Am. J. Hum. Genet. 46:229–241, 1990

Linkage Strategies for Genetically Complex Traits. II. The Power 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 power to detect disease-susceptibility loci through linkage analysis using pairs of affected relatives and affected-unaffected pairs is examined. Allelic identity by descent (ibd) for a completely polymorphic marker for sibling, uncle-nephew, grandparent-grandchild, half-sib, and first-cousin pairs is considered. Affected-unaffected pairs generally represent a poor strategy. For single-locus models, ibd depends on λ_R, the risk ratio for type R relatives compared with population prevalence, and the recombination fraction θ. The ibd for grandparent-grandchild pairs is least affected by recombination, followed by sibs, half-sib, uncle-nephew, and first-cousin pairs. For diseases with large λ values and for small θ values, distant relatives offer greater power. For larger θ values, grandparent-grandchild pairs are best; for small λ values, sibs are best. Additive and multiplicative multilocus models are considered. For the multiplicative model, the same formulas as in the single-locus model apply, except that λ_{iR} (for the *i*th contributing locus) is substituted for λ_R. For the additive model, the deviation from null expectation for ibd is divided among all contributing loci. Compared with the multiplicative model, for an additive model there is usually greater advantage in distant relationships. Multipoint analysis using linked marker loci for affected relative pairs is described. Simultaneous use of multiple markers diminishes the effect of recombination and allows for localization of the disease-susceptibility locus.

Introduction

While the rationale for and execution of genetic linkage analysis of diseases with a clear Mendelian pattern of inheritance is straightforward (if arduous), such analysis of complex, non-Mendelian familial disease remains difficult. Without clear evidence of a major locus effect and its characterization, conventional linkage analysis, which requires specification of mode of inheritance, is ill-founded. Therefore, robust methods of detecting "disease-susceptibility loci," which may have major or minor effects, have been suggested. These methods generally focus on small constellations of affected family members. For a marker locus, demonstration of genotypic concordance among affected relatives in excess of what would be expected is taken as evidence for the existence of a disease-susceptibility locus linked to the marker. The ability to detect such a locus depends on the contribution it makes to the total variation in the trait.

The simplest of such methods utilizes pairs of affected relatives, most commonly siblings. It turns out that the power of such "pair" methods to detect linkage of a disease-susceptibility locus depends only on its contribution to the increase in risk to relatives (λ_R), compared with population prevalence. In the present paper, I discuss the power to detect linkage by using different relative pair designs as a function of the genetic model (single locus, multiple locus), how multiple loci may interact, the observed λ_R's, and the recombination fraction θ between the marker and the disease locus. Marker loci are assumed to be 100% polymorphic; the effect of reduced polymorphism is considered in the following paper.

The notation to be used below is identical to that defined in the preceding paper. K denotes the popula-

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.

0002-9297/90/4602-0003$02.00

tion prevalence; K_R denotes the recurrence risk to a type R relative of an affected individual; and $\lambda_R = K_R/K$. Relationship subscripts are as follows: O = offspring; S = siblings, M = MZ twins; G = grandchildren; H = half-sibs; N = nieces or nephews; and C = first cousins. In addition, α_{Ri} denotes the probability that a pair of type R relatives share i alleles at a random locus identical by descent (ibd).

Probability of Alleles Shared ibd

First, consider the case in which no recombination occurs between the trait locus T and the marker locus M. It was shown by Risch (1987) that the probability z_{Ri} that an affected relative pair will share i alleles ibd at the marker can be written in terms of λ values; that is,

$$\begin{aligned} z_{R0} &= P(\text{ibd}=0 \mid 2 \text{ relatives affected}) \\ &= \frac{P(\text{ibd}=0)\, P(2 \text{ relatives affected} \mid \text{ibd}=0)}{P(2 \text{ relatives affected})} \\ &= \frac{\alpha_{R0}(K \times K)}{K \times K_R} \\ &= \alpha_{R0}/\lambda_R \,. \end{aligned} \tag{1}$$

Similarly,

$$z_{R1} = \alpha_{R1}\lambda_O/\lambda_R \tag{2}$$

and

$$z_{R2} = \alpha_{R2}\lambda_M/\lambda_R \,. \tag{3}$$

Along with the null expectation α_{Ri}, an important parameter for determining power to detect linkage by using affected relative pairs is the deviation δ_{Ri} that the probability of sharing i alleles ibd has from its null expectation; that is, $\delta_{Ri} = z_{Ri} - \alpha_{Ri}$. From equations (1)–(3), the following equations can be derived:

$$\delta_{R0} = -\frac{\alpha_{R0}}{\lambda_R}(\lambda_R - 1) \,; \tag{4}$$

$$\delta_{R1} = \frac{\alpha_{R1}}{\lambda_R}(\lambda_O - \lambda_R) \,; \tag{5}$$

$$\delta_{R2} = \frac{\alpha_{R2}}{\lambda_R}(\lambda_M - \lambda_R) \,; \tag{6}$$

Affected-Unaffected Relative Pairs

Formulas analogous to those derived above for affected relative pairs can also be obtained for an affected individual with an unaffected relative. In this case, let y_{Ri} be the probability that an affected-unaffected type R relative pair share i alleles ibd. Then

$$\begin{aligned} y_{R0} &= P(\text{ibd}=0) \mid 1 \text{ relative affected, 1 unaffected}) \\ &= \frac{P(\text{ibd}=0)\, P(1 \text{ affected, 1 unaffected} \mid \text{ibd}=0)}{P(1 \text{ affected, 1 unaffected})} \\ &= \frac{\alpha_{R0}K(1-K)}{K(1-K_R)} \\ &= \alpha_{R0}\frac{(1-K)}{(1-K_R)} \,; \end{aligned}$$

similarly,

$$y_{R1} = \alpha_{R1}\frac{(1-K_o)}{(1-K_R)}$$

and

$$y_{R2} = \alpha_{R2}\frac{(1-K_M)}{(1-K_R)} \,.$$

Analogous to the case of affected relative pairs, define $\varepsilon_{Ri} = y_{Ri} - \alpha_{Ri}$ for affected-unaffected pairs. Then

$$\varepsilon_{R0} = \alpha_{R0}\frac{K}{1-K_R}(\lambda_R - 1) \,,$$

$$\varepsilon_{R1} = -\alpha_{R1}\frac{K}{1-K_R}(\lambda_O - \lambda_R) \,,$$

$$\varepsilon_{R2} = -\alpha_{R2}\frac{K}{1-K_R}(\lambda_M - \lambda_R) \,.$$

Notice the similarity between the formulas for ε_{Ri} and those for δ_{Ri}. Specifically, $\varepsilon_{Ri} = -K_R/(1-K_R)\,\delta_{Ri}$. As expected, for affected-unaffected pairs the sign of the deviation from the null hypothesis is opposite that for affected pairs. However, also note that the ratio of the magnitudes of the deviations for affected-unaffected pairs compared with affected pairs is $K_R/(1-K_R)$, which is less than 1 provided that $K_R < .5$. In this case, α_{Ri} is the same for affected pairs and affected-unaffected pairs, so the power will depend on δ_{Ri}. Therefore, when the recurrence risk K_R for a given type of relative is low, as is usually the case for more distant relatives and as is often the case even for sibs, there will be much less power in using affected-unaffected pairs

than there is in using affected pairs. For example, even if $K_R = .25$, $K_R/(1-K_R) = 1/3$, so the deviation for affected-unaffected pairs is greatly diminished. Because values for K_R are always less than 50% and usually less than 25% (except for high-penetrance autosomal dominant diseases), affected-unaffected pairs will generally not be efficient as a study design for detecting linkage.

