provide estimates of IBD sharing at each locus

IBS Linkage Test

$$\chi^2_{2df} = \sum_i \frac{\left(N_{IBS=i} - E\left[N_{IBS=i}\right]\right)^2}{E\left(N_{IBS=i}\right)}$$

- $E(N_{IBS=i})$ depends on N and allele frequencies
- Bishop and Williamson (1990)

Likelihood for Sibpair Data

$$L_i \propto \sum_{j=0}^{2} P(IBD = j \mid ASP) P(Genotypes \mid IBD = j) \propto \sum_{j=0}^{2} z_j w_{ij}$$

Risch (1990) defines

$$w_{ij} \propto P(Genotypes_i \mid IBD = j)$$

$$z_i = P(IBD = i \mid \text{affected relative pair})$$

MLS Statistic of Risch (1990)

$$L(z_0, z_1, z_2) = \prod_i \sum_j z_j w_{ij}$$

$$LOD = \log_{10} \prod_i \frac{\hat{z}_0 w_{i0} + \hat{z}_1 w_{i1} + \hat{z}_2 w_{i2}}{\frac{1}{4} w_{i0} + \frac{1}{2} w_{i1} + \frac{1}{4} w_{i2}} = \frac{\chi^2}{2 \ln 10}$$

The MLS statistic is the LOD evaluated at the MLEs of z_0, z_1, z_2

The $\hat{z}_0, \hat{z}_1, \hat{z}_2$ can be estimated using an E-M algorithm

Today ...

- Predicting IBD for affected relative pairs
  - Modeling marginal effect of a single locus
  - Relative risk ratio (λ_R)
- The Possible Triangle for Sibling Pairs
  - Plausible IBD values for affected siblings
  - Refinement of the model of Risch (1990)

Single Locus Model

1. Allele frequencies
   - For normal and susceptibility alleles
2. Penetrances
   - Probability of disease for each genotype

- Useful in exploring behavior of linkage tests
  - A simplification of reality
- Ignore effect of other loci and environment

Penetrance

- $f_{ij} = P(\ Affected \mid G = ij)$
- Probability someone with genotype *ij* is affected
- Models the marginal effect of each locus

Using Penetrances

- Allele frequency p
- Genotype penetrances f_{11}, f_{12}, f_{22}

- Probability of genotype given disease
  - P(G = ij | D) =
- Prevalence
  - K =

Pairs of Individuals

- A genetic model can predict probability of sampling different affected relative pairs
- We will consider some simple cases:
  - Unrelated individuals
  - Parent-offspring pairs
  - Monozygotic twins
- What do the pairs above have in common?

What we might expect …

- Related individuals have similar genotypes
- For a genetic disease…
- Probability that two relatives are both affected must be greater or equal to the probability that two randomly sampled unrelated individuals are affected

Relative Risk and Prevalence

- In relation to affected proband, define
  - K_R prevalence in relatives of type R
  - $\lambda_R = K_R/K$ increase in risk for relatives of type R
- λ_R is a measure of the overall effect of a locus
  - Useful for predicting power of linkage studies

Unrelated Individuals

- Probability of affected pair

$$
\begin{aligned}
P(a \text{ and } b \text{ affected}) &= \mathrm{P}(a \text{ affected})\mathrm{P}(b \text{ affected}) \\
&= \mathrm{P}(\text{affected})^2 \\
&= \left[p^2 f_{11} + 2p(1-p) f_{12} + (1-p)^2 f_{22}\right]^2 \\
&= K^2
\end{aligned}
$$

- For any two related individuals, probability that both are affected should be greater

Monozygotic Twins

- Probability of affected pair

$$
\begin{aligned}
P(MZ \text{ pair affected}) &= \sum_G P(G)P(a \text{ affected} \mid \mathrm{G})P(b \text{ affected} \mid G) \\
&= p^2 f_{11}^2 + 2p(1-p) f_{12}^2 + (1-p)^2 f_{22}^2 \\
&= K_{MZ} K \\
&= \lambda_{MZ} KK
\end{aligned}
$$

