Am. J. Hum. Genet. 46:222–228, 1990

Linkage Strategies for Genetically Complex Traits. I. Multilocus Models

Neil Risch

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

Summary

In order to investigate linkage detection strategies for genetically complex traits, multilocus models of inheritance need to be specified. Here, two types of multilocus model are described: (1) a multiplicative model, representing epistasis (interaction) among loci, and (2) an additive model, which is shown to closely approximate genetic heterogeneity, which is characterized by no interlocus interaction. A ratio λ_R of risk for type R relatives that is compared with population prevalence is defined. For a single-locus model, $\lambda_R - 1$ decreases by a factor of two with each degree of relationship. The same holds true for an additive multilocus model. For a multiplicative (epistasis) model, $\lambda_R - 1$ decreases more rapidly than by a factor of two with degree of relationship. Examination of λ_R values for various classes of relatives can potentially suggest the presence of multiple loci and epistasis. For example, data for schizophrenia suggest multiple loci in interaction. It is shown in the second paper of this series that λ_R is the critical parameter in determining power to detect linkage by using affected relative pairs.

Introduction

The advent of highly polymorphic loci densely mapping the human genome has already led to the identification and mapping of loci for a number of Mendelian disorders. However, the prospect of using these RFLPs for complex, non-Mendelian familial diseases holds even greater promise because of the chronic and highly prevalent nature of these disorders, such as cancer, cardiovascular disease, epilepsy, autoimmune disease, psychiatric disorders, and so on. The importance of identifying susceptibility loci for such diseases lies not only in mapping them to the human genome but also in helping to define both the pattern of inheritance and, ultimately, the disease etiology.

Before undertaking a linkage study for a complex disease, one must have some understanding of the prospects of obtaining evidence for linkage; this will depend on the number of loci involed in disease susceptibility, the frequency of high-risk alleles, the relationship of genotypes at multiple loci on risk, the density of the RFLP marker map, and the polymorphic content of the markers involved. In this series of papers, I attempt to evaluate strategies for detection of linkage of disease loci and mode of inheritance determination by using pairs of relatives. The affected-sib-pair paradigm has been used extensively for HLA-associated diseases (Day and Simons 1976; Suarez et al. 1978). Its success there is largely due to the highly polymorphic nature of the HLA complex of loci. However, Thomson (1986) has shown that RFLPs can be used in the same way as HLA using affected sib pairs. It has been established (Day and Simons 1976; Cantor and Rotter 1986; Weeks and Lange 1988) that other types of relative pairs can also be used for linkage analysis, although they have been utilized less frequently.

For relative pairs, deviations from null expectations for marker-allele sharing depend on mode-of-inheritance parameters and the number of loci involved, as well as on the recombination fraction between the marker and the disease locus. A partial assessment of the number of loci and of the relevant parameters for linkage can be made by examining familial recurrence patterns. The present paper examines the essential parameters for single- and multiple-locus models.

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.

0002-9297/90/4602-0002$02.00

Single-Locus Model

Assume that a single locus with n alleles underlies disease susceptibility. Enumerate the alleles as $g_1, g_2, \ldots, g_n$. Let the population frequency of g_i be t_i for $i = 1, \ldots, n$. Let f_{ij} be the penetrance of genotype g_ig_j. Further, define the random variable X_1 to be 1 if individual 1 is affected, and 0 if unaffected; similarly, define X_2 for a related individual 2 of type R. If the Hardy-Weinberg law is assumed to hold, the population prevalence is given by

$$K = E(X_1) = \sum_{i=1}^{n} \sum_{j=1}^{n} t_i t_j f_{ij} . \quad (1)$$

Define $K_R = E(X_2 \mid X_1 = 1)$ to be the recurrence risk for a type R relative of an affected individual. Then the probability that a proband and type R relative are both affected is $K \times K_R = E(X_1X_2) = \mathrm{Cov}(X_1,X_2) + K^2$. Thus, $K_R = K + (1/K)\mathrm{Cov}(X_1,X_2)$. This formula was first derived by James (1971).