Allele Sharing, Including the Effects of Recombination

Formulas (1)–(6) were derived assuming no recombination between the trait and marker loci. These formulas can be generalized to allow for the effect of recombination. It is assumed that recombination fractions in males and females are equal and can be summarized by the single value θ. In order to calculate the appropriate formulas, the conditional probabilities for a relative pair to share i alleles at the trait locus T given that they share j alleles at the marker locus M ($i,j = 1, \ldots, 3$) is required. These probabilities can be calculated in a straightforward fashion for the different types of relatives. The matrix of probabilities for sibs was given by Haseman and Elston (1972) and Suarez et al. (1978); for other relative types, formulas have been given by Campbell and Elston (1971) and Thompson (1986). Here the conditional probability matrices for sibs and for grandparent-grandchild, uncle-nephew, half-sib, and first-cousin relationships are given in table 1. The parameter $\Psi = \theta^2 + (1-\theta)^2$ (as defined by Haseman and Elston [1972]). It is important to note that for the three types of second-degree relatives the matrices are all different; this fact leads to differences in power to detect linkage by using these relative types when $\theta > 0$.

Let the random variables ibd_m and ibd_t represent ibd at the marker and trait loci, respectively. Then, for an affected relative pair of type R, z_{Ri} can be calculated as follows:

$$z_{Ri} = \frac{P(ibd_m = i)}{P(2 \text{ relatives affected})} P(2 \text{ relatives affected} | ibd_m = i)$$

$$= \frac{\alpha_{Ri}}{K \times K_R} \times \sum_{j=0}^{2} P(2 \text{ relatives affected} | ibd_t = j) \times P(ibd_t = j | ibd_m = i) . \quad (7)$$

Affected Sib Pairs

Applying formula (7) to an affected sib pair and using table 1 and the formula $\lambda_M = 4\lambda_S - 2\lambda_O - 1$ (formula [5] in Risch 1990*a*) leads to

Table 1

Conditional IBD Matrix for Linked-Trait Locus, Given Marker Locus for Various Relatives

	Trait IBD		
Relationship and Marker IBD	2	1	0
Sibs:			
2	ψ^2	$2\psi(1-\psi)$	$(1-\psi)^2$
1	$\psi(1-\psi)$	$\psi^2 + (1-\psi)^2$	$\psi(1-\psi)$
0	$(1-\psi)^2$	$2\psi(1-\psi)$	ψ^2
Grandparent-grandchild:			
1		$1 - \theta$	θ
0		θ	$1 - \theta$
Uncle-nephew:			
1		$\psi(1-\theta) + \frac{1}{2}\theta$	$1 - \frac{1}{2}\theta - \psi(1-\theta)$
0		$1 - \frac{1}{2}\theta - \psi(1-\theta)$	$\psi(1-\theta) + \frac{1}{2}\theta$
Half-sibs:			
1		ψ	$1 - \psi$
0		$1 - \psi$	ψ
First cousins:			
1		$\psi(1-\theta)^2 + \frac{1}{2}\theta^2$	$1 - \frac{1}{2}\theta^2 - \psi(1-\theta)^2$
0		$\frac{1}{3}\left[1 - \frac{1}{2}\theta^2 - \psi(1-\theta)^2\right]$	$\frac{1}{3}\left[2 + \frac{1}{2}\theta^2 + \psi(1-\theta)^2\right]$

$$z_{S0} = \frac{1/4}{K \times K_S}[K^2\psi^2 + KK_O 2\psi(1-\psi) + KK_M(1-\psi)^2]$$
$$= \frac{1}{4} - \frac{1/4}{\lambda_S}(2\psi - 1)[(\lambda_S - 1) + 2(1-\psi)(\lambda_S - \lambda_O)] . \quad (8)$$

Similarly,

$$z_{S1} = \frac{1/2}{K \times K_S}\{K^2\psi(1-\psi) + KK_O[\psi^2 + (1-\psi)^2] + KK_M\psi(1-\psi)\}$$
$$= \frac{1}{2} - \frac{1}{2}(2\psi - 1)^2 \frac{1}{\lambda_S}(\lambda_S - \lambda_O) \quad (9)$$

and

$$z_{S2} = \frac{1/4}{K \times K_S}[K^2(1-\psi)^2 + KK_O 2\psi(1-\psi) + KK_M\psi^2]$$
$$= \frac{1}{4} + \frac{1/4}{\lambda_S}(2\psi - 1)[(\lambda_S - 1) + 2\psi(\lambda_S - \lambda_O)] . \quad (10)$$

Formulas similar to these were also derived by Suarez et al. (1978), although those authors used a parmeterization involving the genetic variance components V_A and V_D and the population prevalence K.

If θ is near zero, then formulas (8) and (10) reduce, respectively, to

$$z_{S0} = \frac{1}{4} - \frac{1}{4}(2\psi - 1)\frac{1}{\lambda_S}(\lambda_S - 1) \quad (11)$$

and

$$z_{S2} = \frac{1}{4} + \frac{1}{4}\psi(2\psi - 1)\frac{1}{\lambda_S}(\lambda_M - \lambda_S) . \quad (12)$$

When $\lambda_S = \lambda_O$ (e.g., $V_D = 0$), formula (11) also holds, but formula (10) becomes

$$z_{S2} = \frac{1}{4} + \frac{1}{4}(2\psi - 1)\frac{1}{\lambda_S}(\lambda_S - 1) . \quad (13)$$

When λ_O is small compared with λ_S, formula (10) becomes

$$z_{S2} = \frac{1}{4} + \frac{1}{4}\frac{1}{\lambda_S}(2\psi - 1)(1 + 2\psi)(\lambda_S - 1) . \quad (14)$$

When θ is near zero, or when $\lambda_S = \lambda_O$, or when $\lambda_S >> \lambda_O$, formulas (9) and (11)–(14) resemble formulas (4)–(6), which assume that $\theta = 0$. In this case, however, the deviations (δ_{Ri}) from expected are multiplied by a function of θ. For example, for z_{S0}, δ_{S0} is multiplied by $(2\psi - 1)$. This result illustrates directly the confounding between recombination and mode of inheritance parameters in determining z_{S2}, z_{S1} and z_{S0}.

Grandparent-Grandchild Pairs

For a grantparent-grandchild pair, formula (7) and table 1 give

$$z_{G0} = \frac{1/2}{K \times K_G}[K^2(1-\theta) + KK_O\theta]$$
$$= \frac{1}{2} - \frac{1}{2}(1-2\theta)\frac{1}{\lambda_G}(\lambda_G - 1)$$
$$= \frac{1}{2} - \frac{1}{2}(1-2\theta)\frac{1}{(\lambda_O + 1)}(\lambda_O - 1) \quad (15)$$

and

$$z_{G1} = 1 - z_{G0}$$
$$= \frac{1}{2} + \frac{1}{2}(1-2\theta)\frac{1}{\lambda_G}(\lambda_G - 1)$$
$$= \frac{1}{2} + \frac{1}{2}(1-2\theta)\frac{1}{(\lambda_O + 1)}(\lambda_O + 1) , \quad (16)$$

if it is assumed that $\lambda_G - 1 = \frac{1}{2}(\lambda_O - 1)$.

Uncle (Aunt)-Nephew (Niece)

For this type of pair, formula (7) and table 1 give

$$z_{N0} = \frac{1/2}{K \times K_N}\left\{K^2\left[1 - \frac{1}{2}\theta - 2\theta(1-\theta)^2\right] + KK_O\left[\frac{1}{2}\theta + 2\theta(1-\theta)^2\right]\right\}$$
$$= \frac{1}{2} - \frac{1}{2}(1-\theta)(1-2\theta)^2\frac{1}{\lambda_N}(\lambda_N - 1)$$
$$= \frac{1}{2} - \frac{1}{2}(1-\theta)(1-2\theta)^2\frac{1}{(\lambda_O + 1)}(\lambda_O - 1) \quad (17)$$

and

$$z_{N1} = 1 - z_{N0}$$
$$= \frac{1}{2} + \frac{1}{2}(1-\theta)(1-2\theta)^2\frac{1}{\lambda_N}(\lambda_N - 1)$$

$$= \frac{1}{2} + \frac{1}{2}(1-\theta)(1-2\theta)^2 \frac{1}{(\lambda_O+1)}(\lambda_O-1), \quad (18)$$

if it is assumed that $\lambda_N - 1 = \frac{1}{2}(\lambda_O-1)$.