- λ_{MZ} will be greater than for any other relationship

Probability for Genotype Pairs

	Child			
Parent	A_1A_1	A_1A_2	A_2A_2	
A_1A_1	p_1^3	$p_1^2p_2$	0	p_1^2
A_1A_2	$p_1^2p_2$	p_1p_2	$p_1p_2^2$	$2p_1p_2$
A_2A_2	0	$p_1p_2^2$	p_2^3	p_2^2
	p_1^2	$2p_1p_2$	p_2^2	N pairs

Probability of Genotype Pairs and Being Affected

	Child			
Parent	A_1A_1	A_1A_2	A_2A_2	
A_1A_1	$p_1^3 f_{11}^2$	$p_1^2 p_2 f_{12} f_{11}$	0	
A_1A_2	$p_1^2 p_2 f_{11} f_{12}$	$p_1 p_2 f_{12}^2$	$p_1 p_2^2 f_{12} f_{22}$	
A_2A_2	0	$p_1 p_2^2 f_{12} f_{22}$	$p_2^3 f_{22}^2$	
				N pairs

Parent Offspring Pairs

- Probability of Affected Pair

$$
\begin{aligned}
P &= P(\textit{parent} \text{ and } \textit{child} \text{ affected}) \\
&= \sum_{G_P} \sum_{G_O} P(G_P, G_O) f_{G_P} f_{G_O} \\
&= \sum_{i} \sum_{j} \sum_{k} P(i,j,k) f_{ij} f_{ik} \\
&= p^3 f_{11}^2 + (1-p)^3 f_{22}^2 + p(1-p) f_{12}^2 + 2p^2(1-p) f_{11} f_{12} + 2p(1-p)^2 f_{22} f_{12} \\
&= KK_O \\
&= \lambda_O KK
\end{aligned}
$$

- λ will be lower for other unilineal relationships
- λ_o will be between 1.0 and λ_{MZ}

Point of Situation

- Probabilities of affected pairs for
  - Unrelated Individuals
  - Monozygotic Twins
  - Parent-Offspring Pairs
- Each of these shares a fixed number of alleles IBD …

For a single locus model...

$$\lambda_{IBD=2} = \lambda_{MZ}$$
$$\lambda_{IBD=1} = \lambda_{O}$$
$$\lambda_{IBD=0} = 1$$

$$K_{IBD=2} = K_{MZ}$$
$$K_{IBD=1} = K_{O}$$
$$K_{IBD=0} = 1$$

- Model ignores contribution of other genes and environment
- Simple model that allows for useful predictions
  - Risk to half-siblings
  - Risk to cousins
  - Risk to siblings

Affected Half-Siblings

- IBD sharing
  - 0 alleles with probability 50%
  - 1 allele with probability 50%

- This gives …

$$\lambda_H = \tfrac{1}{2}\lambda_O + \tfrac{1}{2} = \tfrac{1}{2}(\lambda_O + 1)$$

$$K_H = \tfrac{1}{2}K_O + \tfrac{1}{2}K = \tfrac{1}{2}(K_O + K)$$

Uni-lineal Relationships

$$\lambda_R = P(IBD = 1 \mid R)\lambda_O + P(IBD = 0 \mid R)$$

$$K_R = P(IBD = 1 \mid R)K_O + P(IBD = 0 \mid R)K$$

$P(IBD = 1)$ decreases 50% with increasing degree of relationship

$(\lambda_R - 1)$ also decreases 50% with increasing degree of unilineal relationship

Affected Sibpairs

- IBD sharing …
  - 0 alleles with probability 25%
  - 1 alleles with probability 50%
  - 2 alleles with probability 25%
- This gives …

$$\lambda_S = \tfrac{1}{4}\lambda_{MZ} + \tfrac{1}{2}\lambda_O + \tfrac{1}{4} = \tfrac{1}{4}(\lambda_{MZ} + 2\lambda_O + 1)$$

which implies

$$\lambda_{MZ} = 4\lambda_S - 2\lambda_O - 1$$

Examples: Full Penetrance

Recessive

p	f_{11}	f_{12}	f_{22}	K	Lambdas		
					MZ	Offspring	Sibling
.001	0	0	1	.000001	1000000	1000	250500
.01	0	0	1	.0001	10000	100	2550
.1	0	0	1	.01	100	10	30

