Am. J. Hum. Genet. 66:593–604, 2000

The Relationship between the Sibling Recurrence-Risk Ratio and Genotype Relative Risk

Benjamin A. Rybicki[1] and Robert C. Elston[2]

[1]Department of Biostatistics and Research Epidemiology, Henry Ford Health System, Detroit, and [2]Department of Epidemiology and Biostatistics, Rammelkamp Center for Education and Research, Case Western Reserve University, MetroHealth Medical Center, Cleveland

Summary

The recurrence-risk ratio of disease in siblings, λ_S, is a standard parameter used in genetic analysis to estimate the statistical power for detection of a disease locus. However, the relationship between the underlying risk conferred by a disease-susceptibility allele and λ_S has not been well described. The former is generally quantified as a genotype relative risk, γ, and measures the ratio of disease risks between those with and those without the susceptibility genotype(s). We demonstrate that λ_S varies significantly more with respect to γ and the disease-allele frequency for two-locus multiplicative models than for other two-locus and for single-locus models. For the single- and two-locus dominant-inheritance models that we studied, when a disease-susceptibility allele had a frequency $\geq$.2, λ_S had an upper limit of <10. In general, λ_S values >10 are possible only under recessive inheritance, dominant inheritance with relatively rare (<5%) disease-susceptibility alleles, or when two or more disease loci have alleles acting either epistatically or multiplicatively. We introduce the idea of a restricted sib recurrence-risk ratio (λ^*_S) estimated by restriction of sibships to those ascertained through a proband who already has a putative high-risk allele. A λ^*_S larger than the λ_S value estimated from randomly selected probands can serve as an indirect way of testing whether the posited susceptibility allele increases disease risk. Our results demonstrate that a λ_S of 2–3 may portend successful mapping for a variety of genetic models but that, for some two-locus models, a λ_S as high as 10 does not guarantee underlying genes easily mapped by linkage.

Received September 17, 1999; accepted for publication November 16, 1999; electronically published February 16, 2000.

Address for correspondence and reprints: Dr. Benjamin A. Rybicki, Department of Biostatistics and Research Epidemiology, Henry Ford Health System, 1 Ford Place, 3E, Detroit, MI 48202. E-mail: brybick1@hfhs.org

0002-9297/2000/6602-0027$02.00

Introduction

The aggregation of a disease or physiologic trait in families is the first observable clue for an underlying genetic susceptibility. A standard measure of familial aggregation is the recurrence-risk ratio of disease in relatives of an index case, defined as the risk of disease in relatives of a random individual with disease, divided by the population prevalence of the disease. While the relative recurrence-risk ratio has long been used as a measure of familial aggregation, it has become a linchpin for genetic linkage studies since the three seminal papers by Risch in 1990 (Risch 1990*a*, 1990*b*, 1990*c*). Risch outlined how the relative recurrence-risk ratio, which he denoted as "λ," could be used to infer a genetic model of inheritance (Risch 1990*a*). He also showed that under a multilocus multiplicative model, the total λ can be factored into locus-specific risk ratios (i.e., $\lambda = \lambda_1 \lambda_2 \ldots \lambda_i$), where i = the number of disease loci, so that, under certain assumptions, λ can be used as a parameter to estimate the power of an affected-relative-pair genetic linkage analysis (Risch 1990*a*; Risch 1990*b*).

More recently, Risch and Merikangas (1996) extended this paradigm by showing the genotypic relative risks (defined as the risk to the heterozygote relative to that for a person with no susceptibility alleles) and disease-susceptibility-allele frequencies necessary to conduct anticipated-association studies of the whole genome. The premise behind this work was that linkage analysis will likely be unable to locate many of the remaining genes with modest effects on risk for complex diseases. Others have refuted this idea (Scott et al. 1997), and, at present, linkage analysis remains the standard method for localization of disease genes. Since many researchers use λ_S, the sib recurrence-risk ratio, to estimate the power of a planned linkage study, we investigate here the relationship between the directly estimable λ_S and the frequency and relative risk of the underlying susceptibility genotype(s).

We demonstrate that, in general, the overall λ_S estimated from disease-prevalence data does not by itself provide a reliable parameter for estimating the statistical power of a proposed linkage study. An elevated λ_S can

be the result of one or more susceptibility genotypes inherited in a variety of ways. For instance, when more than one locus is involved, a large λ_S can be the result of genotype-specific effects that are likely undetectable by conventional linkage-analysis methods. On the other hand, we show that, by careful choice of the sample, mapping of disease-susceptibility loci by use of affected-sib-pair (ASP) methods may be possible, even when very modest levels of λ_S are observed.