Half-Sib Pairs

For half-sibs, formula (7) and table 1 give

$$\begin{aligned} z_{H0} &= \frac{\frac{1}{2}}{K \times K_H}[K^2\psi + KK_O(1-\psi)] \\ &= \frac{1}{2} - \frac{1}{2}(2\psi-1)\frac{1}{\lambda_H}(\lambda_H-1) \\ &= \frac{1}{2} - \frac{1}{2}(2\psi-1)\frac{1}{(\lambda_O+1)}(\lambda_O-1) \quad (19) \end{aligned}$$

and

$$\begin{aligned} z_{H1} &= 1 - z_{H0} \\ &= \frac{1}{2} + \frac{1}{2}(2\psi-1)\frac{1}{\lambda_H}(\lambda_H-1) \\ &= \frac{1}{2} + \frac{1}{2}(2\psi-1)\frac{1}{(\lambda_O+1)}(\lambda_O-1), \quad (20) \end{aligned}$$

if it is assumed that $\lambda_H - 1 = \frac{1}{2}(\lambda-1)$.

First-Cousin Pairs

For first cousins, formula (7) and table 1 give

$$\begin{aligned} z_{C0} &= \frac{\frac{3}{4}}{K \times K_C}\left(K^2\left\{\frac{1}{3}\left[2+\frac{1}{2}\theta^2+\psi(1-\theta)^2\right]\right\}\right. \\ &\quad \left. + KK_O\left\{\frac{1}{3}\left[1-\frac{1}{2}\theta^2-\psi(1-\theta)^2\right]\right\}\right) \\ &= \frac{3}{4} - \left[(1-\theta)^4+\theta^2(1-\theta)^2+\frac{1}{2}\theta^2-\frac{1}{4}\right] \times \\ &\quad \frac{1}{\lambda_C}(\lambda_C-1) \\ &= \frac{3}{4} - \left[(1-\theta)^4+\theta^2(1-\theta)^2+\frac{1}{2}\theta^2-\frac{1}{4}\right] \times \\ &\quad \frac{1}{(\lambda_O+3)}(\lambda_O-1) \quad (21) \end{aligned}$$

and

$$\begin{aligned} z_{C1} &= 1 - z_{C0} \\ &= \frac{1}{4} + \left[(1-\theta)^4+\theta^2(1-\theta)^2+\frac{1}{2}\theta^2-\frac{1}{4}\right] \times \\ &\quad \frac{1}{\lambda_C}(\lambda_C-1) \\ &= \frac{1}{4} + \left[(1-\theta)^4+\theta^2(1-\theta)^2+\frac{1}{2}\theta^2-\frac{1}{4}\right] \times \\ &\quad \frac{1}{(\lambda_O+3)}(\lambda_O-1), \quad (22) \end{aligned}$$

if it is assumed that $\lambda_C - 1 = \frac{1}{4}(\lambda_O-1)$.

Power to Detect Linkage

Formulas (8)–(22) give the expected ibd distributions as a function of recombination θ and risk ratios λ. They can therefore be used directly to determine the power to detect linkage for the different types of relative pairs.

For affected sib pairs, formulas (9), (11), and (12) show that, for small values of θ, z_{S0} depends only on λ_S and not on λ_O, whereas z_{S1} and z_{S2} both depend on λ_S and λ_O. Therefore, in general, the power is a function of the two parameters λ_S and λ_O. For a trait which displays a significant dominance variance component (e.g., a rare recessive trait), z_{S1} can be less than $\frac{1}{2}$ and z_{S2} can be greater than $\frac{1}{2}$; in fact, z_{S2} can approach 1. A linkage test based on the observed value of z_{S0} will therefore, under such circumstances, not be as powerful as one based on z_{S2}. However, in the absence of a dominance variance component ($\lambda_S = \lambda_O$), $z_{S1} = \frac{1}{2}$, so $z_{S0} + z_{S2} = \frac{1}{2}$. In this case, a linkage test can be based on either z_{S0} or z_{S2}, with equivalent results.

In fact, most common complex diseases in man show little or no dominance effect, i.e., λ_S and λ_O are similar. This is true, for example, for common cancers, cardiovascular disease, psychiatric disorders, birth defects, and so on. Therefore, I only consider in detail the case of $\lambda_S = \lambda_O$, so that power can be determined in terms of the single parameter λ_S or λ_O.

For other types of relative pairs, there are only two ibd outcomes which depend on the single parameter λ_R. Note that, for a single locus, for second- or third-degree relatives λ_R can be written in terms of λ_O. For example, $\lambda_N = \frac{1}{2}(\lambda_O+1)$ and $\lambda_C = \frac{1}{4}(\lambda_O+3)$ (see the preceding paper [Risch 1990*a*]). Therefore, for every type of relative pair, the power to detect linkage can be calculated in terms of the single parameter λ_O and the recombination fraction θ.

Studies using affected sib pairs have commonly used a significance level of $P = .05$; for example, Blackwelder and Elston (1985) compare the power of three linkage tests for affected sib pairs. The three tests are based on z_{S2}, $z_{S2} + \frac{1}{2}z_{S1}$, and a χ^2 goodness-of-fit test for the

null hypothesis of no linkage. The power of the three tests was comparable, although the second test performed best in many situations. In all cases, a significance level of $P = .05$ was used. This is a less stringent criterion than is used in standard lod-score analysis, namely, a lod score greater than 3. Because, in general, there is no prior hypothesis of linkage for a given marker (except in the case where there is a disease-marker association), I use the more stringent requirement of a lod score of 3 to determine power. Specifically, suppose in a sample of N affected relative pairs n_0 share 0 marker alleles ibd and $N - n_0$ share 1 (or more, for sibs) marker allele ibd. Then I define a maximum lod-score statistic T as follows:

$$T = n_0 \log_{10}\left(\frac{n_0}{N\alpha_{R0}}\right) + (N-n_0)\log_{10}\left(\frac{N-n_0}{N-N\alpha_{R0}}\right). \quad (23)$$

T represents the $\log_{10}$ of the ratio of the maximum likelihood of the observed data to the likelihood under the null hypothesis of no linkage. Evidence for linkage is assumed to be significant when $T > 3$. For sib pairs, $\alpha_{S0} = \frac{1}{4}$; for second-degree relatives (e.g., uncle/nephew), $\alpha_{N0} = \frac{1}{2}$; for third-degree relatives (e.g., first cousins), $\alpha_{C0} = \frac{3}{4}$.

First I consider the power to detect linkage for affected sib pairs as a function of λ_S for different values of N, the total number of pairs. For each value of N, a value W is determined such that if $n_0 \leq W$, then the statistic $T > 3$. For $N = 40$, $W = 1$; for $N = 100$, $W = 10$; for $N = 300$, $W = 48$. For each value of λ_S, z_{S0} is calculated using formula (1); when this value of z_{S0} is used, the power is calculated as the probability of observing $n_0 \leq W$. For $N \leq 60$, the probability is calculated exactly by using the binomial distribution. For $N \geq 80$, the probability is calculated using a normal approximation to the binomial.

The results are given in figure 1. Power to detect linkage for a given value of λ_S can be determined directly from this figure. For example, for $\lambda_S = 2$, a sample size greater than 200 will be required to detect linkage with 80% power. For $\lambda_S = 5$, however, a sample size of 60 will be sufficient to detect linkage with 80% power. This figure represents a *minimum* required sample size because recombination is assumed to be zero, and the marker is assumed to be completely informative (100% polymorphic). The influence of the polymorphic content of the marker is detailed in the following paper.