Dominant

p	f_{11}	f_{12}	f_{22}	K	Lambdas		
					MZ	Offspring	Sibling
.001	0	1	1	.002	500.25	250.50	250.56
.01	0	1	1	.02	50.25	25.50	25.56
.1	0	1	1	.19	5.26	3.02	3.08

Examples: Incomplete Penetrance

Recessive

p	f_{11}	f_{12}	f_{22}	K	Lambdas		
					MZ	Offspring	Sibling
.001	.001	.001	1	.001	2.0	1.0	1.2
.01	.001	.001	1	.001	83.5	1.8	22.0
.1	.001	.001	1	.01	82.8	8.4	25.2

Dominant

p	f_{11}	f_{12}	f_{22}	K	Lambdas		
					MZ	Offspring	Sibling
.001	.001	1	1	.003	223	112	112
.01	.001	1	1	.02	46	23	23
.1	.001	1	1	.19	5	3	3

Examples: Small Effects

Smaller Effects

p	f_{11}	f_{12}	f_{22}	K	Lambdas		
					MZ	Offspring	Sibling
.1	.01	.02	.04	.012	1.2	1.1	1.1
.1	.01	.08	.16	.024	2.6	1.8	1.8
.1	.02	.16	.32	.048	2.6	1.8	1.8
.2	.01	.02	.04	.014	1.2	1.1	1.1
.2	.01	.08	.16	.038	2.1	1.6	1.6
.2	.02	.16	.32	.080	2.1	1.6	1.6

Multiple susceptibility loci...

- λ are upper bound on effect size for one locus
- λ decay rapidly for distant relatives
- If genes act multiplicatively, we can multiply marginal λ together

Another interpretation...

$$\lambda_{IBD=2} = \lambda_{MZ} = \frac{P(affected \mid IBD = 2 \text{ with affected relative})}{P(affected)}$$

$$\lambda_{IBD=1} = \lambda_{O} = \frac{P(affected \mid IBD = 1 \text{ with affected relative})}{P(affected)}$$

$$\lambda_{IBD=0} = 1 = \frac{P(affected \mid IBD = 0 \text{ with affected relative})}{P(affected)}$$

Bayes' Theorem: Predicting IBD Sharing

$$P(IBD = i \mid \text{affected pair}) =$$

$$= \frac{P(IBD = i)P(\text{affected pair} \mid IBD = i)}{\sum_j P(IBD = j)P(\text{affected pair} \mid IBD = j)}$$