Define λ_R as the risk ratio for a type R relative of an affected individual compared with population prevalence; that is,

$$\lambda_R = K_R/K = 1 + (1/K^2)\mathrm{Cov}(X_1,X_2) . \quad (2)$$

James (1971) also indicated that $\mathrm{Cov}(X_1,X_2)$ can be written in terms of the additive genetic variance V_A and of the dominance variance V_D of the penetrances. Specifically, if c_R is twice the kinship coefficient and u_R is the probability the two relatives share two alleles identical by descent, then

$$\lambda_R = 1 + (1/K^2)(c_R V_A + u_R V_D)$$

and

$$\lambda_R - 1 = (1/K^2)(c_R V_A + u_R V_D) . \quad (3)$$

Define the relationship subscripts as follows: M = MZ twin; S = sibling (or DZ twin); 1 = parent (or offspring); 2 = second-degree relative; 3 = third-degree relative. Then $\lambda_M - 1 = (1/K^2)(V_A + V_D)$, $\lambda_S - 1 = (1/K^2)(\frac{1}{2}V_A + \frac{1}{4}V_D)$; $\lambda_1 - 1 = (1/K^2)(\frac{1}{2}V_A)$; $\lambda_2 - 1 = (1/K^2)(\frac{1}{4}V_A)$; $\lambda_3 - 1 = (1/K^2)(\frac{1}{8}V_A)$. Thus,

$$\lambda_1 - 1 = 2(\lambda_2 - 1) = 4(\lambda_3 - 1) . \quad (4)$$

Also,

$$(\lambda_M - 1) - 4(\lambda_S - 1) + 2(\lambda_1 - 1) = 0, \text{ or}$$
$$\lambda_M = 4\lambda_S - 2\lambda_1 - 1 . \quad (5)$$

In addition, if $V_D = 0$, then $\lambda_M - 1 = 2(\lambda_S - 1)$, and $\lambda_S = \lambda_1$.

The implications of these formulas are as follows: If a single disease-susceptibility-locus model applies (with any number of alleles), then the parameter $\lambda_R - 1$ should decrease by a factor of 2 for each decreasing degree of unilineal relationship, using the parent-offspring relationship for the first-degree relative. If the dominance variance is zero (e.g., if parent-offspring risk is the same as for sibs), then the above also holds for MZ twins (as "zero"-degree relatives) and sibs (first-degree relatives). Also, formula (5) must be applicable to any disease for which the familial aggregation is attributable to a single locus.

Two-Locus Models

Now assume that two unlinked loci are involved in disease susceptibility; again I allow for an arbitrary number of alleles and genotypes at each locus. Denote the *genotypes* at the first locus by G_i, $i = 1, \ldots, n$ with corresponding population frequencies p_i and those at the second locus by H_j, $j = 1, \ldots, m$ with corresponding population frequencies q_j. For a pair of relatives of a certain type R, define τ_{kl} as the conditional probability that the relative has genotype l given that the proband has genotype k (i.e., the genotype transition probability). Further, let w_{ij} be the penetrance of genotype G_iH_j; hence, an $n \times m$ matrix W of penetrances can be defined. K, K_R, and λ_R are as defined previously. Then

$$K = \sum_i \sum_j p_i q_j w_{ij} \quad (6)$$

and

$$K \times K_R = \sum_i \sum_j p_i q_j w_{ij} \sum_k \sum_l \tau_{ik}\tau_{jl} w_{kl} . \quad (7)$$

Multiplicative Model

The first two-locus model I consider is a multiplicative model. In this case, the $n \times m$ matrix W can be determined by $n + m$ parameters. Specifically, assume that values $x_1, \ldots, x_n$ and $y_1, \ldots, y_m$ can be defined such that the penetrance $w_{ij} = x_iy_j$. Refer to the x's and y's as "penetrance factors" for loci 1 and 2, respectively. This type of model has been described elsewhere by Hodge (1981). It represents interaction (or epistasis)

between genotypes at two loci. For example, suppose two rare dominant alleles at two distinct loci are required for development of disease; all other combinations of genotypes are at no risk. This situation can be modeled by $x_1 = x_2 = y_1 = y_2 = 1$, $x_3 = y_3 = 0$, where the genotypes at locus 1 are $G_1 = D_1D_1$, $G_2 = D_1d_1$, $G_3 = d_1d_1$, and similarly for locus 2; D_1 is the high-risk allele and d_1 the low-risk allele at locus 1. The multiplicative model leads to simple recurrence-risk formulas. Specifically, formula (6) becomes

$$K = \sum_i \sum_j p_i q_j x_i y_j = \left(\sum_i p_i x_i\right)\left(\sum_j q_j y_j\right) = K_1 \times K_2 , \quad (8)$$

