Am. J. Hum. Genet. 46:242–253, 1990

Linkage Strategies for Genetically Complex Traits. III. The Effect of Marker Polymorphism on Analysis of Affected Relative Pairs

Neil Risch

Departments of Epidemiology and Public Health, and of Human Genetics, Yale University School of Medicine, New Haven, CT

Summary

The results from the second paper of this series are reexamined for markers that are not completely polymorphic. A maximum lod score (MLS) criterion is defined for affected relative pairs. The expected MLS (EMLS) is calculated as a function of the marker polymorphic information content (PIC) for various values of λ_R (relative risk ratio) and different relative types by using simulations. An m-allele model with equal allele frequencies is employed. The EMLS is calculated for two sampling strategies: scheme 1, which uses pairs only, and scheme 2, which also includes additional informative relatives. For scheme 2, the percent of the maximum achievable EMLS (i.e., for a marker with a PIC of 1.0) is approximately equal to the marker PIC value for all relative types. For scheme 1, the EMLS is greatly diminished unless PIC is high, especially for distant relatives. For example, scheme 1 is not cost-effective for sibs unless PIC > .7; for second- and third-degree relatives, PIC must be >.85. Therefore, in general, it will be worthwhile to type additional relatives in linkage studies using affected pairs. The comparative value of sibs versus distant relatives depends on λ_R, recombination θ, and PIC. For large λ_R and PIC values, distant relatives are preferred. Alternatively, for smaller λ_R and PIC values, sibs are best.

Introduction

The preceding paper evaluated the power to detect linkage using pairs of affected relatives. Single and multilocus models of inheritance were considered. In all analyses, however, the marker was assumed to be 100% polymorphic. This would be appropriate for a highly polymorphic system of linked loci, such as HLA, and therefore represents the limiting case of maximal potential power to detect linkage under ideal circumstances. In general, markers are not so polymorphic. Therefore, it is important to reconsider the conclusions from the preceding paper while taking into account the polymorphic content of the marker.

When a genetic marker is not 100% polymorphic, sharing of alleles identical by descent (ibd) by a pair of relatives cannot be determined unequivocally, especially when other pertinent relatives have not been typed. In particular, while it is possible to determine that two relatives share zero alleles ibd, it is theoretically impossible to conclude that two relatives share one or two alleles ibd when the marker is not completely polymorphic, unless other informative relatives have been typed. To account for such uncertainty, a number of approaches have been suggested. Lange (1986) and Weeks and Lange (1988) recommend replacing ibd with identity by state (ibs); in contrast, Thomson (1986) suggested using only fully informative mating types for affected sib pairs.

Here, a maximum-likelihood method, as described by Risch (1989), is used to estimate the ibd probabilities; then a "maximum" lod score (MLS) method is described for relative pairs to assess the significance of linkage. The MLS statistic and its expectation are used to evaluate the relationship between genetic parameters and the power to detect linkage as a function of marker polymorphism.

Let α_i be the prior probability that two relatives share i alleles ibd, and let z_i be the posterior probability that two relatives share i marker alleles ibd given

Received May 23, 1989; revision received September 23, 1989.

Address for correspondence and reprints: Dr. Neil Risch, Department of Epidemiology and Public Health, Yale University School of Medicine, 60 College Street, P.O. Box 3333, New Haven, CT 06510.

0002-9297/90/4602-0004$02.00

that they are both affected ($i = 0,1,2$). The goal is to estimate the z_i from observed marker information.

In general, for a pair of affected relatives, let w_i be the probability of the observed marker phenotypes of the pair (and possibly other relevant relatives) given that they share i marker alleles ibd, $i = 0,1,2$. The likelihood of the observed data for the pair is then

$$L = \sum_{i=0}^{2} z_i w_i \,. \tag{1}$$

If L_j is the likelihood for the jth pair from a total of N independent pairs, then the likelihood for all N pairs is ΠL_j. The likelihood ratio for all N pairs is given by

$$\Lambda = \prod_{j=1}^{N} \left(\sum_{i=0}^{2} z_i w_{ij} \right) \Big/ \left(\sum_{i=0}^{2} \alpha_i w_{ij} \right) , \tag{2}$$

where w_{ij} corresponds to the probability w_i for the jth pair. The log of the likelihood ratio Λ can be interpreted as a lod score:

$$T = \log_{10}\Lambda \,. \tag{3}$$

In this case, the lod score is a function of both recombination and mode of inheritance. An MLS can be obtained by maximizing equation (3) with respect to z_i, $i = 0,1,2$. By analogy to the conventional significance criterion of an MLS >3 (Morton 1955), we can also apply the same criterion to T.

The z_i's can be estimated by maximum likelihood. This is equivalent to maximizing the lod score because the denominator in the lod score is constant. One simple method is by gene counting, or the E-M algorithm (Dempster et al. 1977). By this approach, the following recursive formula can be used:

$$z_i' = \frac{1}{N} \sum_{j=1}^{N} \frac{z_i w_{ij}}{z_0 w_{0j} + z_1 w_{1j} + z_2 w_{2j}} \,. \tag{4}$$

For a given value of j, the absolute values of w_{ij} are not important—only the *relative* values for $i = 0,1,2$.

As an example, consider the sample data for sib pairs given in table 1. Of a total of $N = 74$ pairs, only 32 have definite information about ibd; the remaining data are equivocal. For 22 pairs, the outcome was twice as likely if ibd = 2 than if ibd = 1 (e.g., two A_1A_2 children from an $A_1A_1 \times A_1A_2$ cross); for 13 pairs, the outcome was twice as likely if ibd = 0 than if ibd = 1 (e.g., A_1A_1 and A_1A_2 children from an $A_1A_1 \times A_1A_2$ cross); for seven pairs, the outcome was equally likely for ibd = 0 or ibd = 2 (e.g., both children A_1A_2 from an $A_1A_2 \times A_1A_2$ cross). When formula (4) is used, the maximum-likelihood estimates of the z_i's are $\hat{z}_2 = .512$, $\hat{z}_1 = .366$, $\hat{z}_0 = .122$, and the MLS = 2.65. Using the 32 definite cases only gives $z_2 = .44$, $z_1 = .50$, $z_0 = .06$, and the MLS = 2.20.

Table 1

Sample Data for MLS Method (for 74 sib pairs)

No. of Sib Pairs	w_2	w_1	w_0
14	1	0	0
16	0	1	0
2	0	0	1
22	½	¼	0
13	0	½	1
7	½	0	½

Linkage Information as a Function of PIC

In examining the relationship between PIC (Botstein et al. 1980) of a marker locus and the EMLS and power to detect linkage (MLS > 3.0), I have chosen the following types of relative pairs: sibs, uncle(aunt)-nephew(niece), grandparent-grandchild, half-sibs, and first cousins. Both theoretical arguments and simulations have been employed. The marker locus is assumed to be codominant with m alleles. Generally, two approaches are considered: (1) marker information is available for the relative pair only, and (2) marker information is also available for additional relevant relatives.