$$= \frac{\lambda_{IBD=i}}{\sum_j P(IBD = j)\lambda_{IBD=i}}$$

Sibpairs
Expected Values for z_0, z_1, z_2

$$z_0 = 0.25\frac{1}{\lambda_s}$$

$$z_1 = 0.50\frac{\lambda_o}{\lambda_s}$$

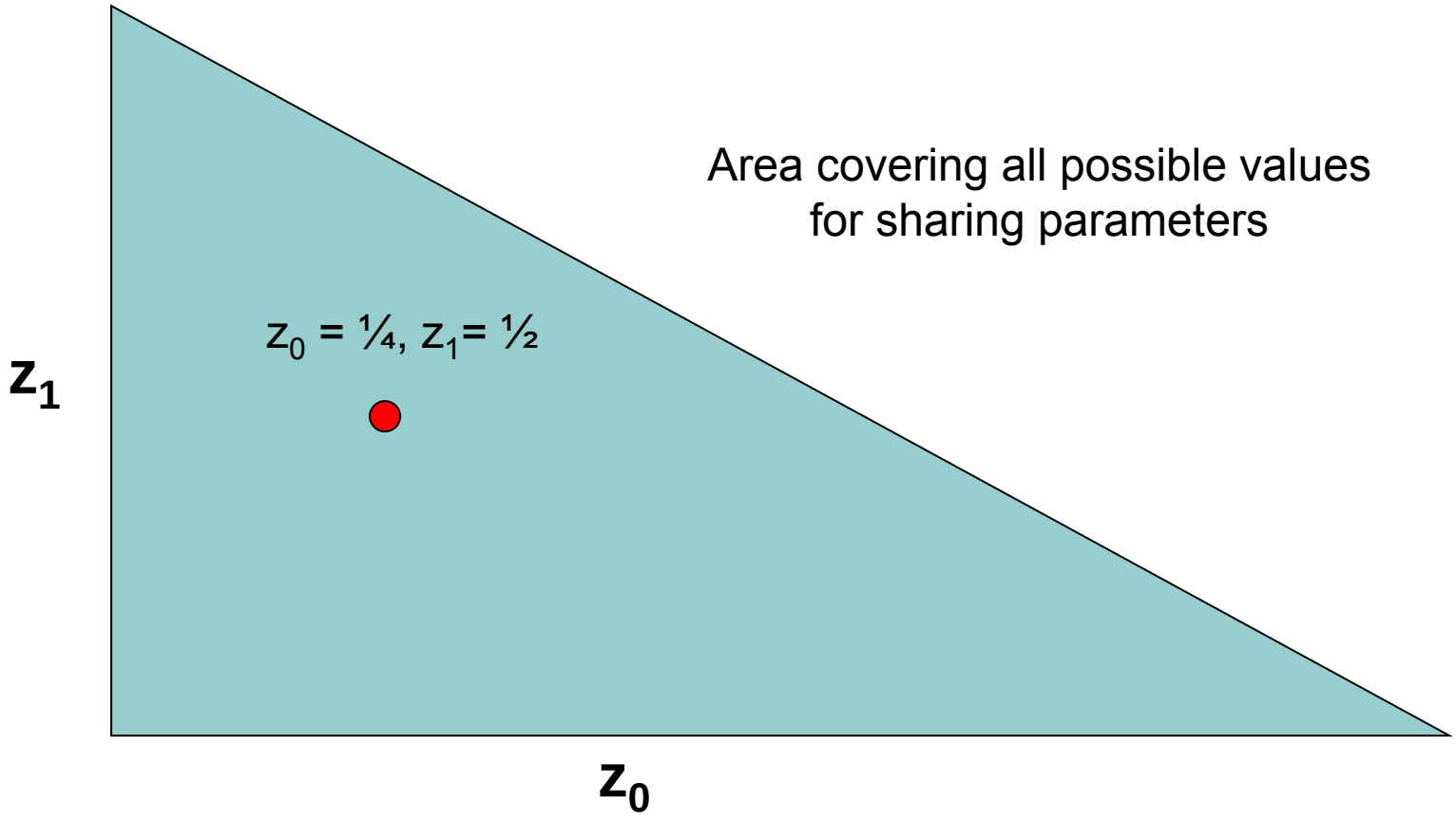
$$z_2 = 0.25\frac{\lambda_{MZ}}{\lambda_s}$$