where $K_1 = \Sigma p_i x_i$ and $K_2 = \Sigma q_j y_j$ are the "prevalence factors" defined in terms of the penetrance factors. Similarly, formula (7) becomes

$$\begin{aligned} K \times K_R &= \sum_i \sum_j p_i q_j x_i y_j \sum_k \sum_l \tau_{ik}\tau_{jl} x_k y_l \\ &= \sum_i \sum_k p_i x_i \tau_{ik} x_k \sum_j \sum_l q_j y_j \tau_{jl} y_l \\ &= K_1 \times K_{1R} \times K_2 \times K_{2R} , \end{aligned} \quad (9)$$

where K_{1R} and K_{2R} are defined analogously to the single-locus case but by using the penetrance factors x_i and y_j, respectively. Dividing equation (9) by equation (8), we obtain $K_R = K_{1R}K_{2R}$, and again dividing by equation (8) we get

$$\lambda_R = \lambda_{1R} \times \lambda_{2R} . \quad (10)$$

In other words, for a multiplicative model, the risk ratio λ_R is the product of "risk ratio factors" defined in terms of the penetrance factors for the two contributing loci.

Notice that for this type of model, formulas (4) and (5) no longer hold. In fact, the values $\lambda_R - 1$ decrease more rapidly than by a factor of two for each decreasing degree of relationship. Because formula (4) applies to each of the contributing loci, we have instead

$$\begin{aligned} \lambda_2 &= \lambda_{12}\lambda_{22} = \tfrac{1}{2}(\lambda_{11}+1)\tfrac{1}{2}(\lambda_{21}+1) \\ &= \tfrac{1}{4}(\lambda_{11}+1)(\lambda_{21}+1) \end{aligned}$$

and

$$\lambda_3 = \tfrac{1}{16}(\lambda_{11}+3)(\lambda_{21}+3) . \quad (11)$$

For example, suppose $\lambda_{11} = \lambda_{21} = 4$. Then $\lambda_1 = 16$, $\lambda_2 = 6.25$, and $\lambda_3 \approx 3.06$. Notice that $(\lambda_2 - 1)/(\lambda_1 - 1) = .35$ and that $(\lambda_3 - 1)/(\lambda_2 - 1) \approx .39$.

Additive Model

For this two-locus model, it is assumed that values x_i, $i = 1, \ldots, n$ and y_j, $j = 1, \ldots, m$ can be defined so that the penetrance $w_{ij} = x_i + y_j$. In this case, the x's and y's are referred to as "penetrance summands." I assume that the x's and y's are restricted to the range [0,1]. This model may appear unrealistic because, for certain combinations of genotypes, it can lead to penetrances greater than one; however, I will show in the next section that it gives an excellent approximation to a model of genetic heterogeneity. Define the "prevalence summands" $K_1 = \Sigma_i p_i x_i$ and $K_2 = \Sigma_j q_j y_j$. When the additive relationship is used, formula (6) for prevalence becomes

$$K = \sum_i \sum_j p_i q_j (x_i + y_j) = \sum_i p_i x_i + \sum_j q_j y_j = K_1 + K_2 . \quad (12)$$

The recurrence-risk formula (7) becomes

$$\begin{aligned} K \times K_R &= \sum_i \sum_j p_i q_j (x_i + y_j) \sum_k \sum_l \tau_{ik}\tau_{jl}(x_k + y_l) \\ &= \sum_i \sum_k p_i x_i \tau_{ik} x_k + \sum_j \sum_l q_j y_j \tau_{jl} y_l \\ &\quad + \sum_j \sum_l q_j \tau_{jl} y_l \sum_i p_i x_i \sum_k \tau_{ik} \\ &\quad + \sum_i \sum_k p_i \tau_{ik} x_k \sum_j q_j y_j \sum_l \tau_{jl} \\ &= K_1 K_{1R} + K_2 K_{2R} + 2K_1 K_2 , \end{aligned} \quad (13)$$

where K_{1R} and K_{2R} are defined analogously to K_R but by using the penetrance summands x_i and y_j, respectively. Both the fact that the allele frequencies sum to one and the following equalities were used to obtain formula (13):