Equation (11) gives the formula for z_{S0} including recombination; the deviation $\delta_{S0} = -\frac{1}{4}(2\psi-1)(\lambda_S-1)/\lambda_S$. This deviation is equivalent to that corresponding to a value λ_S' with $\theta = 0$, where $(\lambda_S'-1)/\lambda_S' = (2\psi-1)(\lambda_S-1)/\lambda_S$ or, equivalently,

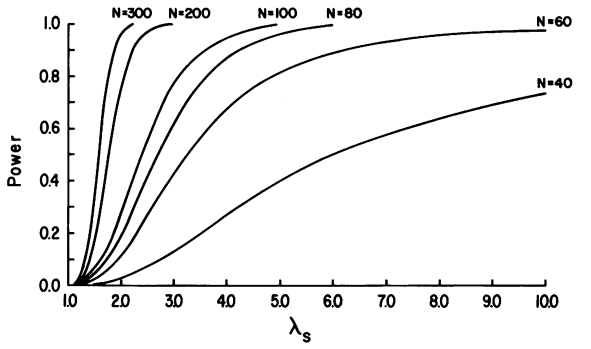

Figure 1 Power to detect linkage as a function of λ_S, by using N affected sib pairs and assuming a fully informative marker and $\theta = 0$.

$$\lambda_S' = \frac{\lambda_S}{1 + 4\theta(1-\theta)(\lambda_S-1)}. \quad (24)$$

Formula (24) can be used in conjunction with figure 1 to determine the power of detecting linkage when $\theta > 0$. For example, consider again $\lambda_S = 2$. When $\theta = .05$, $\lambda_S' = 1.68$ and a sample size of $N = 300$ gives only about 75% power; for $\theta = .10$, $\lambda_S' = 1.47$ and the power for $N = 300$ drops to less than 40%. For $\lambda_S = 5$ and $\theta = .05$, $\lambda_S' = 2.84$ and $N > 100$ is required for 80% power; for $\theta = .10$, $\lambda_S' = 2.05$ and $N > 200$ is required for 80% power.

The effect of recombination is illustrated graphically in figure 2. Three values of N are considered—$N = 300$ (solid lines), $N = 100$ (long dashes), and $N = 40$ (short dashes)—and three values of θ are given—$\theta = 0$, .05, and .10. For small sample sizes ($N = 40$), the effect of recombination can be devastating. Even for $\lambda_S = 10$, with recombination of only 5% the power drops from 74% to 23%. With a moderate sample size ($N = 100$), the effect of recombination on reducing power is serious for $\lambda_S < 4$ but not as serious for $\lambda_S > 4$, provided $\theta \leq .05$. When $\theta = .10$, the power is only 10% even for $\lambda_S = 10$. For a large sample ($N = 300$), recombination up to 5% has only a minor effect on reducing power, but with $\theta = .10$ the effect is serious for $\lambda_S \leq 2$.

The power for other types of relative pairs can be determined in a similar fashion by using formulas (15)–(22). As in the case of sibs, for each type of relationship a λ_O' can be defined by assuming that $\theta =$

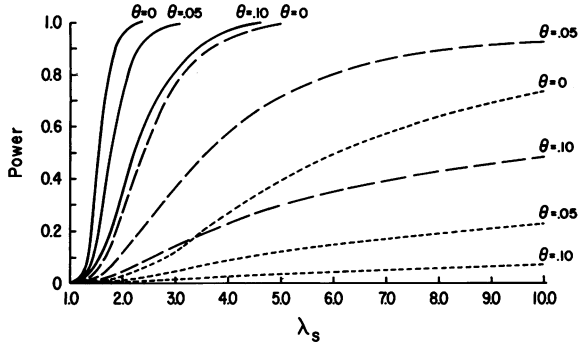

Figure 2 Power to detect linkage as a function of λ_S for three values of θ, by using N affected sib pairs and assuming a fully informative marker. Solid lines, N = 300; dashed lines, N = 100; dotted lines, N = 40.

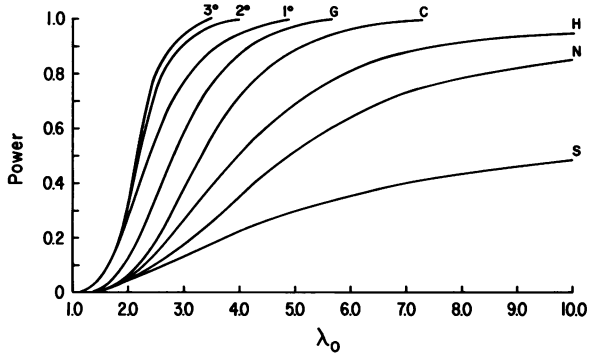

Figure 3 Power to detect linkage as a function λ_O, by using 100 affected relative pairs and assuming a fully informative marker. Lines are denoted as follows: 3° = third-degree relatives, θ = 0; 2° = second-degree relatives, θ = 0; 1° = sibs, θ = 0; C = first cousins, θ = .1; H = half-sibs, θ = .1; N = nephews, θ = .1; S = sibs, θ = .1.

0, corresponding to the value λ_o with recombination fraction θ.

The power to detect linkage by using affected pairs of second- or third-degree relatives is determined in the same manner as described above for sib pairs. Power is calculated for a sample of N = 100 relative pairs, under the assumptions θ = 0 and θ = .10. When θ = 0, all second-degree relatives give the same results; for θ = .10, however, the different types of relatives offer different power. The results are shown in figure 3. When θ = 0, the power curves for second- and third-degree relatives are quite similar; the curve for sibs (first-degree relative) lies below the other two. When θ = .10, the curve for first-, second-, and third-degree relatives are now quite distinct. The greatest power is for grandparent-grandchild pairs, followed by cousins, half-sibs, uncle-nephew, and sibs. The differences among these types of relatives are not minor. For example, for a value of λ_O = 5, the power ranges from 98% for grandparent-grandchild pairs down to 30% for sib pairs.

When θ = 0, the absolute deviation $|\delta_{SO}|$ for sibs is $\frac{1}{4}(\lambda_o - 1)/\lambda_o$. For a unilineal relative of type R, let r = one-quarter of the inverse of the kinship coefficient; for example, for half-sibs, r = 2; for first cousins, r = 4; for second cousins, r = 16. From formula (4) above and from formula (3) of the preceding paper (Risch 1990*a*), the absolute deviation for a relative of type R is $|\delta_{RO}| = \frac{1}{r}(r-1)(\lambda_o - 1)/(\lambda_o + r - 1)$. From this formula, the following relationship holds:

$$|\delta_{SO}| = \frac{1}{2}\frac{\lambda_o + 1}{\lambda_o}|\delta_{NO}| = \frac{1}{3}\frac{\lambda_o + 3}{\lambda_o}|\delta_{CO}|, \qquad (25)$$

where N and C correspond to second- and third-degree relatives, respectively. Therefore, $|\delta_{NO}| > |\delta_{SO}|$ for $\lambda_o > 1$, $|\delta_{CO}| < |\delta_{NO}|$ for $\lambda_o < 3$, and $|\delta_{CO}| > |\delta_{NO}|$ for $\lambda_o > 3$. Although the value of $|\delta_{RO}|$ has a major impact on determining the power to detect linkage for a type of relative pair, the null expectation α_{RO} for no allele sharing also plays a role (e.g., see Bishop and Williamson 1990); for example, although λ_o values between 1 and 3 give slightly larger values of $|\delta_{RO}|$ for second- than for third-degree relatives, slightly greater power is obtained with third-degree relatives, as illustrated in figure 3.

When $\theta > 0$, distant relatives are not automatically preferable to close ones because for distant relatives there are more opportunities for crossing-over to segregate ibd at the marker locus from ibd at the disease locus. Examination of formulas (8)–(22) offers some clues. Consider the coefficient involving the deviation δ_{RO}. If $\theta < .10$, then we can reasonably ignore terms including second or higher powers of θ. Then, for each type of relationship, the approximate coefficients are as follows: sibs, $1 - 4\theta$; grandparent-grandchild, $1 - 2\theta$; uncle-nephew, $1 - 5\theta$; half-sibs, $1 - 4\theta$; cousins, $1 - 5\frac{1}{3}\theta$. Therefore, the coefficient decreases fastest with θ for cousins, followed by uncle-nephew; next are sibs and half-sibs, which are equal, followed by grandparent-grandchild. This is the reason that grandparent-grandchild pairs are least sensitive to the effects of recombination.