In particular, for scheme 2 refer to figure 1. For sibs (e.g., individuals 5 and 7), the parents (individuals 1 and 2) are also included; for grandparent-grandchild (e.g., individuals 1 and 10), the grandparent's spouse (individual 2) and the parents of the grandchild (individuals 6 and 7) are included; for uncle(aunt)-neph-

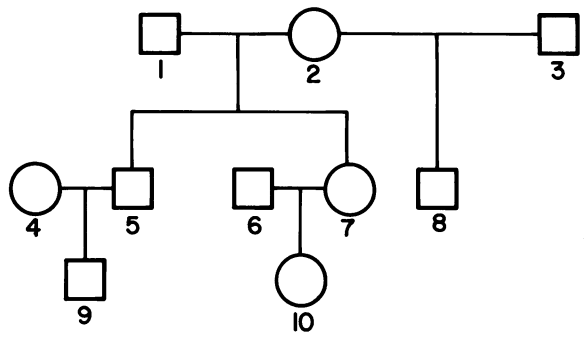

Figure 1 Pedigree depicting relatives for typing, when scheme 2 is used.

ew(niece) (e.g., individuals 5 and 10), the parents of the uncle(aunt) are included (individuals 1 and 2) as well as the parents of the nephew(niece) (individuals 6 and 7); for half-sibs (e.g., individuals 7 and 8), all three parents (e.g., individuals 1–3) are included; for first cousins (e.g., individuals 9 and 10), both sets of parents (individuals 4–7) and the common set of grandparents (individuals 1 and 2) are included.

To simplify the evaluation of the relationship with PIC, I consider only the case of m alleles of equal frequency ($1/m$), for $m = 2$ to $m = \infty$. While other situations with similar PIC values may not give identical results, reasonable generality can be assumed.

Information on Pairs Only (scheme 1)

For a pair of relatives, it is possible to define seven different ibs relationships. Consider the four distinct alleles A_i, A_j, A_k, and A_l. The seven ibs relationships are enumerated as follows: (1) A_iA_i,A_iA_i, (2) A_iA_i,A_iA_j, (3) A_iA_i,A_jA_j, (4) A_iA_i,A_jA_k, (5) A_iA_j,A_iA_j, (6) A_iA_j,A_iA_k, and (7) A_iA_j,A_kA_l. These seven states encompass all possible arrangements of ibs between and within individuals. Let the frequency of allele i be p_i. Then the probability of each of the seven possible arrangements as a function of the ibd parameters z_0, z_1, and z_2 can be calculated and is given in table 2. The lod score for each outcome is obtained by taking its probability (as a function of the z_i's) and dividing by the same probability, substituting α_i for z_i, and then taking the log (to base 10). The lod scores for the seven outcomes are also listed in table 2. The total probability for each of the seven outcomes is obtained by summing the probability in column two of table 2 over all possible combinations of p_i,p_j,p_k, and p_l. When $p_i = 1/m$ for $i = 1, \ldots, m$ is assumed, this amounts to doing the following: first setting p_i, p_j, p_k, and p_l each equal to $1/m$ in all formulas and then multiplying the probability for outcome (1) by m; for outcome (2) by $m(m-1)$; for outcomes (3) and (5) by $m(m-1)/2$; for outcomes (4) and (6) by $m(m-1)(m-2)/2$; and for outcome (7) by $m(m-1)(m-2)(m-3)/8$. The division by 2 for outcomes (3)–(6) and by 8 for outcome (7) is due to symmetry of alleles in these configurations. The values for outcomes (3), (4), and (7) can be combined into a single number (designated outcome [3]), since these cases have identical lod scores. The results of these calculations are given in table 3. The EMLS and the probability that MLS > 3 can be calculated using the lod scores and probabilities listed in table 3.

Using Information on Additional Relatives (scheme 2)

Calculating expected lod scores and power is inherently more difficult when additional relatives are available. For this case, different mating types for the parents must be considered. For example, refer to figure

Table 2

Probabilities and Lod Scores for ibs Arrangements

Arrangement	Probability	Lod Score
(1) $A_iA_i,A_iA_i \ldots$	$z_2p_i^2 + z_1p_i^3 + z_0p_i^4$	$\log \frac{z_2 + z_1p_i + z_0p_i^2}{\alpha_2 + \alpha_1p_i + \alpha_0p_i^2}$
(2) $A_iA_i,A_iA_j \ldots$	$z_1(2p_i^2p_j) + z_0(4p_i^3p_j)$	$\log \frac{z_1 + 2z_0p_i}{\alpha_1 + 2\alpha_0p_i}$
(3) $A_iA_i,A_jA_j \ldots$	$z_0(2p_i^2p_j^2)$	$\log \frac{z_0}{\alpha_0}$
(4) $A_iA_i,A_jA_k \ldots$	$z_0(4p_i^2p_jp_k)$	$\log \frac{z_0}{\alpha_0}$
(5) $A_iA_j,A_iA_j \ldots$	$z_2(2p_ip_j) + z_1(p_ip_j)(p_i + p_j) + z_0(4p_i^2p_j^2)$	$\log \frac{2z_2 + z_1(p_i+p_j) + z_0(4p_ip_j)}{2\alpha_0 + z_1(p_i+p_j) + z_0(4p_ip_j)}$
(6) $A_iA_j,A_iA_k \ldots$	$z_1(2p_ip_jp_k) + z_0(8p_i^2p_jp_k)$	$\log \frac{z_1 + 4z_0p_i}{\alpha_1 + 4\alpha_0p_i}$
(7) $A_iA_j,A_kA_l \ldots$	$z_0(8p_ip_jp_kp_l)$	$\log \frac{z_0}{\alpha_0}$

Table 3

Probabilities of Lod Score Outcomes for *m*-allele Model

Lod Score	Probability
(1) $\log \frac{z_2 + z_1/m + z_0/m^2}{\alpha_2 + \alpha_1/m + \alpha_0/m^2}$	$z_2/m + z_1/m^2 + z_0/m^3$
(2) $\log \frac{z_1 + 2z_0/m}{\alpha_1 + 2\alpha_0/m}$	$z_1\frac{2(m-1)}{m^2} + z_0\frac{4(m-1)}{m^3}$
(3) $\log \frac{z_0}{\alpha_0}$	$z_0 \frac{(m-1)(m^2-3m+3)}{m^3}$
(5) $\log \frac{z_2 + z_1/m + 2z_0/m^2}{\alpha_2 + \alpha_1/m + 2\alpha_0/m^2}$ . . .	$z_2\frac{(m-1)}{m} + z_1\frac{(m-1)}{m^2} + 2z_0\frac{(m-1)}{m^3}$
(6) $\log \frac{z_1 + 4z_0/m}{\alpha_1 + 4\alpha_0/m}$	$z_1\frac{(m-1)(m-2)}{m^2} + z_0\frac{4(m-1)(m-2)}{m^3}$

1. For sib pairs (individuals 5 and 7), all possible mating types for the parents (individuals 1 and 2) must be considered. For uncle(aunt)-nephew(niece) pairs (e.g., individuals 5 and 10), the mating type of individuals 1 and 2 as well as the genotype of individual 6 are needed. For half-sibs (individuals 7 and 8), the three parents (individuals 1–3) are required; for grandparent-grandchild pairs (e.g., individuals 1 and 10), the remaining parent (individual 6) and grandparent (individual 2) are needed. For first cousins (individuals 9 and 10), the shared grandparents (individuals 1 and 2) and parents (individuals 4–7) need to be included.