$$\sum_k \tau_{ik} = \sum_l \tau_{jl} = 1 ,$$

$$\sum_j \sum_l q_j \tau_{jl} y_l = \sum_l q_l y_l = K_2 ,$$

and

$$\sum_i \sum_k p_i \tau_{ik} x_k = K_1 .$$

Dividing equation (13) by K^2, I obtain

$$\lambda_R = \frac{1}{K^2}(K_1 K_{1R} + K_2 K_{2R} + 2K_1 K_2)$$

$$= 1 + \frac{1}{K^2}(K_1K_{1R} - K_1^2) + \frac{1}{K^2}(K_2K_{2R} - K_2^2)$$
$$= 1 + \left(\frac{K_1}{K}\right)^2(\lambda_{1R} - 1) + \left(\frac{K_2}{K}\right)^2(\lambda_{2R} - 1),$$

or

$$\lambda_R - 1 = \left(\frac{K_1}{K}\right)^2(\lambda_{1R} - 1) + \left(\frac{K_2}{K}\right)^2(\lambda_{2R} - 1), \quad (14)$$

where $\lambda_{1R} = K_{1R}/K_1$ and $\lambda_{2R} = K_{2R}/K_2$. In other words, for an additive model, $\lambda_R - 1$ is a weighted sum of similar terms for each contributing locus, where the weight is the square of the proportion of total prevalence attributable to that locus. Note also for this model that formulas (4) and (5) still hold because they hold for each contributing locus; that is,

$$\lambda_1 - 1 = \left(\frac{K_1}{K}\right)^2(\lambda_{11} - 1) + \left(\frac{K_2}{K}\right)^2(\lambda_{21} - 1)$$
$$= 2\left(\frac{K_1}{K}\right)^2(\lambda_{12} - 1) + 2\left(\frac{K_2}{K}\right)^2(\lambda_{22} - 1)$$
$$= 2(\lambda_2 - 1)$$

and, similarly, $\lambda_1 - 1 = 4(\lambda_3 - 1)$. Therefore, for this type of model the predicted pattern of recurrence risks in relatives is identical to that of the single-locus model.

Genetic Heterogeneity Model

For this two-locus model, I assume that loci 1 and 2 are independent causes of disease; an individual can be affected through possessing a predisposing genotype at either locus. I again define for the two loci marginal penetrances x_i, $i = 1, \ldots, n$ and y_j, $j = 1, \ldots, m$ whose values range from 0 to 1. The penetrance w_{ij} is defined as $1 - (1-x_i)(1-y_j) = x_i + y_j - x_iy_j$; that is, the probability that an individual is affected is the probability that he or she becomes affected through either locus mechanism. This model is more realistic than the additive model because penetrances greater than one are not possible. The prevalence formula (6) now becomes

$$K = \sum_i \sum_j p_iq_j\,[1-(1-x_i)(1-y_j)]$$
$$= 1 - \sum_i \sum_j p_i\,(1-x_i)q_j(1-y_j)$$
$$= 1 - (1-K_1)(1-K_2)$$
$$= K_1 + K_2 - K_1K_2, \quad (15)$$

where K_1 and K_2 are the prevalence summands when one uses x_i and y_j, respectively, as they have been defined above. The recurrence-risk formula (7) is given by

$$K \times K_R = \sum_i\sum_j p_iq_j\sum_k\sum_l \tau_{ik}\tau_{jl}[1-(1-x_i)(1-y_j)] \times [1-(1-x_k)(1-y_l)]$$
$$= 1 - 2(1-K_1)(1-K_2) + (1-2K_1+K_1K_{1R})(1-2K_2+K_2K_{2R}), \quad (16)$$

where K_{1R} and K_{2R} are defined analogously to K_R by using the penetrance summands x_i and y_j, respectively. In the Appendix, it is shown that formula (16) reduces to

$$\lambda_R - 1 = \left(\frac{K_1}{K}\right)^2(\lambda_{1R} - 1) + \left(\frac{K_2}{K}\right)^2(\lambda_{2R} - 1) - \left(\frac{K_1}{K}\right)\left(\frac{K_2}{K}\right)[2K_1(\lambda_{1R} - 1) + 2K_2(\lambda_{2R} - 1) - K_1K_2(\lambda_{1R}\lambda_{2R} - 1)]. \quad (17)$$

In general, the last term is small compared with the first two, especially when the prevalence summands K_1 and K_2 are small. Hence, formula (17) for the genetic heterogeneity model is well approximated by the additive-model formulation given in the previous section.