It is also possible to consider θ and λ_o simultaneously for the different types of relatives. Generally, when θ = .05, grandparent-grandchild pairs have greatest power up to about λ_o = 4, after which first cousins have greater power. At θ = .10, grandparent-grandchild

pairs are always more powerful than first cousins; half-sib and uncle-nephew pairs are more powerful than cousins only for small values (<2) of λ_o, and even then differences are very slight.

Multilocus Multiplicative Model

It is important to evaluate the prospects for detecting disease-susceptibility loci through linkage analysis when more than a single locus is involved. I first consider the multiplicative model of two unlinked disease-susceptibility loci that was described in the preceding paper (Risch 1990*a*) along with the relevant notation (see also Risch 1987). For this model, the prevalence $K = K_1K_2$ and $\lambda_R = \lambda_{1R}\lambda_{2R}$ (subscripts 1 and 2 denote loci 1 and 2, respectively). Consider the case where a marker locus is linked to disease locus 1; at first assume that $\theta = 0$. Let z_{Ri} be the probability that the pair of affected relatives share i marker alleles ibd. Then

$$z_{R0} = \frac{\alpha_{R0}P(\text{both affected} \mid \text{ibd}=0 \text{ at locus } 1)}{P(\text{both affected})} = \frac{\alpha_{R0}K_1K_1K_2K_{2R}}{K_1K_{1R}K_2K_{2R}} = \frac{\alpha_{R0}}{\lambda_{1R}} = \alpha_{R0} - \frac{\alpha_{R0}}{\lambda_{1R}}(\lambda_{1R}-1)\,. \quad (26)$$

Similarly,

$$z_{R1} = \frac{\alpha_{R1}\lambda_{1o}}{\lambda_{1R}} = \alpha_{R1} + \frac{\alpha_{R1}}{\lambda_{1R}}(\lambda_{1o}-\lambda_{1R}) \quad (27)$$

and

$$z_{R2} = \frac{\alpha_{R2}\lambda_{1M}}{\lambda_{1R}} = \alpha_{R2} + \frac{\alpha_{R2}}{\lambda_{1R}}(\lambda_{1M}-\lambda_{1R})\,. \quad (28)$$

Notice that formulas (26)–(28) are identical to the single-locus formulas (1)–(3), except that λ_{1R} replaces λ_R. In fact, all of the formulas which include the effect of recombination (i.e., formulas [7]–[24]) also generalize in this case, where λ_R is replaced by λ_{1R} for each type of relative. Therefore, all the previous figures and comments regarding power for the different types of relatives continue to hold true for this case. In fact, these results also generalize to any number of loci, provided that there is a multiplicative relationship among the contributing loci. For the additive model, however, the situation is quite different.

Multilocus Additive Model

Consider first the additive model of two unlinked loci as described in the preceding paper (Risch 1990*a*). As was shown there, $K = K_1 + K_2$ and $KK_R = K_1K_{1R} + K_2K_{2R} + 2K_1K_2$, leading to $\lambda_R - 1 = (K_1/K)^2(\lambda_{1R}-1) + (K_2/K)^2(\lambda_{2R}-1)$. Assume that a marker locus is linked to disease locus 1 with recombination fraction $\theta = 0$. Then, using logic similar to that given for the derivation of formulas (13) and (14) in the preceding paper, I obtain the probabilities z_{Ri} of sharing i alleles ibd at the marker locus as follows:

$$z_{R0} = \frac{\alpha_{R0}P(\text{both affected} \mid \text{ibd}=0 \text{ at locus } 1)}{P(\text{both affected})} = \alpha_{R0}\left(\frac{K_1K_1+K_2K_{2R}+2K_1K_2}{K_1K_{1R}+K_2K_{2R}+2K_1K_2}\right) = \alpha_{R0}\left(1-\frac{K_1K_{1R}-K_1^2}{KK_R}\right) = \alpha_{R0} - \alpha_{R0}\left(\frac{K_1}{K}\right)^2\frac{1}{\lambda_R}(\lambda_{1R}-1)\,. \quad (29)$$

Similarly,

$$z_{R1} = \alpha_{R1}\left(\frac{K_1K_{1O}+K_2K_{2R}+2K_1K_2}{K_1K_{1R}+K_2K_{2R}+2K_1K_2}\right) = \alpha_{R1}\left(1+\frac{K_1K_{1O}-K_1K_{1R}}{KK_R}\right) = \alpha_{R1} + \alpha_{R1}\left(\frac{K_1}{K}\right)^2\frac{1}{\lambda_R}(\lambda_{1O}-\lambda_{1R}) \quad (30)$$

and

$$z_{R2} = \alpha_{R2}\left(\frac{K_1K_{1M}+K_2K_{2R}+2K_1K_2}{K_1K_{1R}+K_2K_{2R}+2K_1K_2}\right) = \alpha_{R2}\left(1+\frac{K_1K_{1M}-K_1K_{1R}}{KK_R}\right) = \alpha_{R2} + \alpha_{R2}\left(\frac{K_1}{K}\right)^2\frac{1}{\lambda_R}(\lambda_{1M}-\lambda_{1R})\,. \quad (31)$$

These formulas are not directly analogous to formulas (1)–(3). For example, in the single-locus model, the deviation $\delta_{R0} = (\alpha_{R0}/\lambda_R)(\lambda_R-1)$, while here it is $(\alpha_{R0}/\lambda_R)(K_1/K)^2(\lambda_{1R}-1)$. However, formula (14) of the preceding paper shows that the deviation δ_{1R0} for locus 1 is that proportion of the total deviation (i.e., $\delta_{R0} = [\alpha_{R0}/\lambda_R][\lambda_R-1]$) attributable to locus 1.

Model-dependent Strategies

The differences between formulas (26)–(28) for the multiplicative model and formulas (29)–(31) for the additive model have important implications with regard to strategies for identifying linked loci. For a given locus in the multiplicative model (under the assumption that $\lambda_S = \lambda_O$ and $\theta = 0$), a formula analogous to formula (25) holds; when λ_{1O} is substituted for λ_O,

$$|\delta_{1S0}| = \frac{1}{2}\frac{\lambda_{1O}+1}{\lambda_{1O}}|\delta_{1N0}| = \frac{1}{3}\frac{\lambda_{1O}+3}{\lambda_{1O}}|\delta_{1C0}| . \quad (32)$$

For a given locus in the additive model (again under the assumption that $\lambda_S = \lambda_O$ and $\theta = 0$), $|\delta_{1S0}| = \alpha_{S0}(K_1/K)^2(\lambda_{1O}-1)/\lambda_S = \frac{1}{4}(K_1/K)^2(\lambda_{1O}-1)/\lambda_O$. For second-degree relatives, $|\delta_{1N0}| = \alpha_{N0}(K_1/K)^2(\lambda_{1N}-1)/\lambda_N = \frac{1}{2}(K_1/K)^2(\lambda_{1O}-1)/(\lambda_O+1)$, and, for third-degree relatives, $|\delta_{1C0}| = \alpha_{C0}(K_1/K)^2(\lambda_{1C}-1)/\lambda_C = \frac{3}{4}(K_1/K)^2(\lambda_{1O}-1)/(\lambda_O+3)$. Hence, for the additive model,

$$|\delta_{1S0}| = \frac{1}{2}\frac{\lambda_O+1}{\lambda_O}|\delta_{1N0}| = \frac{1}{3}\frac{\lambda_O+3}{\lambda_O}|\delta_{1C0}| . \quad (33)$$

If the contributing locus has only a small total effect on λ_O while λ_O itself is large, formulas (32) and (33) can lead to significantly different relationships among the δ values; for example, if $\lambda_{1O} = 2$ but $\lambda_O \geq 20$, then for the additive model the ratio $\delta_{1S0}:\delta_{1N0}:\delta_{1C0}$ is 1:2:3, whereas for the multiplicative model the ratio is $1:\frac{4}{3}:\frac{6}{5}$. The implication is that for the multiplicative model the comparative power to detect linkage by using various types of relatives is the same as has been described for a single-locus model, provided that λ_{1O} is substituted for λ_O; that is, it depends on the risk ratio for that locus only. For the additive model, however, comparative power to detect linkage with different relative types depends on the total λ value and not on the λ value for that locus alone.