The general approach is as follows: For each mating type for individuals 1 and 2, all possible outcomes for the remaining relatives of interest are considered. The probability of each outcome is calculated given that the relative pair shares two, one, or zero alleles ibd. This gives the probability of the given outcome under the assumption of ibd values z_2, z_1, and z_0. The same probability can be calculated using instead the null hypothesis parameters α_2, α_1, and α_0. Taking the log of the ratio of these two probabilities gives the lod score for that observation. For all observations with identical lod scores, their probabilities are summed, yielding a total probability for a given lod score within the specified mating type. There are seven distinguishable mating types based on ibs. These are the same as the distinct types of sib-pair arrangements listed in table 2.

Sib Pairs

Consider first the case of sib pairs and parents. There are assumed to be m marker alleles which are enumerated $A_1, \ldots, A_m$. The seven possible distinct mating types and their population frequencies are listed in table 4. Also listed in table 4 are the possible lod score outcomes and their total probability of occurrence. As an example of how these probabilities are calculated, consider mating type (2). There are three possible outcomes in the sibs: (*a*) both are A_iA_i; (*b*) one is A_iA_i and the other is A_iA_j; and (*c*) both are A_iA_j. The probability of outcome (*a*) when the sibs share two alleles ibd is ½; the probability when they share one allele ibd is ¼; and the probability when they share zero alleles ibd is 0. Therefore, the probability of outcome (*a*) is $z_2 \times ½ + z_1 \times ¼$, which leads to a lod score of $\log[(2z_2 + z_1)/(2\alpha_2 + \alpha_1)]$. The probability of outcome (*b*) when the sibs share two alleles ibd is 0; the probability when they share one allele ibd is ½; the probability when they share zero alleles ibd is 1. Therefore, the probability of outcome (*b*) is $z_1 \times ½ + z_0$, leading to a lod score of $\log[(2z_0 + z_1)/(2\alpha_0 + \alpha_1)]$. Finally, the probability of outcome (*c*) when the sibs share two alleles ibd is ½; the probability when they share one allele ibd is ¼; the probability when they share zero alleles ibd is 0. Therefore, the probability of outcome (*c*) is $z_2 \times ½ + z_1 \times ¼$, leading to a lod score of $\log[(2z_2 + z_1)/(2\alpha_2 + \alpha_1)]$. Hence, there are two possible lod score outcomes; $\log[(2z_2 + z_1)/(2\alpha_2 + \alpha_1)]$ occurs with frequency $z_2 + ½\, z_1$, and $\log[(2z_0 + z_1)/(2\alpha_0 + \alpha_1)]$ occurs with frequency $z_0 + ½\, z_1$. The possible lod score outcomes and their probabilities can be calculated in a similar fashion for the other matings and are listed in table 4.

By examining table 4, one can observe that there are a total of six distinct nonzero lod scores. A total probability for each lod score can be obtained by summing over all mating types. Again, I am assuming that the m alleles have equal frequency ($1/m$), simplifying the

Table 4

Lod Scores and Probabilities for Sib Pairs, When Scheme 2 Is Used

Mating type	Frequency	Lod Score	Probability
(1) $A_iA_i \times A_iA_i$	p_i^4	0	1
(2) $A_iA_i \times A_iA_j$	$4p_i^3p_j$	$\log \frac{2z_2 + z_1}{2\alpha_2 + \alpha_1}$	$z_2 + \frac{1}{2}z_1$
		$\log \frac{2z_0 + z_1}{2\alpha_0 + \alpha_1}$	$z_0 + \frac{1}{2}z_1$
(3) $A_iA_i \times A_jA_j$	$2p_i^2p_j^2$	0	1
(4) $A_iA_i \times A_jA_k$	$4p_i^2p_jp_k$	$\log \frac{2z_2 + z_1}{2\alpha_2 + \alpha_1}$	$z_2 + \frac{1}{2}z_1$
		$\log \frac{2z_0 + z_1}{2\alpha_0 + \alpha_1}$	$z_0 = \frac{1}{2}z_1$
(5) $A_iA_j \times A_iA_j$	$4p_i^2p_j^2$	$\log \frac{z_2}{\alpha_2}$	$\frac{1}{2}z_2$
		$\log \frac{z_1}{\alpha_1}$	z_1
		$\log \frac{z_0}{\alpha_0}$	$\frac{1}{2}z_0$
		$\log \frac{z_2 + z_0}{\alpha_2 + \alpha_0}$	$\frac{1}{2}z_2 + \frac{1}{2}z_0$
(6) $A_iA_j \times A_iA_k$	$8p_i^2p_jp_k$	$\log \frac{z_2}{\alpha_2}$	z_2
		$\log \frac{z_1}{\alpha_1}$	z_1
		$\log \frac{z_0}{\alpha_0}$	z_0
(7) $A_iA_j \times A_kA_l$	$8p_ip_jp_kp_l$	$\log \frac{z_2}{\alpha_2}$	z_2
		$\log \frac{z_1}{\alpha_1}$	z_1
		$\log \frac{z_0}{\alpha_0}$	z_0

summations. The final probabilities corresponding to each of the six nonzero lod score outcomes are given in table 5. The values in this table will be used to determine power to detect linkage and EMLSs.

Grandparent-Grandchild Pairs

Consider individuals 1 and 10 in the pedigree of figure 1. If individual 7 (the intervening parent) is homozygous, there is no information regarding sharing of alleles ibd by individuals 1 and 10. If individual 7 is a heterozygote, it will be possible to determine unequivocally whether persons 1 and 10 share an allele ibd, provided that individuals 1, 2, and 7 do not have the same heterozygous genotype and that individuals 6, 7, and 10 also do not have the same heterozygous genotype. Under all other circumstances, there is no information about allele sharing between individuals 1 and 10. The probability that individual 7 has the heterozygous genotype A_iA_j and that individuals 1 and 2 are not both A_iA_j and also that individuals 6 and 10 are not both A_iA_j *is* $2p_ip_j(1\text{-}p_ip_j)^2$. When this expression is summed over all i and $j \neq i$, one obtains $[(m\text{-}1)/m](1\text{-}m^{-2})^2 = (1\text{-}m^{-2})^2(1\text{-}m^{-1})$. Therefore, the probability that a grandparent-grandchild pair unequivocally share one allele ibd is $z_1(1\text{-}m^{-2})^2(1\text{-}m^{-1})$, which has a corresponding lod score of $\log(z_1/\alpha_1)$. Hence, for this case, there are two possible lod score outcomes, which are shown in table 6 along with their corresponding probabilities.