$$1 \le \lambda_o \le \lambda_s \le \lambda_{MZ} \text{ for any genetic model}$$

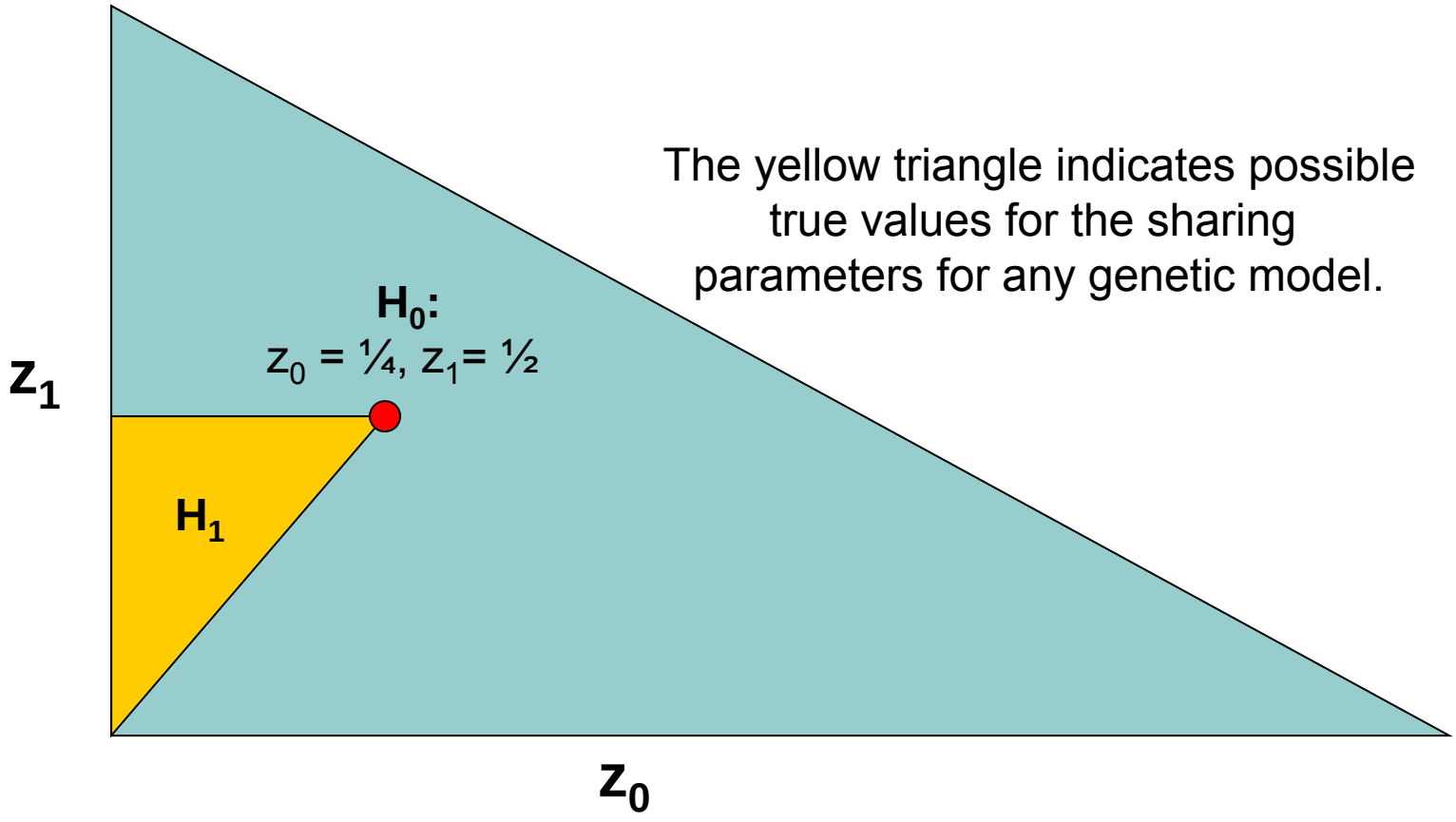
Maximum LOD Score (MLS)

- Powerful test for genetic linkage
- Likelihood model for IBD sharing
  - Accommodates partially informative families
- MLEs for IBD sharing proportions
  - Can be calculated using an E-M algorithm
- Shortcoming:
  - Sharing estimates may be implausible

Possible Triangle



Possible Triangle



The yellow triangle indicates possible true values for the sharing parameters for any genetic model.

Intuition

- Under the null
  - True parameter values are (¼, ½, ¼)
  - Estimates will wobble around this point
- Under the alternative
  - True parameter values are within triangle
  - Estimates will wobble around true point