Consider the following example: Suppose that $\lambda_O = \lambda_S = 10$ and that two loci of identical effect contribute to disease risk. For the multiplicative model, $\lambda_{1S} = \lambda_{1O} = \sqrt{10} \approx 3.162$, giving $|\delta_{1S0}| \approx .1709$, $|\delta_{1N0}| \approx .2597$ and $|\delta_{1C0}| \approx .2633$. With a sample of 100 relative pairs and $\theta = 0$, the power to detect linkage to one of the loci is 84% for sib pairs, 96% for second-degree relatives, and 96% for third-degree relatives. For the additive model, $|\delta_{1S0}| = .1125$, $|\delta_{1N0}| \approx .2045$, and $|\delta_{1C0}| \approx .2596$. With 100 relative pairs and $\theta = 0$, the power to detect linkage is 17% for sib pairs, 67% for second-degree relatives, and 97% for third-degree relatives. In such a situation, therefore, there may be a high probability of detecting linkage by using cousin pairs but a low probability by using sib pairs.

As shown before, the effect of recombination varies among relationships, even within the same degree. However, the principle that the power may be made substantially greater by using more distant relatives when genetic heterogeneity (additive model) is present presumably still holds, at least for small θ values.

This result can be given an intuitive basis. Consider a genetic heterogeneity model whereby a number of distinct, rare alleles lead to disease. Although the overall recurrence in relatives is high compared with population rates, the familiality attributable to any given locus is small. If distant relatives share the same marker allele ibd, there is considerably more evidence for linkage than if close relatives do. The prior probability that two affected relatives share the same disease allele decreases with the distance of the relationship. However, if the disease alleles are rare, the probability that distant affected relatives share the same allele may still be quite high.

In general, the existence of multiple contributing loci can greatly reduce the power of detecting a given locus. The deterioration of power is usually greater for an additive model than for a multiplicative one. For example, in an additive model, with n loci of equal effect, if $|\delta_{S0}| = \frac{1}{4}(\lambda_S-1)/\lambda_S$, the δ value for each locus is $|\delta_{S0}|/n$. If $n = 4$, then $|\delta_{S0}|/n$ must be $<.0625$; similarly, for first cousins, $|\delta_{C0}|/n < .1875$. These values would make linkage detection difficult when using this approach. For a multiplicative model, however, if λ_S is large, the impact may not be as severe. For example, if $\lambda_S = 16$ and $n = 4$, then $\lambda_{1S} = 2$ and $|\delta_{S0}| = .125$, which is double the value for the additive model.

A comparison of marker allele ibd for different degrees of relative can also theoretically distinguish between an additive or multiplicative multilocus model (or, in other words, suggest the presence of epistasis). Consider the following example involving insulin dependent diabetes mellitus (IDDM) and HLA: For this disorder the risk ratio λ_S for sibs has been estimated to be 15 (Risch 1987); Payami et al. (1985) have estimated z_{S0} to be .07, on the basis of data on several hundred affected sib pairs. Therefore, $\delta_{1S0} = .07 - .25 = -.18$. For the multiplicative model, $\lambda_{10} = .25/.07 = 3.5$ for HLA. Therefore, formula (32) gives $\delta_{1N0} = -.28$, or $z_{N0} = .22$ and $\delta_{1C0} = -.29$, or $z_{C0} = .46$. By contrast, for the additive model, formula (33) gives $\delta_{1N0} = -.34$, or $z_{N0} = .16$ and $\delta_{1C0} = -.45$, or $z_{C0} = .30$. Thus, the best discrimination

would be obtained for cousins, although such data are lacking. Barbosa et al. (1977, 1980) collected data for multigenerational families from which affected uncle (aunt)-nephew (niece) pairs can be extracted. Of 24 such pairs, five were discordant for HLA, for an estimated $z_{N0} = .21$, a value closer to that for the multiplicative model than to that for the additive one, although, given the small sample size, the observed number of discordant pairs is not significantly different from either prediction. A larger sample of more distantly related affected pairs could further address this issue, although epistasis is also suggested by the large drop in risk from first-degree to second-degree relatives (Wagener et al. 1982).

The above considerations illustrate that the expected amount of sharing of marker alleles depends both on recombination and on complexities of inheritance. A question then arises as to how to combine linkage evidence from different types of relative pairs if we have such in our data. One approach is to merely add maximum "lod scores" across relative pairs; however, this does not take into account constraints, imposed by the genetic model and recombination, on the sets of possible values of ibd for the different types of relative pairs. Therefore, one must be cautious in adding maximum lod scores and keep in mind that the result may be anticonservative because the maximum total lod score, accounting for genetic constraints, may be less than the total maximum lod score. Examination of the observed ibd values for the different relatives should, however, reveal any major discrepancies from genetic constraints.

In the above analyses, it is also assumed that relative pairs are statistically independent. This will certainly be the case if each pair comes from a different family. When a family has more than two affected relatives, however, not all pairs will be independent, depending on relationships. In general, nonindependence of ibd among relative pairs is a complicated, unresolved issue.

Two Loci with Two Markers

When two contributing disease-susceptibility loci have been identified through linked markers, the joint ibd matrix can provide insight into the relationship between the two loci. For example, for the multiplicative model, the joint ibd matrix also has a multiplicative structure, namely,

$$z_{Rij} = \left(\alpha_{Ri}\frac{\lambda_{1i}}{\lambda_{1R}}\right)\left(\alpha_{Rj}\frac{\lambda_{2j}}{\lambda_{2R}}\right),$$

where $\lambda_{1i} = \lambda_{1M}$ for $i = 2$, $\lambda_{1i} = \lambda_{1O}$ for $i = 1$, and $\lambda_{1i} = 1$ for $i = 0$. This formula holds when $\theta = 0$ for both loci. However, an analogous multiplicative formula also holds when $\theta > 0$.

When the penetrance structure is additive, the joint ibd matrix itself is not additive. However, a matrix whose ij entry is defined as

$$\frac{z_{Rij}}{\alpha_{Ri}\alpha_{Rj}} - 1 = \frac{1}{\lambda_R}\left(\frac{K_1}{K}\right)^2(\lambda_{1i} - \lambda_{1R}) + \frac{1}{\lambda_R}\left(\frac{K_2}{K}\right)^2(\lambda_{2j} - \lambda_{2R})$$

does have an additive structure. Again, the additivity holds true even when $\theta > 0$.

Multipoint Analysis

The previous discussion considered linkage with affected relative pairs by using a single marker locus. However, with chromosomes densely marked with RFLPs, it is important to consider the impact that using multiple linked marker loci has on affected pair analysis. Multipoint analysis not only increases the power of detecting linkage and locating a disease-susceptibility locus but also allows for an estimate of λ_R, even when $\theta \neq 0$ between any of the marker loci and the disease locus, provided that marker loci flanking the disease locus can be obtained.

Consider a map of marker loci which are completely polymorphic. Suppose that ibd for each marker locus is determined for a set of affected relative pairs. Then the disease locus is situated nearest to the marker showing maximum ibd. The interval in which the disease locus resides is determined by flanking markers, although it is difficult, in general, to locate a disease-predisposing allele with great accuracy.