By considering the pedigree in figure 1, one can observe that the ibd and marker relationship for half-sibs (e.g., individuals 7 and 8) is the same as that for grandparent-grandchild. Therefore, the values given in table 6 apply equally to half-sib pairs.

Table 5

Lod Scores and Probabilities for *m*-allele Model for Sib Pairs, When Scheme 2 Is Used

Lod Score	Probability
$\log \frac{2z_2 + z_1}{2\alpha_2 + \alpha_1}$	$(2z_2 + z_1)\frac{(m-1)}{m^2}$
$\log \frac{2z_0 + z_1}{2\alpha_0 + \alpha_1}$	$(2z_0 + z_1)\frac{(m-1)}{m^2}$
$\log \frac{z_2 + z_0}{\alpha_2 + \alpha_0}$	$(z_2 + z_0)\frac{(m-1)}{m^3}$
$\log \frac{z_2}{\alpha_2}$	$z_2\frac{(m-1)(m^2-m-1)}{m^3}$
$\log \frac{z_1}{\alpha_1}$	$z_1\frac{(m-1)^2}{m^2}$
$\log \frac{z_0}{\alpha_0}$	$z_0\frac{(m-1)(m^2-m-1)}{m^3}$

Table 6

Lod Scores and Probabilities for *m*-allele Model for Grandparent-Grandchild Pairs, When Scheme 2 Is Used

Lod Score	Probability
$\log \frac{z_1}{\alpha_1}$	$z_1(1-\frac{1}{m^2})^2(1-\frac{1}{m})$
$\log \frac{z_0}{\alpha_0}$	$z_0(1-\frac{1}{m^2})^2(1-\frac{1}{m})$

Uncle(aunt)-Nephew(niece) Pairs

As a model, consider individuals 5 and 10 in the pedigree in figure 1. To determine all possible lod score outcomes for each mating type (for individuals 1 and 2), one must include individuals 7 and 6 as well. For each mating type, all possible outcome genotypes for individuals 5, 7, and 10 were enumerated. The probability of each such genotype configuration when the uncle and niece share one allele ibd and zero alleles ibd, respectively, was calculated. These probabilities were then weighted by z_1 and $z0$, respectively, and were added to obtain the probability of that outcome under the alternative hypothesis (linkage); the same was done using α_1 and α_0, respectively, to obtain the probability under the null hypothesis (no linkage). The log ratio of these values is the lod score for that observation. The total probability for each distinct lod score was obtained by summation. In fact, there are only four distinct lod score outcomes, as shown in table 7.

For example, consider mating type (2). The probability that the uncle (individual 5) is A_iA_i and that the mother (individual 7) is A_iA_i and that the niece (individual 10) is A_iX (X being any allele) is $\frac{3}{8}$ if individuals 5 and 10 share one allele ibd and is $\frac{1}{8}$ if they share zero alleles ibd. Hence, the lod score for this observation is $\log[(\frac{3}{8}z_1 + \frac{1}{8}z_0)/(\frac{3}{8}\alpha_1 + \frac{1}{8}\alpha_0)] = \log[(3z_1 + z_0)/(3\alpha_1 + \alpha_0)]$ with probability $\frac{3}{8}z_1 + \frac{1}{8}z_0$. If the uncle is genotype A_iA_j and the mother is A_iA_i and the niece is A_iX, then the probability when the uncle and niece share one allele ibd is

Table 7

Lod Scores and Probabilities for Uncle/Nephew Pairs, When Scheme 2 Is Used

Mating type	Frequency	Lod Score	Probability
(1) $A_iA_i \times A_iA_i$	p_i^4	0	1
(2) $A_iA_i \times A_iA_j$	$4p_i^3p_j$	$\log \frac{3z_1 + z_0}{3\alpha_1 + \alpha_0}$	$\frac{1}{8}(3z_1 + z_0)(2 - Q)$
		$\log \frac{3z_0 + z_1}{3\alpha_0 + \alpha_1}$	$\frac{1}{8}(3z_0 + z_1)(2 - Q)$
		$\log \frac{z_1}{\alpha_1}$	$\frac{1}{4}z_1Q$
		$\log \frac{z_0}{\alpha_0}$	$\frac{1}{4}z_0Q$
(3) $A_iA_i \times A_jA_j$	$2p_i^2p_j^2$	0	1
(4) $A_iA_i \times A_jA_k$	$4p_i^2p_jp_k$	$\log \frac{z_1}{\alpha_1}$	$\frac{1}{2}z_1Q$
		$\log \frac{z_0}{\alpha_0}$	$\frac{1}{2}z_0Q$
		$\log \frac{3z_1 + z_0}{3\alpha_1 + \alpha_0}$	$\frac{1}{8}(3z_1 + z_0)(1 - Q)$
		$\log \frac{3z_0 + z_1}{3\alpha_0 + \alpha_1}$	$\frac{1}{8}(3z_0 + z_1)(1 - Q)$
(5) $A_iA_j \times A_iA_j$	$4p_i^2p_j^2$	$\log \frac{z_1}{\alpha_1}$	$\frac{1}{4}z_1(1 + Q)$
		$\log \frac{z_0}{\alpha_0}$	$\frac{1}{4}z_0(1 + Q)$
(6) $A_iA_j \times A_iA_k$	$8p_i^2p_jp_k$	$\log \frac{z_1}{\alpha_1}$	$\frac{1}{8}z_1(4 + 3Q)$
		$\log \frac{z_0}{\alpha_0}$	$\frac{1}{8}z_0(4 + 3Q)$
(7) $A_iA_j \times A_kA_l$	$8p_ip_jp_kp_l$	$\log \frac{z_1}{\alpha_1}$	$\frac{1}{2}z_1(1 + Q)$
		$\log \frac{z_0}{\alpha_0}$	$\frac{1}{2}z_0(1 + Q)$