Extension to Additional Loci

The formulas given in the previous sections for the two-locus multiplicative and additive models can be readily extended to include an arbitrary number of loci. For example, assume there are L loci. For the multiplicative model, the prevalence formula (8) becomes

$$K = K_1K_2 \ldots K_L, \quad (18)$$

where K_i is the prevalence factor for the ith locus. Simi-

larly, the recurrence-risk formula (10) can be generalized to L loci to give

$$\lambda_R = \lambda_{1R}\lambda_{2R}. \; . \; .\lambda_{LR} , \qquad (19)$$

where λ_{iR} is defined using the penetrance factor for the ith locus. Also, formula (11) generalizes to

$$\lambda_2 = \left(\frac{1}{2}\right)^L \prod_{i=1}^{L} (\lambda_{i1} + 1) ,$$

$$\lambda_3 = \left(\frac{1}{4}\right)^L \prod_{i=1}^{L} (\lambda_{i1} + 3) ,$$

and so on.

Consider a model where each of the λ_{i1}'s is near unity. Then

$$\lambda_2 = \prod_{i=1}^{L} \left(\frac{\lambda_{i1}+1}{2}\right) = \prod_{i=1}^{L} \left(1 + \frac{\lambda_{i1}-1}{2}\right)$$
$$\approx \exp\left[\frac{1}{2}\sum_{i=1}^{L} (\lambda_{i1} - 1)\right] .$$

Similarly,

$$\lambda_1 \approx \exp\left[\sum_{i=1}^{L} (\lambda_{i1} - 1)\right] .$$

Both of the above approximations hold for $\lambda_{i1} - 1$ near zero for each i. Therefore, $\lambda_2 \approx \lambda_1^{1/2}$. Similarly, $\lambda_3 \approx \lambda_1^{1/4}$, and so on. If $V_D = 0$, then $\lambda_M \approx \lambda_1^2$. In other words, in a multiplicative model with a large number of loci each with small effect, the risk ratios decrease by a power of one-half with each degree of relationship. For example, if the risk ratio for parents/offspring is 10, the ratio for second-degree relatives (e.g., half-sibs) is 3.16 and that for third-degree relatives (e.g., first cousins) is 1.78.

General Multilocus Model

The multiplicative and additive models were considered for their mathematical simplicity and relevance to real biological situations. However, for a given trait, neither of these models may be appropriate. A general formulation was given by James (1971) that applies to any number of loci and any number of alleles. For this case, one needs to define epistatic sources of variance of the penetrances. Specifically, define V_{sAtD} as the variance due to an sth-order interaction of additive components and tth-order interaction of dominance components (Kempthorne 1957). Then, for a type R relative of an affected individual,

$$\lambda_R - 1 = \frac{1}{K^2}\left[\sum_{n=1}^{\infty} \sum_{s+t=n} (c_R)^s (u_R)^t \, V_{sAtD}\right] . \qquad (20)$$

When $u_R = 0$, or under strict additivity within loci,

$$\lambda_R - 1 = \frac{1}{K^2}\left[\sum_{s=1}^{\infty} (c_R)^s \, V_{sA}\right] . \qquad (21)$$

Formula (21) shows that $\lambda_R - 1$ decreases by greater than a factor of two with each degree of relationship if and only if epistatic variance components are present. The rate of decrease depends on the number of underlying loci and on the degree of epistasis. In general, the lower-order variance terms are larger than the higher-order ones. Dempster and Lerner (1950) showed that large epistatic variance components are obtained only when heritability is high and when prevalence is low (equivalently, large λ_R values).

Example–Schizophrenia

There is currently considerable interest in understanding the genetic component in major psychiatric illness, particularly for severe conditions such as schizophrenia. The advent of numerous polymorphic markers (RFLPs) in humans offers the possibility of using linkage analysis to detect loci which contribute to susceptibility to schizophrenia. However, it is important to know at the outset the power that is likely to obtain for such an analysis. An assessment can be made based on familial recurrence patterns. For schizophrenia, recurrence risks have been estimated for a range of first-, second-, and third-degree relatives, as well as for MZ and DZ twins. A summary of the different studies has been given by McGue et al. (1983). When a population lifetime incidence of 0.85% was assumed, risk ratios were derived from the age-adjusted risks presented by McGue et al. (1983) for offspring, sibs, MZ and DZ twins, half-sibs, nieces/nephews, grandchildren, and first cousins. The numbers are given in table 1. The risk to sibs is consistently lower than that to offspring; this can be partially explained by a reproductive disadvantage of affected individuals, which tends to diminish the recurrence risk for sibs (Risch 1983). Also, the DZ twin concordance is generally higher than sibling