Relationship between λ and Genotype Relative Risk

As a measure of genetic risk, the limitations of λ are well known: both shared genes and shared environment likely contribute to familial aggregation of a complex disease. More generally, we can say that λ is subject to exposure misclassification, since the "exposure" effect that λ is supposed to measure is the risk conferred by susceptibility gene(s) shared with the proband. To better illustrate the relationship between λ and the underlying genotype relative risk, γ, consider each parameter separately. We can define γ in epidemiological terms as the risk of disease for individuals with a given susceptibility genotype, divided by the risk of disease for those without the susceptibility genotype (i.e.—with the terminology appropriate for an environmental factor—the risk in the exposed, divided by the risk in the unexposed). If $g+$ symbolizes those with the susceptibility genotype (or, more generally, the set of such genotypes), $g-$ symbolizes those without the susceptibility genotype (or the set of such genotypes), and D represents disease, then, expressed in terms of probability,

$$\gamma = \frac{\Pr(D \mid g+)}{\Pr(D \mid g-)} \quad . \tag{1}$$

On the other hand, λ is the risk of disease in the relatives of an individual with disease, divided by the population prevalence of disease; that is,

$$\lambda = \frac{\Pr(D \mid AR)}{\Pr(D)} \quad , \tag{2}$$

where "AR" denotes "affected relative," and the denominator of λ, $\Pr(D)$, is the population prevalence of disease. If a disease is determined solely by genetic factors, disease prevalence can be simply expressed in terms of genotypic frequencies and conditional probabilities; that is, $\Pr(D) = p_G \Pr(D|g+) + p_0 \Pr(D|g-)$, where p_G and p_0 represent, respectively, the frequencies of those with and without the susceptibility genotype(s) and sum to 1. Substituting this value for $\Pr(D)$ into the expression for λ, equation (2), and solving for the conditional probability of disease if one has the susceptibility genotype, $\Pr(D|g+)$, we find that the result is

$$\Pr(D \mid g+) = \frac{\{[\Pr(D \mid AR)]/\lambda\} - p_0 \Pr(D \mid g-)}{p_G} \quad .$$

This value for $\Pr(D|g+)$ can be substituted into the expression for γ, equation (1), so that γ can be expressed in terms of λ, genotypic frequencies, and disease conditional probabilities; that is,

$$\gamma = \frac{\{[\Pr(D \mid AR)]/\lambda\} - p_0 \Pr(D \mid g-)}{p_G(D \mid g-)} \quad . \tag{3}$$

To give an example of the possible large differences in magnitude between λ and γ, consider a disease for which $\lambda_S = 5$ and the disease prevalence is 1/1,000. If the risk of disease in those without the susceptibility genotype is 1/10,000 and 90% of the population does not have the susceptibility genotype, then γ, according to the expression (3) above, would be 91, a value almost 20 times larger than λ_S.

Single-Locus Model

To further explore the relationship between λ and γ with regard to a purely genetic model, consider first the simplest case of a dominant allele that increases risk of disease. If we assume that all other alleles at the locus in question make an equally lower contribution to disease risk, then the population prevalence of the disease can be expressed in terms of genotypic frequencies and baseline disease risk. If K_0 is the prevalence of disease for those without the susceptibility allele, and K_G is the prevalence of disease for those with the susceptibility allele—that is, $K_G = K_0 g$—then, under Hardy-Weinberg equilibrium, the population prevalence of disease for a dominantly inherited allele is $K_G(p^2 + 2pq) + K_0 q^2$, where p is the frequency of the disease allele and $q = 1 - p$. This term is the denominator of λ. If we consider the specific case of sibs, then the numerator of λ_S can be expressed as three separate probabilities (corresponding to the proband having two, one, or no disease alleles), which sum to the total expected prevalence of disease in sibs of affected patients (for a detailed explanation of how the numerator for λ_S is derived, see the Appendix).