⅛; the probability when they share zero alleles ibd is ⅜. Hence, the lod score for this observation is $\log[(z_1 + 3z_0)/(\alpha_1 + 3\alpha_0)]$ with probability $\frac{1}{8}z_1 + \frac{3}{8}z_0$. When the uncle is A_iA_i and the mother is A_iA_j and the niece inherits A_i from the mother, the probability is ⅛ when individuals 5 and 10 share one or zero alleles ibd; hence, the lod score is 0. Similarly, when the uncle is A_iA_j and the mother is A_iA_j and the niece inherits A_i, the probability is ⅛ when individuals 5 and 10 share one or zero alleles ibd; again the lod score is 0. When the uncle and mother are both A_iA_j and the niece is known to have inherited A_j from the mother, the probability is ¼ when individuals 5 and 10 share one allele ibd and 0 when they share zero alleles ibd, leading to a lod score of $\log(z_1/\alpha_1)$. The allele the niece inherits from the mother is determinable, provided that the father (individual 6) is not also A_iA_j and has not transmitted A_i to his daughter; this latter event has probability $1 - 2p_ip_j \times \frac{1}{2} = 1 - p_ip_j$. I denote this probability by the parameter Q. Hence, the lod score is $\log(z_1/\alpha_1)$ with probability $\frac{1}{4}z_1Q$. When the uncle is A_iA_i and the mother is A_iA_j and the niece is known to have inherited A_j from the mother, the probability is ¼ when individuals 5 and 10 share zero alleles ibd and is 0 when they share one allele ibd. The allele the niece inherits from the mother is determinable with probability Q. Thus, the lod score is $\log(z_0/\alpha_0)$ with probability $\frac{1}{4}z_0Q$. When the uncle is A_iA_j and the mother is A_iA_j and the allele inherited by the niece from her mother cannot be determined (i.e., the father is also A_iA_j and transmits the opposite allele from the mother—with probability p_ip_j), the probability of the uncle and mother's genotypes when individuals 5 and 10 share one allele ibd is ⅜; the probability when they share zero alleles ibd is ⅛. Hence, in this case, the lod score is $\log[(3z_1 + z_0)/(3\alpha_1 + \alpha_0)]$ with probability $(\frac{3}{8}z_1 + \frac{1}{8}z_0)(1-Q)$. Similarly, when the father is A_iA_i and the mother is A_iA_j and the niece's inheritance cannot be determined, the lod score is $\log[(3z_0 + z_1)/(3\alpha_0 + \alpha_1)]$ with probability $\frac{1}{8}(3z_0 + z_1)(1-Q)$. Therefore, the total probability of the lod score $\log[(3z_1 + z_0)/(3\alpha_1 + \alpha_0)]$ is $\frac{1}{8}(3z_1 + z_0)(2-Q)$, and for the lod score $\log[(3z_0 + z_1)/(3\alpha_0 + \alpha_1)]$ it is $\frac{1}{8}(3z_0 + z_1)(2-Q)$. The other mating types can be analyzed in a similar fashion, and the results are given in table 7. The parameter Q given in the formulas for the other mating types may involve different alleles. However, when all allele frequencies are equal to $1/m$, $Q = 1 - m^{-2}$ for any combination of alleles. Summing over all possible allelic combinations for nonzero lod scores gives the total probabilities given in table 8. These are the formulas used to calculate the EMLSs and power to detect linkage for an uncle(aunt)-nephew(niece) pair.

Table 8

Lod Scores and Probabilities for *m*-allele Model for Uncle-Nephew Paris, When Scheme 2 Is Used

Lod Score	Probability
$\log\frac{3z_1 + z_0}{3\alpha_1 + \alpha_0}$	$\frac{1}{4}(3z_1 + z_0)\frac{(m-1)}{m^4}(1 + 2m)$
$\log\frac{3z_0 + z_1}{3\alpha_0 + \alpha_1}$	$\frac{1}{4}(3z_0 + z_1)\frac{(m-1)}{m^4}(1 + 2m)$
$\log\frac{z_1}{\alpha_1}$	$z_1\frac{(m-1)}{m^5}(m^4 - \frac{1}{2}m^3 - \frac{3}{2}m^2 + \frac{1}{2})$
$\log\frac{z_0}{\alpha_0}$	$z_0\frac{(m-1)}{m^5}(m^4 - \frac{1}{2}m^3 - \frac{3}{2}m^2 + \frac{1}{2})$

First-Cousin Pairs

The logic for generating the possible lod scores and their probabilities for first cousins is the same as that used for the other relatives described above, although the actual calculations are somewhat more intricate. In this case, there are five distinct nonzero lod score outcomes. The possible outcomes and their probabilities for each mating type are given in table 9. The total probabilities obtained by summing over all mating types, when m alleles with equal frequency are assumed, are given in table 10. These probabilities are used to generate the EMLSs and power to detect linkage for first-cousin pairs.

Simulations

Simulations were performed to determine the power to detect linkage for the different types of relatives as a function of the polymorphic content of the marker locus by using schemes 1 and 2. Four different values of recurrence risk for first-degree relatives (λ_0) were considered: $\lambda_0 = 2, 3, 5$, and 10. Three different values were considered for N: $N = 40, 100$, and 300. Seven different values were considered for m: $m = 2, 3, 4, 5, 8, 10$, and 20; these values of m correspond to PIC values of .38, .59, .70, .77, .86, .89, and .95, respectively. Four types of relative were considered: sibs, grandparent-grandchild (same as half-sibs), uncle-nephew, and first cousins. Therefore, a total of 336 cases were considered. For each case, 5,000 replicate samples of size N were generated; the distribution of lod score outcomes for each sample was obtained using the probabilities given in tables 3, 5, 6, 8, and 10. A simple

Table 9

Lod Scores and Probabilities for Cousin Pairs, When Scheme 2 Is Used

(1) $A_iA_i \times A_iA_i$	p_i^4	0	1
(2) $A_iA_i \times A_iA_j$	$4p_i^3p_j$	$\log \frac{1 - \frac{2}{3}z_0}{1 - \frac{2}{3}\alpha_0}$	$\frac{1}{8}(1 - \frac{2}{3}z_0)[Q^2 + 4Q(1 - Q)]$
		$\log \frac{1 - \frac{4}{9}z_0}{1 - \frac{4}{9}\alpha_0}$	$\frac{3}{8}(1 - \frac{4}{9}z_0)[1 + (1 - Q)^2]$
		$\log \frac{z_1}{\alpha_1}$	$\frac{1}{4}z_1Q^2$
		$\log \frac{z_0}{\alpha_0}$	$\frac{1}{6}z_0Q(Q + 2)$
		$\log \frac{1 + \frac{4}{3}z_0}{1 + \frac{4}{3}\alpha_0}$	$\frac{1}{4}(1 + \frac{4}{3}z_0)(1 - Q)$
(3) $A_iA_i \times A_jA_j$	$2p_i^2p_j^2$	0	1
(4) $A_iA_i \times A_jA_k$	$4p_i^2p_jp_k$	$\log \frac{1 - \frac{2}{3}z_0}{1 - \frac{2}{3}\alpha_0}$	$\frac{1}{2}(1 - \frac{2}{3}z_0)[Q^2 + 2Q(1 - Q)]$
		$\log \frac{1 - \frac{4}{9}z_0}{1 - \frac{4}{9}\alpha_0}$	$\frac{3}{4}(1 - \frac{4}{9}z_0)(1 - Q)^2$
		$\log \frac{z_1}{\alpha_1}$	$\frac{1}{2}z_1Q^2$
		$\log \frac{z_0}{\alpha_0}$	$\frac{1}{6}z_0[5Q^2 + 4Q(1 - Q)]$
		$\log \frac{1 + \frac{4}{3}z_0}{1 + \frac{4}{3}\alpha_0}$	$\frac{1}{4}(1 + \frac{4}{3}z_0)(1 - Q)^2$
(5) $A_iA_j \times A_iA_j$	$4p_i^2p_j^2$	$\log \frac{1 - \frac{2}{3}z_0}{1 - \frac{2}{3}\alpha_0}$	$\frac{1}{4}(1 - \frac{2}{3}z_0)(Q + 1)^2$
		$\log \frac{z_0}{\alpha_0}$	$\frac{1}{6}z_0(Q + 1)^2$
(6) $A_iA_j \times A_iA_k$	$8p_i^2p_jp_k$	$\log \frac{1 - \frac{2}{3}z_0}{1 - \frac{2}{3}\alpha_0}$	$\frac{1}{8}(1 - \frac{2}{3}z_0)(4 + 6Q - 7Q^2)$
		$\log \frac{1 - z_0}{1 - \alpha_0}$	$\frac{5}{8}(1 - z_0)Q^2$
		$\log \frac{z_0}{\alpha_0}$	$\frac{1}{24}z_0(Q^2 + 12Q + 8)$
(7) $A_iA_j \times A_kA_l$	$8p_ip_jp_kp_l$	$\log \frac{1 - \frac{2}{3}z_0}{1 - \frac{2}{3}\alpha_0}$	$\frac{1}{2}(1 - \frac{2}{3}z_0)(1 - Q)(1 + 3Q)$
		$\log \frac{1 - z_0}{1 - \alpha_0}$	$(1 - z_0)Q^2$
		$\log \frac{z_0}{\alpha_0}$	$\frac{1}{3}z_0(2Q + 1)$