Table 1

Multilocus Multiplicative Models for Schizophrenia

Risk Ratio[a]	Observed	Model Prediction[b]						
		I	II	III	IV	V	VI	VII
λ_O	10.0	10.0	10.0	10.0	10.0	10.0	10.0	10.0
λ_S	8.6							
λ_M	52.1	19.0	100.0	75.0	55.6	43.8	56.3	42.2
λ_D	14.2							
λ_H	3.5							
λ_N	3.1							
λ_G	3.3							
(λ_2).....	3.2	5.50	3.16	3.35	3.65	3.95	3.56	3.77
λ_C.....	1.8	3.25	1.78	1.87	2.03	2.20	1.96	2.07

[a] Definitions of subscripts: O = offspring; D = DZ twins; H = half-sibs; N = niece/nephew; G = grandchild; C = first cousins. All other subscripts are as defined in the text.

[b] Definitions of models: I—one locus, $\lambda_{1o} = 10.0$; II—infinite loci, each with small effect; III—$\lambda_{1o} = 2.0$, infinite other loci; IV—$\lambda_{1o} = 3.0$, infinite other loci; V—$\lambda_{1o} = 4.0$, infinite other loci; VI—$\lambda_{1o} = \lambda_{2o} = 2.0$, infinite other loci; VII—$\lambda_{1o} = \lambda_{2o} = \lambda_{3o} = 2.0$, infinite other loci.

risk, which may reflect either an environmental impact, which is more similar for twins than for sibs, or alternatively an ascertainment bias in twin series of schizophrenia. In any event, because sib risk is not higher than that for offspring, there is no evidence for a V_D component for this disease. Therefore, only a single parameter ($\lambda_o = \lambda_S = \lambda_1$) is necessary to predict all relative types. Predictions for various multiplicative multilocus models are given in table 1. A compromise value of $\lambda_1 = 10.0$ was chosen to correspond to the value $\lambda_o = 11.0$ for offspring and $\lambda_S = 8.6$ for sibs. The λ values for the three types of second-degree relatives are quite similar, and a weighted average value (λ_2) of 3.2 was used.

Predictions for a single-locus model with λ_1 fixed at 10.0 are given in table 1 under column I. A clear pattern emerges—the MZ twin ratio is seriously underestimated and the ratios for second- and third-degree relatives are overestimated. This pattern suggests an interaction of multiple loci (epistasis) that leads to a faster decrease in λ with increasing distance of relationship. Predictions for a multiplicative model specifying an infinite number of loci with small effect is given in table 1 under column II. White consistency is seen for second- and third-degree relatives, the MZ ratio is overestimated. In an attempt to determine the largest plausible single-locus contribution and the possible number of loci involved, the following additional models were examined for consistency with the data: III—one locus with $\lambda_{1o} = 2.0$, the remaining loci all of small effect; IV—one locus with $\lambda_{1o} = 3.0$, the remaining loci of small effect; V—one locus with $\lambda_{1o} = 4.0$, the remaining loci of small effect; VI—two loci with $\lambda_{1o} = \lambda_{2o} = 2.0$, all other loci of small effect; VII—three loci with $\lambda_{1o} = \lambda_{2o} = \lambda_{3o} = 2.0$, the remaining loci of small effect. While the models with one major locus and with other minor loci improve the correspondence with observed ratios, a value of $\lambda_{1o} = 4.0$ (model V) begins to show a sizeable discrepancy from the observed ratios in the same direction as does the single locus model; therefore, one locus with a value of λ greater than 3.0 is unlikely. Models specifying a single locus with a λ value of 3.0 and with all other loci of small effect or models with two or three loci with λ values of 2.0 (models VI and VII) appear to be most consistent with the observed data.

Although a multiplicative model may not strictly apply for schizophrenia, formula (21) for the general multilocus model also suggests that significant epistatic variance components are necessary to explain the observed drop in risk with degree of relationship. If only two loci are involved, then $(\lambda_M - 1):(\lambda_o - 1):(\lambda_2 - 1):(\lambda_C - 1) = (V_A + V_{AA}):(\frac{1}{2}V_A + \frac{1}{4}V_{AA}):(\frac{1}{4}V_A + \frac{1}{16}V_{AA}):(\frac{1}{8}V_A + \frac{1}{64}V_{AA})$. A very high ratio of V_{AA} to V_A (intense epistasis) would be necessary to obtain the observed descent by degree of relationship given in Table 1. With three or more loci, the required intensity of epistasis would be less.