Figure 1*A* and *B* shows the values of λ_S for a single-locus model, over a range of γ values and allele frequencies, for dominant and recessive models. For a dominantly inherited susceptibility allele (fig. 1*A*), λ_S values >5 are not possible for γ values <100 unless the disease-allele frequency is <5%. For a disease allele with a frequency of 1%, the increase in λ_S is approximately linear for γ values of 10–50. Overall, the relationship between λ_S and γ is sigmoidal, with λ_S approaching an upper limit as γ increases, that is inversely related to the disease-allele frequency. At the higher disease-susceptibility-

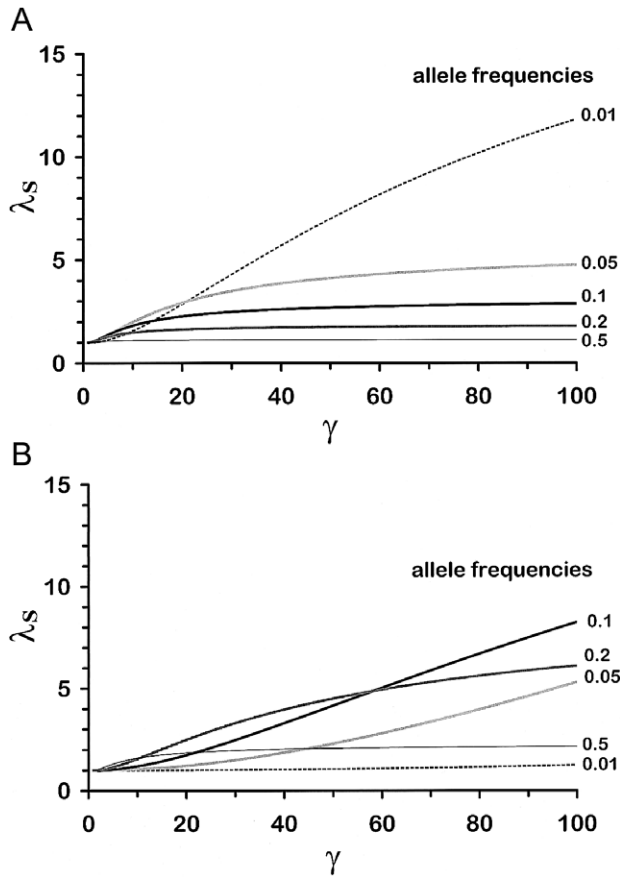

Figure 1 *A*, Relationship between λ_S and γ for disease-susceptibility allele frequencies of .01–.5 under a single-locus dominant-inheritance model. *B*, Relationship between λ_S and γ for disease-susceptibility-allele frequencies of .01–.5 under a single-locus recessive-inheritance model.

allele frequencies ($p \geq .05$), λ_S rises rapidly with respect to γ and then quickly levels off at modest levels of λ_S.

For a single recessive disease allele (fig. 1*B*), the expected values of λ_S at disease-allele frequencies $\geq .1$ are generally higher than those in the dominant case at corresponding values of γ, but overall the same type of relationship holds; that is, for large values of γ, the highest values of λ_S are at the lowest allele frequencies (data not shown). For instance, for γ values <50, the λ_S values for a disease allele at frequency $p = .2$ are greater than at $p = .1$, but for γ values >50 the reverse is true. A similar cross in curves would be observed between $p = .01$ and each of the other four curves if the *X*-axis of figure 1*B* were extended to higher values of γ. In general, under single-locus inheritance, any given value of λ_S can result from more than one inheritance pattern. For example, a λ_S value of 5 is possible for either a dominant allele with a γ value of ~30 and frequency .01 or a recessive allele with a γ value of ~60 and frequency .1–.2.

Two-Locus Model

For most complex diseases, it is likely that more than one disease locus exists. In a two-locus model, apart from having to account for more susceptibility genotypes, one must also consider the type of joint action between loci (for the derivation of λ_S for the two-locus model, see the Appendix). Consider first the situation in which two unlinked loci act epistatically. Figure 2 shows the expected values of λ_S under three different models of epistasis. In model 1, both dominant alleles, A and B, must be present for disease risk to be increased. Therefore, of the nine possible genotypes, four have an increased risk of disease. In model 2, A and B act in a recessive manner, and both must be present in the homozygous state for disease risk to be increased. Under this model, only one of the nine possible genotypes increases disease risk. In model 3, three or more of the disease alleles must be present for disease risk to be increased (i.e., three of the nine possible genotypes increase disease risk). This model is similar to a simple additive polygenic threshold model. Models 2 and 3 are comparable on the basis of expected values of λ_S. This is not surprising, because model 2 is also similar to an additive polygenic threshold model, requiring at least four disease alleles to be present for disease risk to be increased. Model 1 differs from models 2 and 3 in that, for a greater range of allele frequencies (.05–.2), higher values of λ_S are expected over this range of γ values. Epistatic models 1–3 have λ_S values of magnitude similar to those of the single-locus models depicted in figure 1*A* and *B* for γ values ≤ 100. The difference between the two lies in the upper limits of λ_S that are higher for the two-locus epistatic models than for the single-locus models as $\gamma \to \infty$.