maximization routine was then used to calculate the MLS and maximizing values z_i. The EMLS was determined by calculating the average MLS for the 5,000 samples, and expected values for z_i were obtained in the same fashion. The power to detect linkage was calculated as the proportion of the 5,000 samples with MLS > 3.0.

Results

The EMLS values for 100 relative pairs when a PIC value of 1 and no recombination are assumed were calculated, as a function of λ_0, for first- (sibs), second-, and third-degree relatives. The results are given in figure 2. For values of $\lambda_0 \leqslant 3$, the three types of relatives give

Table 10

Lod Scores and Probabilities for *m*-allele Model for Cousin Pairs, When Scheme 2 Is Used

Lod Score	Probability
$\log \frac{1 - \frac{2}{3}z_0}{1 - \frac{2}{3}\alpha_0}$	$(1 - \frac{2}{3}z_0)\frac{1}{2} \frac{(m-1)}{m^7} (5m^5 - m^4 - 12m^3 + 3m^2 + 6m - 2)$
$\log \frac{1 - \frac{4}{9}z_0}{1 - \frac{4}{9}\alpha_0}$	$(1 - \frac{4}{9}z_0)\frac{3}{2} \frac{(m-1)}{m^7}(m^4 + m - 1)$
$\log \frac{1 - z_0}{1 - \alpha_0}$	$(1 - z_0) \frac{(m-1)}{m^2}(1 - \frac{1}{m^2})^2(m - \frac{3}{2})$
$\log \frac{z_0}{\alpha_0}$	$z_0 \frac{(m-1)}{m^6} \frac{1}{3}(3m^5 + \frac{1}{2}m^4 - 5m^3 - 3m^2 + 2m + \frac{3}{2})$
$\log \frac{1 + \frac{4}{3}z_0}{1 + \frac{4}{3}\alpha_0}$	$(1 + \frac{4}{3}z_0) \frac{(m-1)}{m^7} (m^2 + \frac{1}{2}m - 1)$

comparable EMLS values, although values are greater for the closer relatives (sibs) than for the more distant relatives (cousins). When $\lambda_0 > 3$, the opposite is true: EMLS increases with degree of relationship. For example, at a λ_0 value of 10, the EMLS is 1.5 times greater in third-degree than in second-degree relatives and is 2.5 times greater than in sibs.

Examination of the EMLS values under varying conditions of λ_0 and N revealed that the percentage of the maximum EMLS (corresponding to a PIC value of 1) was reasonably independent of the conditions and only a function of the number of marker alleles (PIC value). This was true both for scheme 1 (pairs only) and for scheme 2 (additional relatives). In figure 3 I have plotted the percentage of the maximal EMLS (PIC = 1), as a function of the PIC value of the marker locus, for first-, second-, and third-degree relatives. The solid lines represent scheme 2, and the broken lines represent scheme 1. For scheme 2, the curves are quite similar, although that for sibs is slightly higher than those for other relatives, followed by grandparent-grandchild (and half-sibs), followed by uncle-nephew and first cousins, who gave similar results. For scheme 1, the line for sibs is substantially above those for other relatives, particularly at moderate PIC values.

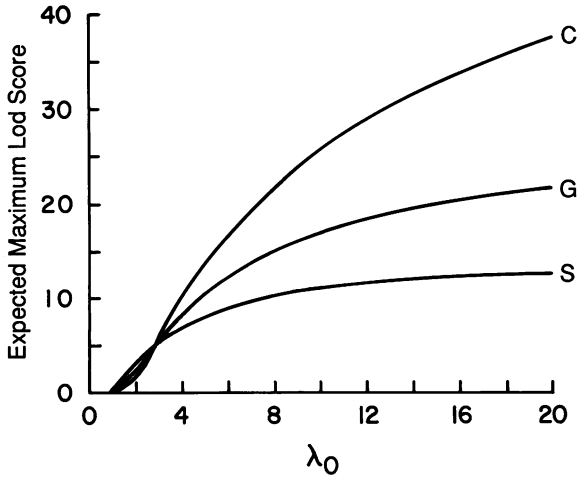

Figure 2 EMLSs as a function of λ_0 for 100 pairs of sibs (S), grandparents-grandchildren (G), and first cousins (C).

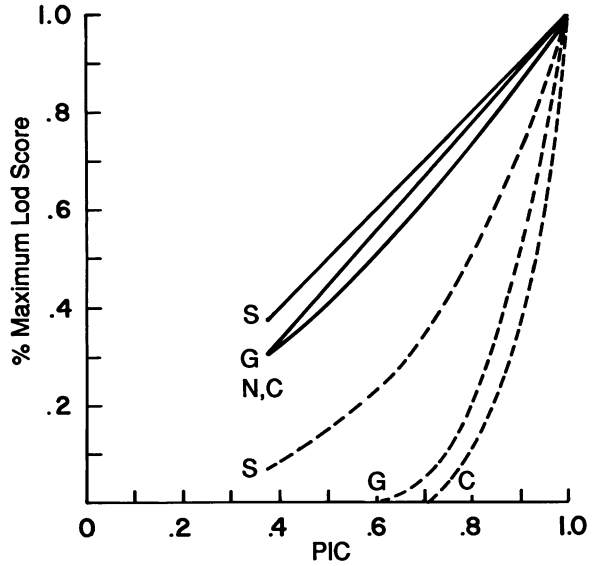

Figure 3 Percent of maximum possible EMLS as a function of PIC for sibs (S), grandparents-grandchildren (G), uncles-nephews (N), and first cousins (C). Dotted line denotes use of scheme 1; solid line denotes use of scheme 2.