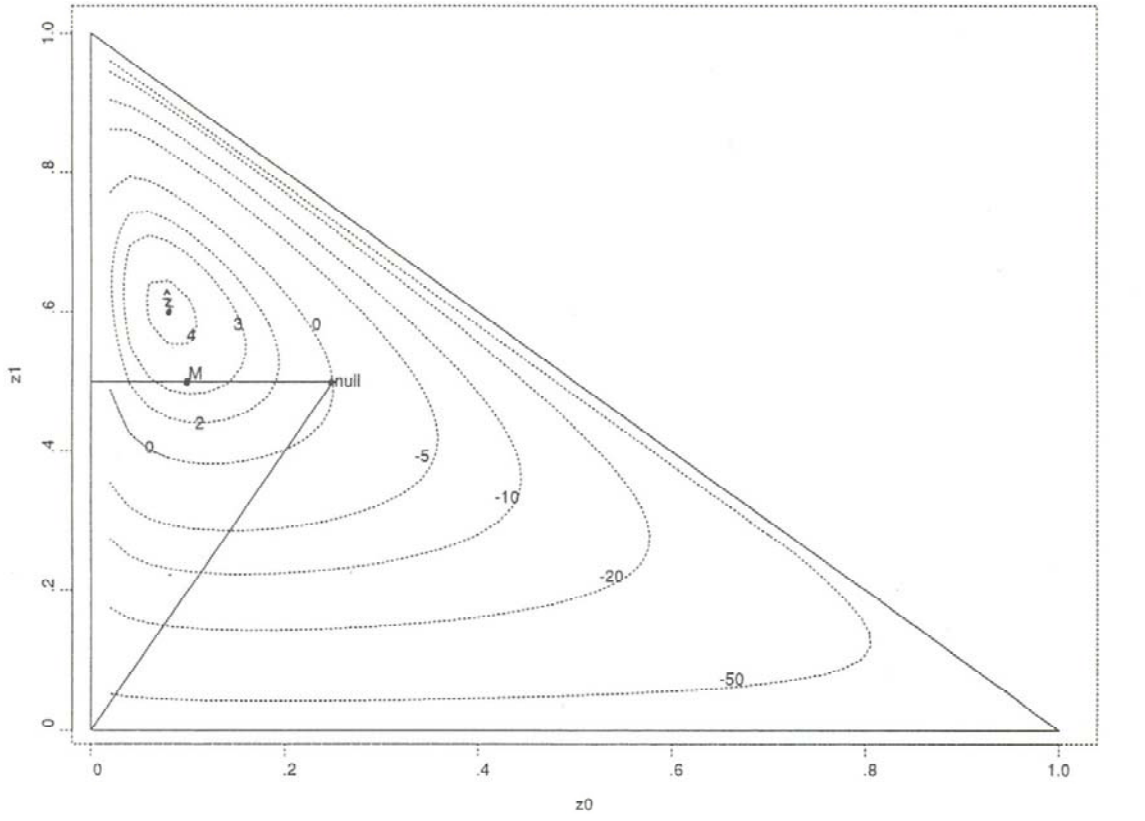
Idea (Holmans, 1993)

- Testing for linkage
  - Do IBD patterns suggest a gene is present?
- Focus on situations where IBD patterns are compatible with a genetic model
  - Restrict maximization to possible triangle

The possible triangle method

1. Estimate z_0, z_1, z_2 without restrictions
2. If estimate of $z_1 > ½$ then …
   a) Repeat estimation with $z_1 = ½$
   b) If this gives $z_0 > ¼$ then revert to null (MLS=0)
3. If estimates imply $2z_0 > z_1$ then …
   a) Repeat estimation with $z_1 = 2z_o$
   b) If this gives $z_0 > ¼$ then revert to null (MLS=0)
4. Otherwise, leave estimates unchanged.

Possible Triangle



Holman's Example:

IBD	Pairs
0	8
1	60
2	32

MLS = 4.22 (overall)
MLE = (0.08,0.60,0.32)

MLS = 3.35 (triangle)
MLE = (0.10,0.50,0.40)

MLS Combined With Possible Triangle

- Under null, true **z** is a corner of the triangle
  - Estimates will often lie outside triangle
  - Restriction to the triangle decreases MLS
  - MLS threshold for fixed type I error decreases
- Under alternative, true **z** is within triangle
  - Estimates will lie outside triangle less often
  - MLS decreases less
  - Overall, power should be increased

Example

- Type I error rate of 0.001
- LOD of 3.0 with unrestricted method
  - Risch (1990)
- LOD of 2.3 with possible triangle constraint
  - Holmans (1993)
  - For some cases, almost doubles power

Recommended Reading

- Holmans (1993)
  Asymptotic Properties of
  Affected-Sib-Pair Linkage Analysis
  Am J Hum Genet **52:**362-374

- Introduces possible triangle constraint
- Good review of MLS method

Reference

- Risch (1990)
  Linkage strategies for genetically complex traits. I. Multi-locus models.
  Am. J. Hum. Genet. ***46:****222-228*

- Recurrence risks for relatives.
- Examines implications of multi-locus models.