Figures 3 and 4 examine two different two locus models, dominant-dominant and recessive-recessive, with additive (fig. 3*A* and *B*) and multiplicative (fig. 4*A* and *B*) interlocus action on γ. As with the single-locus models, five different susceptibility-allele frequencies were examined, in which each locus has one susceptibility allele and both are at the same frequency in the population. For additive action, the overall genotype-specific relative risk of two loci, A and B, equals the sum of the two separate genotype relative risks minus 1: $g_{AB} = g_A + g_B - 1$. For multiplicative action, the overall genotype-specific relative risk equals the product of the two separate genotype relative risks: $g_{AB} = g_A \times g_B$. In these two-locus additive or multiplicative models, the relationship between λ_S and γ is similar to that of the single-locus models. In the case of dominant inheritance at two disease loci with additive action between loci (fig. 3*A*), the values of λ_S for γ values of 1–100 have a pattern similar to that of the single-locus dominant model (fig. 1*A*), but the single-locus model has slightly higher values of λ_S at γ values <100. The same general relationship between

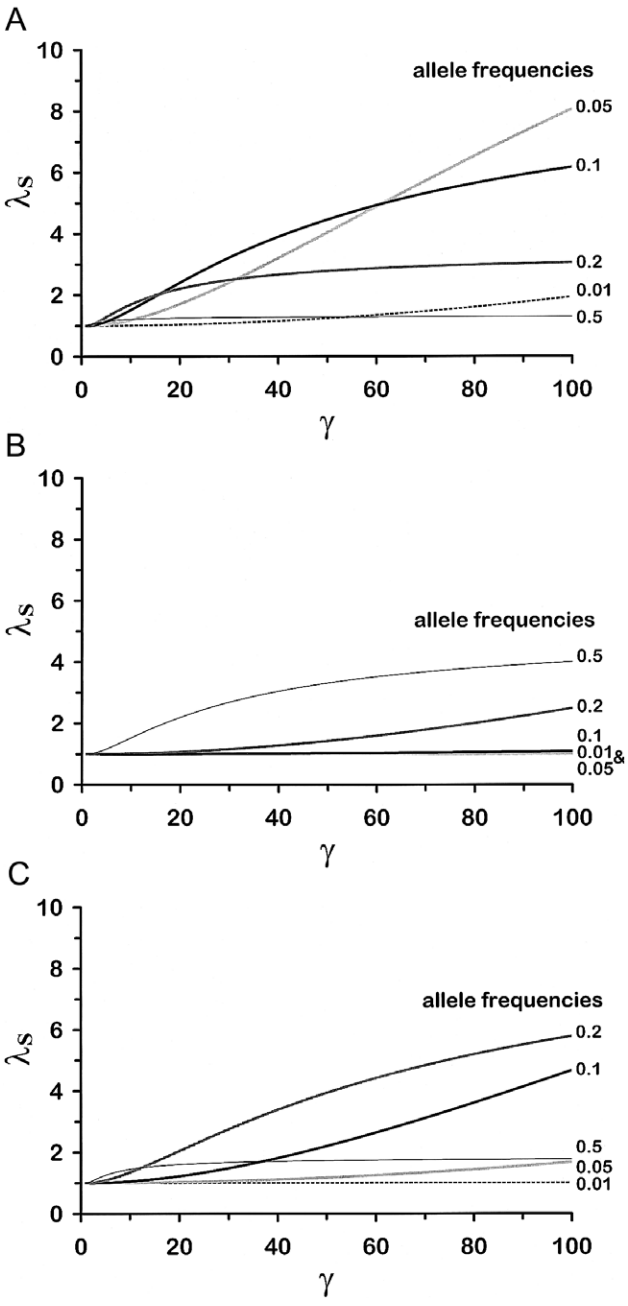

Figure 2 *A*, Relationship between λ_S and γ, for disease-susceptibility-allele frequencies of .01–.5, under a two-locus epistatic-inheritance model (epistatic model 1), in which at least one disease-susceptibility allele from each locus must be present for increased disease risk. *B*, Relationship between λ_S and γ, for disease-susceptibility-allele frequencies of .01–.5, under a two-locus epistatic-inheritance model (epistatic model 2), in which two copies of a disease-susceptibility allele from each locus must be present for increased disease risk. *C*, Relationship between λ_S and γ, for disease-susceptibility allele frequencies of .01–.5, under a two-locus epistatic-inheritance model (epistatic model 3), in which a total of three or more copies of a disease-susceptibility allele from both loci must be present for increased disease risk.

the single-locus and two-locus additive models holds true for recessive inheritance (fig. 1*B* vs. fig. 3*B*), in which comparable λ_S values are slightly higher or lower for the two-locus model, depending on the allele frequency.