However, the major implication of figure 3 is the dramatic loss of information when only relative pairs are used (scheme 1), without other relatives. This is especially true for second- and third-degree relatives but is also true for sibs, although to a lesser degree. For example, to obtain 50% of the maximum possible EMLS requires a PIC value above .9 for second- and third-degree relatives and above .8 for sibs, when scheme 1 is used. When scheme 2 is used, the corresponding PIC value is .5 for sibs and .6 for the other relatives. Because the majority of RFLPs to be used in linkage studies are likely to have PIC values below .8, in general it is clearly preferable to use scheme 2.

Figures 2 and 3 can be combined to give EMLS values for different relatives, as a function of PIC for different values of λ_0 and N. For example, figure 4 gives the corresponding curves for $\lambda_0 = 3$ and $N = 100$. In this case, the curves for scheme 2 are similar for the different types of relatives. Again, scheme 1 is only practical for PIC values greater than .8 for sibs or greater than .9 for other relatives. Figure 5 gives the curves for the case $\lambda_0 = 10$ and $N = 40$. Here the EMLS values are

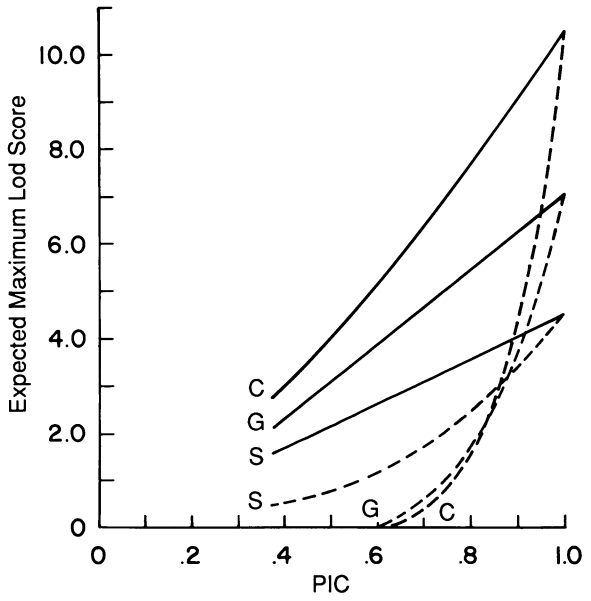

Figure 5 EMLS as a function of PIC for 40 pairs of sibs (S), grandparents-grandchildren (G), and first cousins (C), when $\lambda_0 = 10$ is assumed. Dotted line denotes use of scheme 1; solid line denotes use of scheme 2.

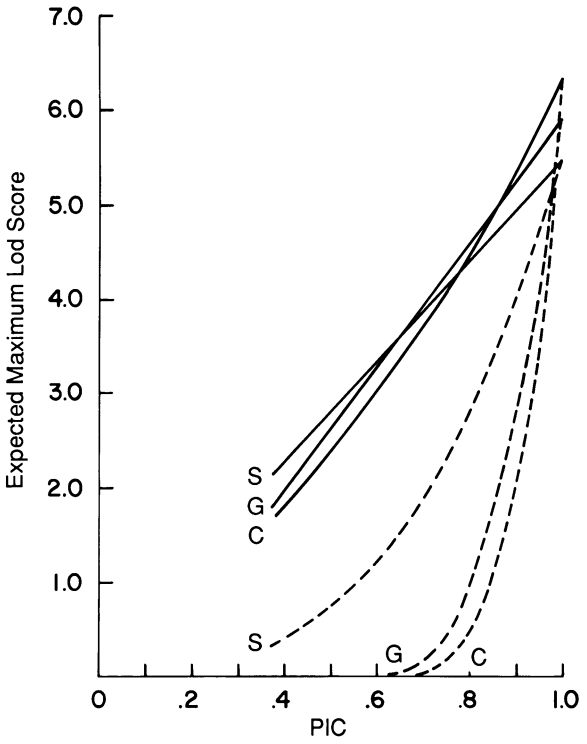

Figure 4 EMLS as a function of PIC for 100 pairs of sibs (S), grandparents-grandchildren (G), and first cousins (C), assuming that $\lambda_0 = 3$. Dotted line denotes use of scheme 1; solid line denotes use of scheme 2.

greatest for cousins, followed by second-degree relatives and then by sibs for scheme 2. The effect of PIC on power (probability of MLS > 3.0) was also examined for these two cases. Figure 6 shows the results for $\lambda_0 = 3$ and $N = 100$. Here the power curves are quite steep as a function of PIC, and the three degrees of relatives are quite similar. Figure 7 gives the results for $\lambda_0 = 10$ and $N = 40$. As expected, power is greatest for cousins, followed by the second- and first-degree relatives. Again, the effect of PIC on power is quite dramatic.

Figures 4 and 5 indicate that the optimal strategy in terms of type of relative to use in a linkage study depends on underlying conditions. If the λ_0 value is large, the more distant relatives will be of greater value, provided that the PIC value for the linked marker is quite high (>.9) or if additional relatives can be typed. For smaller values of λ_0, different relatives give comparable levels of information when additional relatives are typed. In general, however, if additional relatives are not typed and the PIC values are not high (say, <.8), siblings are the preferred relationship. One also needs to keep in mind that recombination affects the various relative types to a different extent, so that, for example, grandparent-grandchild pairs may be the most informative over a broad range of circumstances.

One may also wish to consider the number of additional relatives that need to be typed for each relation-

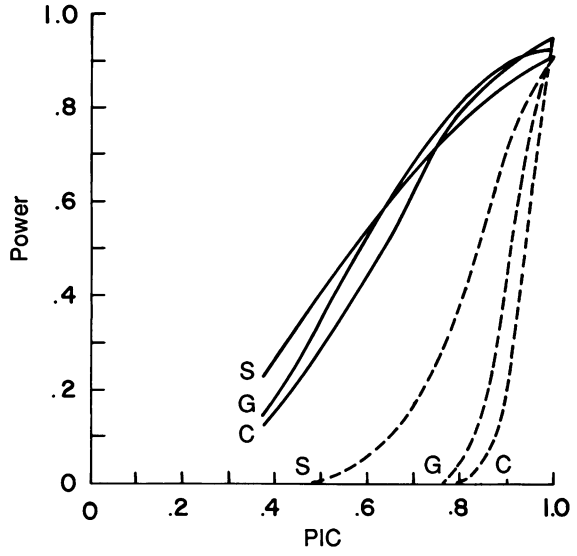

Figure 6 Power to detect linkage as a function of PIC for 100 pairs of sibs (S), grandparents-grandchildren (G), and first cousins (C), when $\lambda_0 = 3$ is assumed. Dotted line denotes use of scheme 1; solid line denotes use of scheme 2.