While it might be argued that the steep decline in risk with degree of relationship is attributable, at least

in part, to nongenetic familial determinants, such an argument has little impact on the conclusion regarding the magnitude of contributing loci. Even if some of the familial determinants are not genetic, a single locus accounting for a large proportion of the familial aggregation of schizophrenia is not compatible with the observed family data.

Discussion

In this first paper of the series, I have described parameterizations of single- and multiple-locus models of inheritance of disease susceptibility. Attention was focused on ratios of risk for relatives compared with population prevalence; it will be shown that these risk ratios are the essential parameters for determining power to detect linkage by using affected relative pairs.

When, compared with population prevalence, there is a high ratio of risk to first-degree relatives (say, greater than fivefold), there is the potential for evaluating the consistency of recurrence-risk patterns with single-locus inheritance. However, it has been shown that genetic heterogeneity models lead to predictions identical to those of single-locus models; therefore, when using such data, it is impossible ever to conclude that only one locus is involved in disease risk. However, when recurrence patterns are not consistent with single-locus inheritance, it is possible to suggest the presence of multiple contributing loci. For diseases characterized by a relatively lower risk to first-degree relatives (say less than fivefold), compared with population prevalence, a discriminatory analysis of this type will not, in general, be possible, because all genetic models will give similar predictions, even when epistatic variance components are present. In this case, and in general, consideration of larger constellations of relatives may offer additional insight into the number of contributing loci; however, as will be shown in the subsequent paper, power to detect linkage by using relative pairs depends only on simple recurrence risks.

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 made numerous improvements and corrections. Financial support was provided by NIGMS grant GM39812.

Appendix

From formula (16),

$$\begin{aligned} K \times K_R - K^2 &= (1-2K_1+K_1K_{1R})(1-2K_2+K_2K_{2R}) \\ &\quad - (1-K_1)^2(1-K_2)^2 \\ &= [(1-K_1)^2+K_1^2(\lambda_{1R}-1)] \times \\ &\quad [(1-K_2)^2+K_2^2(\lambda_{2R}-1)] \\ &\quad - (1-K_1)^2(1-K_2)^2 \\ &= (1-K_2)^2K_1^2(\lambda_{1R}-1) \\ &\quad + (1-K_1)^2K_2^2(\lambda_{2R}-1) \\ &\quad + K_1^2K_2^2(\lambda_{1R}-1)(\lambda_{2R}-1) \\ &= (1-2K_2)K_1^2(\lambda_{1R}-1) \\ &\quad + (1-2K_1)K_2^2(\lambda_{2R}-1) \\ &\quad + K_1^2K_2^2(\lambda_{1R}\lambda_{2R}-1) \\ &= K_1^2(\lambda_{1R}-1) + K_2^2(\lambda_{2R}-1) \\ &\quad - K_1K_2[2K_1(\lambda_{1R}-1)+2K_2 \\ &\quad (\lambda_{2R}-1)-K_1K_2(\lambda_{1R}\lambda_{2R}-1)] \,. \end{aligned}$$

Dividing both sides by K^2 gives formula (17).

References

Cantor RM, Rotter JI (1987) Marker concordance in pairs of distant relatives: a new method of linkage analysis for common diseases. Am J Hum Genet 41:A252

Day NE, Simons MJ (1976) Disease susceptibility genes—their identification by multiple case family studies. Tissue Antigens 8:109–119

Dempster ER, Lerner IM (1950) Heritability of threshold characters. Genetics 35:212–236

Hodge SE (1981) Some epistatic two-locus models of disease. I. Relative risks and identity by descent distributions in affected sib pairs. Am J Hum Genet 33:381–395

James JW (1971) Frequency in relatives for an all-or-none trait. Ann Hum Genet 35:47–48

Kempthorne O (1957) An introduction to genetic statistics. John Wiley & Sons, New York

McGue M, Gottesman II, Rao DC (1983) The transmission of schizophrenia under a multifactorial threshold model. Am J Hum Genet 35:1161–1178.

Risch N (1983) Estimating morbidity risks in relatives: the effect of reduced fertility. Behav Genet 13:441–451

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

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