The main difference in the relationship between λ_S and γ for the two-locus compared with the single-locus models is quantitative rather than qualitative, as illustrated by the two multiplicative-action models investigated (fig. 4*A* and *B*), in which, for corresponding values of γ, the resulting λ_S values are much higher than those in the single-locus model. The two multiplicative two-locus models result in λ_S values that are significantly different quantitatively from those of the additive two-locus models. The λ_S values for the recessive-recessive model are more variable with disease-allele frequency than the λ_S values for the dominant-dominant model but are at comparable levels for γ values of ~100.

Generalized Expression for λ_S

Table 1 shows the absolute values for the limit of λ_S (for derivation, see the Appendix) for the different genetic models investigated in figures 1–4. In general, the limit of λ_S as $\gamma \to \infty$ increases with decreasing susceptibility-allele frequency, recessive inheritance, and the number of susceptibility loci. However, when two loci have additive effects on disease risk, the limit of λ_S is actually lower than that in the single-locus case. The λ_S limits shown in this table vary greatly, depending on the underlying genetic model and susceptibility-allele frequency. For instance, the λ_S limit for a dominant two-locus additive model with each susceptibility allele at a 50% frequency is 1.07, whereas the λ_S limit for both the two-locus recessive epistatic model (model 2) and the multiplicative model, with susceptibility alleles at 1%, is 6,503,755. In fact, the latter two models have the same λ_S limits at all susceptibility-allele frequencies. The only difference between the two is that the multiplicative model reaches its λ_S limit sooner than does the epistatic model (compare figs. 2*B* and 4*B*). Likewise, the two-locus epistatic model (model 1) and the multiplicative dominant models have virtually the same λ_S limits, with the multiplicative model again reaching its λ_S limit sooner than does the epistatic model.

λ_S *Conditional on an Observed Susceptibility Genotype*

On the basis of the models thus far investigated for a complex disease with one or two susceptibility loci, the estimated value of λ_S has only an indirect correlation with the underlying γ. In situations in which a putative susceptibility locus exists, it may be possible to test whether an elevated λ_S can be explained by an allele at