Consider first the case of three marker loci—A, B, and C—which are informative in nonoverlapping families. Suppose we know from other data that the order of the three loci is A–B–C and that the recombination fraction between A and B is .05 and that that between B and C is .03. Assume that 100 distinct pairs of affected grandparents and grandchildren have been obtained for each marker locus. Suppose that at locus A 84 of the 100 pairs share one allele ibd and that 16 are discordant; for locus B, 90 pairs share one allele ibd and 10 are discordant; for locus C, 88 pairs share one allele ibd and 12 are discordant. Recall from formula (16) that $z_{G1} = \frac{1}{2} + \frac{1}{2}(1-2\theta)(\lambda_G - 1)/\lambda_G$ and $z_{G0} = 1 - z_{G1}$.

Let x be the map distance from locus A to the disease locus D of interest; because of the short recombination distances, I assume that map distance and recombination fraction are synonymous. Then the likelihood L of the observed allele sharing for the three marker loci (when constant terms are ignored) is given by

$$\begin{aligned} L = & \left[\frac{1}{2} - \frac{1}{2}\frac{1}{\lambda_G}(\lambda_G - 1)(1-2x)\right]^{16} \\ & \times \left[\frac{1}{2} + \frac{1}{2}\frac{1}{\lambda_G}(\lambda_G - 1)(1-2x)\right]^{84} \\ & \times \left[\frac{1}{2} - \frac{1}{2}\frac{1}{\lambda_G}(\lambda_G - 1)(1-2|x-.05|)\right]^{10} \\ & \times \left[\frac{1}{2} + \frac{1}{2}\frac{1}{\lambda_G}(\lambda_G - 1)(1-2|x-.05|)\right]^{90} \\ & \times \left[\frac{1}{2} - \frac{1}{2}\frac{1}{\lambda_G}(\lambda_G - 1)(1-2|x-.08|)\right]^{12} \\ & \times \left[\frac{1}{2} + \frac{1}{2}\frac{1}{\lambda_G}(\lambda_G - 1)(1-2|x-.08|)\right]^{88} . \end{aligned} \quad (34)$$

Formula (34) requires iteration to obtain maximum-likelihood estimates of λ_G and x. In this case, the computer program MAXLIK (Kaplan and Elston 1972) was used. The solution is $\hat{x} = 0.0616 \pm .0281$ and $\hat{\lambda}_G = 4.856 \pm 0.951$. Notice that the maximum-likelihood estimate of x locates the disease locus D between loci B and C; however, given the large standard error on the estimate, the disease locus could also reasonably lie between A and B or distal to C. Although an estimate of λ_G was obtained, it is easy to translate it into λ_O by using the formula $\lambda_O - 1 = \frac{1}{2}(\lambda_G - 1)$ and the invariance property of maximum-likelihood estimators; namely, $\hat{\lambda}_O = 8.712 \pm 1.902$. Because there are three loci and only two parameters, it is also possible to perform a χ^2 goodness-of-fit test; that is, using formula (16) and the estimates of x and λ_G, I obtain the following expected values for z_{G1} for loci A, B, and C, respectively: .848, .888, and .882. The corresponding χ^2 statistic is $\chi^2_1 = 0.20$, indicating a good fit.

Now consider the case where all three marker loci are informative in 100 affected grandparent-grandchild pairs. Suppose that the numbers sharing a marker allele ibd for the three loci are as follows: A, B, and C—79; A and B only—3; A and C only —0; B and C only—8; A only—2; B only—0; C only—1; none—7. Here again, 84 pairs share an A allele ibd, 90 share a B allele ibd, and 88 share a C allele ibd. Suppose the disease locus D lies between loci B and C. On the assumption that double crossovers do not occur within intervals A-B or B-C, the above categories can be classified as follows: Those cases where alleles of B and C are shared (87 cases) also share an allele at the disease locus ibd; in those cases where alleles at B and C are not shared (nine cases), the disease allele is also not shared ibd. The remaining cases are ambiguous because crossing-over has occurred between B and C. If x is the distance to locus D from locus A, then the probability that an allele at locus D segregates with locus B, given that a crossover has occurred, is $y = (x-.05)/.03$; the probability that it segregates with C is $1 - y = (.08-x)/.03$. Therefore, in those cases ($n = 3$) where an allele at locus B is shared ibd but not at locus C, the probability that an allele of locus D is also shared ibd is y, and $y - 1$ is the probability that it is not. In those cases ($n = 1$) where an allele at C is shared ibd but not at B, the probability that an allele of locus D is also shared is $1 - y$, and y is the probability that it is not. Therefore, the likelihood (aside from a constant) of the above data can be written as

$$\begin{aligned} L &= [1-z_{G0}]^{87}[z_{G0}]^9[z_{G0}y+(1-z_{G0})(1-y)]^3 \\ &\quad [y(1-z_{G0})+(1-y)z_{G0}] \\ &= [1-z_{G0}]^{87}[z_{G0}]^9[1-z_{G0}-y+2z_{G0}y]^3 \\ &\quad [y+z_{G0}-2yz_{G0}] . \end{aligned} \quad (35)$$

The maximum-likelihood solution of equation (35) can be obtained easily from the equations

$$\hat{z}_{G0} = \frac{9}{(87+9)} = \frac{9}{96} = .09375 \quad (36)$$

and

$$\hat{y} + \hat{z}_{G0} - 2\hat{y}\hat{z}_{G0} = \frac{1}{(1+3)} = .250 . \quad (37)$$

Therefore, $\hat{y} = .192$. Equation (36) gives $\hat{\lambda}_G = .5/\hat{z}_{G0} = 5.333$. Also, $\hat{x} = (.03)\hat{y} + .05 = .0558$. Standard errors for $\hat{\lambda}_G$ and $\hat{y}$ can be found by using Taylor's theorem (Elandt-Johnson 1971). For example, $\mathrm{Var}(\lambda_G) = \frac{1}{4}\mathrm{Var}(1/\hat{z}_{G0}) \approx \frac{1}{4}(\hat{z}_{G0})^{-4}\mathrm{Var}(\hat{z}_{G0}) = .25(.09375)^{-4}(1/96)(.09375)(.90625) = 2.864$; therefore, the standard error for $\hat{\lambda}_G$ is $\sqrt{2.864} = 1.692$. Hence, the estimate for λ_o is 9.667 ± 3.384.

The standard error for y is found as follows: In equation (37), the value .25 represents the outcome of one success in four trials for a binomial random variable b. The solution of equation (37) can be written as $\hat{y} = \frac{1}{2} - \frac{1}{2}[(1-2\hat{b})/(1-2\hat{z}_{G0})]$. Therefore,

$$\begin{aligned}\mathrm{Var}(\hat{y}) &= \frac{1}{4}\,\mathrm{Var}\left(\frac{1-2\hat{b}}{1-2\hat{z}_{G0}}\right)\\ &\approx \frac{1}{4}\left(\frac{1-2\hat{b}}{1-2\hat{z}_{G0}}\right)^2 \left\{\frac{\mathrm{Var}(1-2\hat{b})}{(1-2\hat{b})^2} + \frac{\mathrm{Var}(1-2\hat{z}_{G0})}{(1-2\hat{z}_{G0})^2}\right\}\\ &= (.0947)\left\{\frac{.1875}{.25} + \frac{.0035}{.660}\right\}\\ &= .0715\,.\end{aligned}$$

Hence, $\sigma_y = .267$. Thus, the standard error for $\hat{x}$ is $\sigma_x = (.03)\sigma_y = .0080$. Although the most likely location for locus D is between B and C, it can also be reasonably placed between A and B.

In the case of looking at the three loci (A, B, and C) independently—i.e., at a total of 300 relative pairs—the standard errors were $\sigma_\lambda = 0.951$ and $\sigma_x = 0.0281$. To obtain standard errors for an equivalent sample size ($n = 300$) in the analysis of three loci simultaneously, the previously derived standard errors need to be divided by $\sqrt{3}$; this gives values $\sigma_\lambda = 0.977$ and $\sigma_x = 0.0046$. Hence, the precision of the estimate of λ_G is comparable, whether simultaneous or independent analysis is performed. However, the precision of location of the disease locus is much greater in the simultaneous analysis.