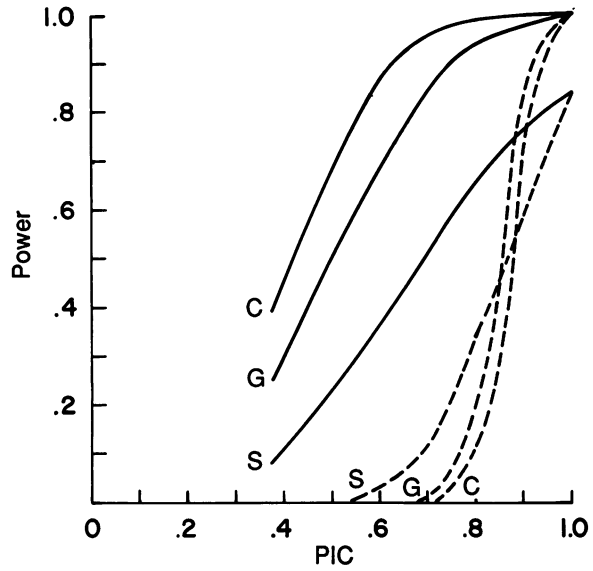

Figure 7 Power to detect linkage as a function of PIC for 40 pairs of sibs (S), grandparents-granchildren (G), and first cousins (C), when $\lambda_0 = 10$ is assumed. Dotted line denotes use of scheme 1; solid line denotes use of scheme 2.

ship. For sibs this number is two; for grandparent-grandchild and half-sib pairs, the number is three; for uncle-nephew pairs, the number is four; and for first cousins, the number is six. However, for the more distant relatives, one can probably obtain most of the available information with fewer relatives. For example, with first-cousin pairs, the two parents (who are siblings) and the shared grandparents (total of four relatives) would appear to be most critical, with the unrelated parents relatively less important.

Discussion

Several different statistics for detecting linkage by using affected sib pairs have been proposed. Three such tests were considered by Blackwelder and Elston (1985), for a fully informative marker similar to HLA. When the marker is less than fully polymorphic, a number of different approaches have been suggested. For example, Thomson (1986) suggested using only affected sib pairs when the parental mating type is fully informative, so that ibd can be determined unequivocally. Alternatively, Lange (1986) and Weeks and Lange (1988) suggest replacing ibd by ibs.

Here, an MLS approach based on likelihood is recommended. Presumably, such an approach makes maximal use of the available information with respect to allelic ibd for affected relative pairs. Although differences among methods when marker information is available for a pair only is probably small, when additional relatives are typed, they can significantly impact the information on allele sharing for the affected pair, especially when a marker is not highly polymorphic. This fact is clearly illustrated in figure 3. Scheme 1, which uses marker information for the affected pair only, yields considerably lower expected lod scores than does scheme 2, in which additional relatives are typed. For example, for affected sib pairs, the EMLS for scheme 2 is at least double that for scheme 1, up to a PIC value of .7. Since scheme 2 involves the typing of four individuals (the sib pair and parents)—as opposed to typing of two individuals for scheme 1— scheme 2 holds an advantage over scheme 1 in terms of expected lod score per person typed, up to a PIC value of .7. Since for most of the markers yet identified the PIC value is less than .7, typing of parents of affected sib pairs is recommended. Similarly, scheme 2 for half-sib or grandparent-grandchild pairs requires typing three additional relatives, i.e., 2.5 times as many total relatives as are required by scheme 1; however, the expected lod score for scheme 2 is at least 2.5 times that for scheme 1 for PIC values up to .85; the same PIC value applies to first cousins as well. Therefore, in general, it is cost-effective for additional available relatives to be typed to increase the information about allele sharing by the affected pair. Furthermore, if such information does ex-

ist in a sample, it is very wasteful not to make use of it (e.g., as would be the case for ibs methods).

Figures 4–7 illustrate that choice of optimal relative type for affected pairs depends on the value for λ_0 and PIC, as well as on whether additional relatives are included. For smaller λ_0 values (say, <5), sibs are always optimal when scheme 1 is used; the same is true for scheme 2 unless the marker is highly polymorphic. Even when λ_0 is large (say, >10), if scheme 1 is used, siblings offer the greatest power unless the PIC value is very high; however, if additional relatives are typed (scheme 2), distant relatives are advantageous.

When making an assessment by using the λ_0 value for a particular disease, one should take into account the effect of recombination. As described in the preceding paper (e.g., see formula [24] in Risch 1990), for a given recombination fraction θ an equivalent λ_0 can be determined; this value will differ for each type of relative because the effect of recombination differs among relatives. For example, consider $\lambda_0 = 5$ and $\theta = .05$. Then equivalent λ_0 values are, respectively, 2.84 for sibs, 4.00 for grandparent-grandchild, 3.35 for half-sibs, 3.11 for uncle-nephew, and 2.59 for first cousins. Consulting figure 2 gives the following EMLSs (when a 100% polymorphic marker is assumed and 100 relative pairs are used): 5.0 for sibs, 8.0 for grandparent-grandchild, 6.0 for half-sibs, 5.5 for uncle-nephew, and 4.0 for first cousins. Figure 3 indicates that the ranking of these EMLSs will not change for lower PIC values when scheme 2 is used; when scheme 1 is used, however, sibs will be the preferred relative type unless PIC is greater than .8.

In summary, figures 2 and 3 should prove useful for investigators designing a linkage study based on pairs of affected relatives. For a given sample size, λ_0, and recombination fraction θ, the expected lod score can be determined, for both schemes 1 and 2, for the different types of relative pairs. When using a value for λ_0, it is important to remember that this value corresponds to the case of a single locus underlying familial aggregation; if multiple loci contribute, then the relevant value of λ_0 for a given locus may be considerably smaller.

Acknowledgments

The author is deeply indebted to Dr. Timothy Bishop for his generosity of ideas, criticism, and support, which have greatly contributed to this work. He is also grateful to Drs. Catherine Falk, Ken Kidd, Jurg Ott, and Stephanie Sherman for many helpful discussions. Dr. Michael Boehnke suggested numerous improvements and corrections. Financial support was provided by NIGMS grant GM 39812.

References

Blackwelder WC, Elston RC (1985) A comparison of sib-pair linkage tests for disease susceptibility loci. Genet Epidemiol 2:85–97

Botstein D, White RL, Skolnick M, Davis RW (1980) Construction of a genetic linkage map in man using restriction fragment length polymorphisms. Am J Hum Genet 32: 314–331

Dempster AP, Laird NM, Rubin DB (1977) Maximum likelihood from incomplete data via the EM algorithm. J R Stat Soc, Ser B 39:1–22

Lange K (1986) The affected sib-pair method using identity by state relations. Am J Hum Genet 39:148–150

Morton NE (1955) Sequential tests for the detection of linkage. Am J Hum Genet 7:277–318

Risch N (1989) Genetics of IDDM: evidence for complex inheritance with HLA. Genet Epidemiol 6:143–148

——— (1990) Linkage strategies for genetically complex traits. II. The power of affected relative pairs. Am J Hum Genet 46:229–241

Thomson G (1986) Determining the mode of inheritance of RFLP-associated diseases using the affected sib-pair method. Am J Hum Genet 39:207–221

Weeks DE, Lange K (1988) The affected-pedigree-member method of linkage analysis. Am J Hum Genet 42:315–326