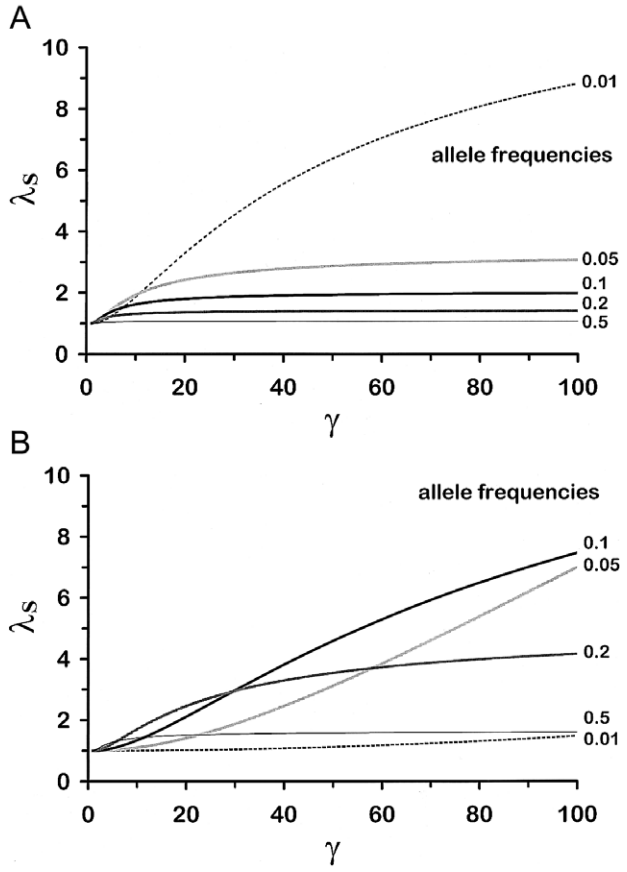

Figure 3 *A,* Relationship between λ_S and γ, for disease-susceptibility-allele frequencies of .01–.5, under a two-locus dominant-additive-inheritance model. *B,* Relationship between λ_S and γ, for disease-susceptibility-allele frequencies of .01–.5, under a two-locus recessive-additive-inheritance model.

this locus. This can be done by restriction of the sibships to those in which the proband has the genotype with suspected increased risk. If the genotype contributes to an increased λ_S, then, when families in which the susceptibility allele likely is not acting on disease risk (i.e., families in which the proband does not have the "at risk" genotype) are eliminated, the overall value in the numerator of λ_S should increase. We denote the sib recurrence-risk ratio for probands having a selected genotype as "λ^*_S." If a putative disease-susceptibility alleles exists, estimates of λ^*_S can serve as a preliminary test of linkage.

Table 2 shows the values of λ_S and λ^*_S for selected inheritance models, for four different combinations of γ and allele frequency. For both single-locus models considered, restricting the probands to those with one or more susceptibility alleles increases the sib recurrence-risk ratio from 10% to >100%. λ^*_S continues to increase when families are restricted to those ascertained through probands with both susceptibility alleles. However, for susceptibility alleles with low ($\leq$10%) frequencies, this would likely require a prohibitively large sample.

The increase of λ^*_S over λ_S when families are restricted to those ascertained through probands with one or more susceptibility alleles, under a two-locus model, depends on the mode of inheritance and type of joint action between the loci. In general, for all three epistatic models, the increase in λ^*_S over λ_S is greater than that for the two single-locus models considered. In some instances, if λ^*_S were not estimated, the evidence for a genetic effect likely would not be detectable. For example, in epistatic model 2, at the allele frequency and γ values shown in table 2, λ_S values are barely >1, but λ^*_S values are generally $\geq$2.

In the two-locus models, the increase in λ^*_S over λ_S is greater in the multiplicative models than in the corresponding additive models. With regard to mode of inheritance, the increase in λ^*_S over λ_S is generally greater

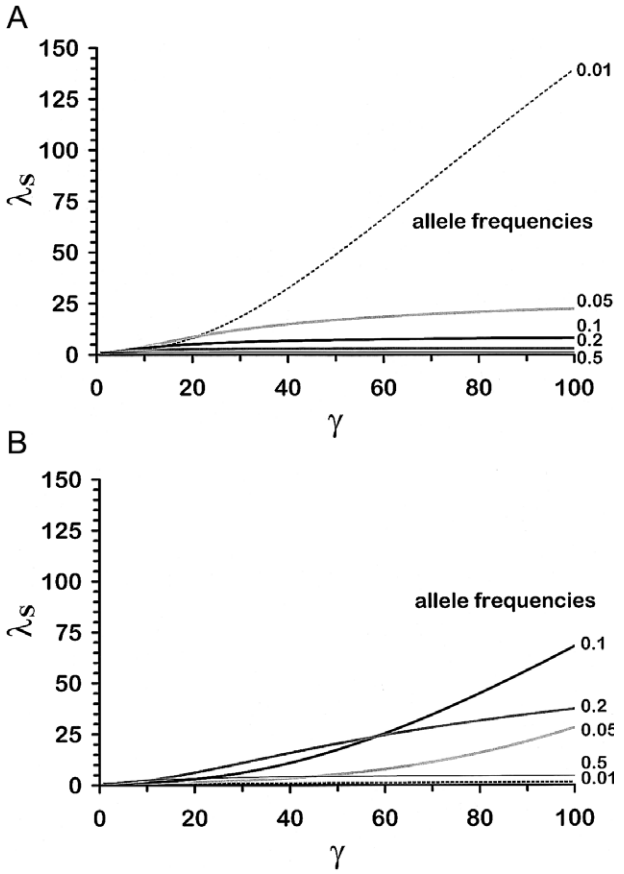

Figure 4 *A,* Relationship between λ_S and γ, for disease-susceptibility-allele frequencies of .01–.5, under a two-locus dominant-multiplicative -inheritance model. *B,* Relationship between λ_S and γ, for disease-susceptibility-allele frequencies of .01–.5, under a two-locus recessive-multiplicative-inheritance model.