Discussion

The power to detect linkage by using affected pairs of relatives depends only on the risk ratios λ_R and on no other genetic model parameters (such as number of alleles at a given locus, gene frequencies, and penetrances). These risk ratios can be estimated from family studies, provided either that a suitable control group is included or that an appropriate estimate of general population risk can be obtained. Usually recurrence-risk estimates for first-degree relatives are available; often values for second- and third-degree relatives have also been obtained. These values can then be applied directly to figures 1–3 to obtain estimates of power for detecting linkage for a given sample size. However, a number of important issues must be addressed.

First, the values given assume a 100% polymorphic marker; hence the power estimate should be considered an upper bound. There are two approaches to remedy the loss of power due to lack of marker polymorphism: (1) the use of multiple, linked markers in a multipont analysis and (2) the use of additional relatives beyond the affected relative pair; for the given relative pair, data on such relatives can help distinguish between identity by state and ibd (this matter is discussed further in the following paper).

It is also important to note that the direct application of λ_R values to figures 1–3 assumes that the value of λ_R can be attributed to the effect of a single locus (this also means that no environmental effect contributes to familial aggregation). In reality, this may not be the case. Complex diseases are often characterized by multiple contributing loci. This fact reduces the prospects (or increases the necessary sample size) for detecting such loci through linkage analysis.

How multiple loci relate to one another can have an important impact on sampling strategy. For example, if there is strong epistasis among loci (as in a multiplicative model), close relatives (sibs) may be preferred to distant relatives. On the other hand, if there is no epistasis (e.g., genetic heterogeneity), it may be advantageous to sample relatives who are more distant (e.g., first cousins). As an example, consider the schizophrenia example described in the preceding paper (Risch 1990*a*). The recurrence data for various classes of relatives suggest an epistatic model of multiple contributing loci. If one were to simply use the value $\lambda_S = \lambda_O = 10$ in figures 1–3, one would then grossly underestimate the sample size necessary to detect linkage. In fact, table 1 from the preceding paper suggests that no single locus can have a $\lambda_O > 3$, and in fact the data are more compatible with λ values $\sim\leqslant 2$. If this is the case, very large samples may be required to detect, through linkage, susceptibility loci for this disease.

While small values of λ_R argue against linkage studies using affected relative pairs, two alternative strategies are possible. First, if the disease can be subcategorized into a form which is more familial, giving higher values for λ_R, power can be increased. Also, the genetic model giving rise to the observed λ_R values is important. One scenario is that genes are common and that penetrance differences among genotypes are not large; in this case, linkage detection will be difficult by any scheme. However, if the small λ_R values are due to a rare allele with high penetrance, in conjunction with a high frequency of sporadic cases, the possibility of detecting linkage is good, provided that pedigrees with the segregating high-risk allele can be obtained. A case in point is breast cancer. Although the lifetime risk to a female first-degree relative of a breast-cancer case is only about twofold above general population risk, early-onset breast cancer is much more familial (Schwartz et al. 1985; Claus et al., in press). Furthermore, for this disease, segregation analysis suggests a rare dominant allele with high penetrance, in conjunction with

sporadic cases (Williams and Anderson 1984; Bishop et al., 1988; Newman et al. 1988; Claus et al., submitted). Hence, either selecting pairs of female relatives with early-onset disease or obtaining extended pedigrees likely to be segregating a high-risk allele would be appropriate strategies in this case.

Figure 2 illustrates the potentially damaging effect that recombination has on power to detect linkage. The effect of recombination, however, can be overcome to some extent by the use of multiple linked markers in a multipoint analysis. Multipoint analysis can also be used to locate the disease-susceptibility locus and to estimate the λ value corresponding to that locus. Figure 2 indicates that a 5-cM map would probably suffice, as the disease locus could be no more than 2.5 cM from the nearest marker, close enough to avoid a sizable loss of power through recombination.

Figure 3 demonstrates that sibs are not necessarily the optimal type of relative pair to sample for linkage studies. When λ_O is large, distant relatives can offer greater power. This is particularly true for grandparent-grandchild pairs, which are least sensitive to the disruptive effects of recombination. In practice, however, it may be difficult to obtain a sizable sample of such pairs. Also, it is important to note that the curves in figure 3 were derived by assuming a 100% polymorphic marker. In fact, when a marker is less polymorphic the impact will be different for different types of relative pairs; distant relatives will be affected the most, sibs the least. Therefore, marker polymorphism also needs to be considered when one is deciding on a sampling strategy. The relationship between marker polymorphism and power is the subject of the following paper (Risch 1989*b*).

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 GM39812.

References

Barbosa J, Chern MM, Anderson VE, Noreen H, Johnson S, Reinsmoen N, McCarty R, et al (1980) Linkage analysis between the major histocompatibility system and insulin-dependent diabetes in families with patients in two consecutive generations. J Clin Invest 65:592–601

Barbosa J, King R, Noreen H, Yunis EJ (1977) The histocompatibility system in juvenile, insulin-dependent diabetic multiplex kindreds. J Clin Invest 60:989–998.

Bishop DT, Cannon-Albright L, McLellan T, Gardner EJ, Skolnick MH (1988) Segregation and linkage analysis of nine Utah breast cancer pedigrees. Genet Epidemiol 5:151–169

Bishop DT, Williamson JA (1990) The power of identity-by-state methods for linkage analysis. Am J Hum Genet 46:254–265

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

Campbell MA, Elston RC (1971) Relatives of probands: models for preliminary genetic analysis. Ann Hum Genet 35:225–236

Claus EB, Risch NR, Thompson WD. Age of onset as an indicator of familial risk of breast cancer. Am J Epidemiol (in press)

——— Autosomal dominant inheritance of early onset breast cancer: implications for risk prediction (submitted-*b*)

Elandt-Johnson RC (1971) Probability models and statistical methods in genetics. John Wiley & Sons, New York

Haseman JK, Elston RC (1972) The investigation of linkage between a quantitative trait and a marker locus. Behav Genet 2:3–19

Kaplan EB, Elston RC (1972) A subroutine package for maximum likelihood estimation (MAXLIK). Institute of Statistics mimeo ser, no 823, University of North Carolina, Chapel Hill

Newman B, Austin MA, Lee M, King MC (1988) Inheritance of human breast cancer: evidence for autosomal dominant transmission in high-risk families. Proc Natl. Acad Sci USA 85:1–5

Payami H, Thomson G, Motro U, Louis EJ, Hudes E (1988) The affected sib method. IV. Sib trios. Ann Hum Genet 49:303–314

Risch N (1987) Assessing the role of HLA-linked and unlinked determinants of disease. Am J Hum Genet 40:1–14

——— (1990*a*) Linkage strategies for genetically complex traits. I. Multilocus models. Am J Hum Genet 46:222–228

——— (1990*b*) Linkage strategies for genetically complex traits. III. The effect of marker polymorphism on analysis of affected relative pairs. Am J Hum Genet 46:242–253

Schwartz AG, King MC, Belle SH, Satariano WA, Swanson CM (1985) Risk of breast cancer to relatives of young breast cancer patients. J Natl Cancer Inst 75:665–668

Suarez BK, Rice J, Reich T (1978) The generalized sib pair IBD distribution: its use in the detection of linkage. Ann Hum Genet 42:87–94

Thompson EA (1986) Pedigree analysis in human genetics. The Johns Hopkins University Press, Baltimore

Wagener DK, Sachs JM, LaPorte RE, MacGregor JM (1982) The Pittsburg study of insulin-dependent diabetes mellitus: risk for diabetes among relatives of IDDM. Diabetes 31:225–236

Williams WR, Anderson DE (1984) Genetic epidemiology of breast cancer: segregation analysis of 200 Danish pedigrees. Genet Epidemiol 1:7–20