Table 1

Limit of λ_S as $\gamma \to \infty$, at Different Disease-Susceptibility-Allele Frequencies

	Limit for Disease-Susceptibility-Allele Frequency of				
Genetic Model	.01	.05	.1	.2	.5
Single-locus models:					
Dominant	25.56	5.57	3.08	1.84	1.14
Recessive	2,550.00	110.25	30.25	9.00	2.25
Two-locus models:					
Epistatic model 1[a]	621.76	30.95	9.45	3.38	1.30
Epistatic model 2[b]	6,503,755.00	12,155.00	915.06	81.00	5.06
Epistatic model 3[c]	33,037.00	326.38	51.98	10.00	1.82
Additive dominant	13.28	3.28	2.04	1.42	1.07
Multiplicative dominant	653.51	31.01	9.46	3.38	1.30
Additive recessive	1,275.00	55.62	15.62	5.00	1.63
Multiplicative recessive	6,503,775.00	12,155.00	915.06	81.00	5.06

[a] Disease risk increases only when at least one susceptibility allele is present at each locus.
[b] Disease risk increases only when both susceptibility alleles are present at each locus.
[c] Disease risk increases when three or more susceptibility alleles are present at both loci.

in recessive models than in dominant models. In these two-locus models, the increase in λ^*_S over λ_S can be striking or modest, depending on the model and underlying genotype effect on disease risk. For example, the λ^*_S for the two-locus dominant additive model in which the proband must have both putative disease-susceptibility alleles is only 26% greater than the standard λ_S when the high-risk alleles have a frequency of 20%. On the other hand, the corresponding λ^*_S for the recessive multiplicative model, when the putative disease allele has a 5% frequency, increases over λ_S anywhere from 100% to 400%, depending on the magnitude of γ. For the two-locus dominant additive and multiplicative models, the increase in λ^*_S over λ_S is generally more modest than that observed in either the single-locus or epistatic models.

Sarcoidosis as an Example

The occurrence of sarcoidosis in African American families can serve as an example of how an estimate of λ^*_S can be applied. Sarcoidosis is a multisystem granulomatous disorder that is of unknown etiology and that occurs more frequently and with more severity in African Americans (Edmondstone and Wilson 1985; Rybicki et al. 1997*a*; Rybicki et al. 1998). Although no major gene has been identified, the clustering of sarcoidosis in families and case-control associations studies support a complex etiology with a significant genetic component (Rybicki et al. 1997*b*). Elsewhere, we have reported tentative associations between the angiotensin-converting enzyme (ACE) D allele on chromosome 17 (Maliarik et al. 1998) and two microsatellite repeat markers, IL-1A*137 and F13A*188 (Rybicki et al. 1999), on chromosomes 2 and 6, respectively. We have found that λ_S for sarcoidosis is ~3.37, on the basis of the ratio of disease prevalence in African American sibs of sarcoidosis cases and the population from which these cases were drawn (B. A. Rybicki and M. C. Iannuzzi, unpublished data). Table 3 shows that the estimate of λ^*_S for the ACE D allele is slightly larger than λ_S when the sibships are restricted to those ascertained through ACE I/D or D/D probands but that it is smaller when only sibships ascertained through D/D probands are considered. Likewise, our estimate of λ^*_S is lower in sibships ascertained through probands with at least one copy of F13A*188. The largest estimate of λ^*_S is for those sibships in which the proband has at least one copy of the IL-1a*137 allele. Although the increases in λ^*_S over λ_S were modest and although, because of the small sample, λ^*_S for F13A*188 and IL-1A*137 homozygotes could not be reliably estimated, under the assumption that all three markers are in complete linkage with the locus in question, these results tend to support the IL-1a locus over the ACE and F13A loci as being a sarcoidosis-susceptibility locus.

Relationship between λ_S, γ, and Allele Sharing

What overall value of λ_S is required for the mapping of disease-susceptibility genes? On the basis of a mean test (Blackwelder and Elston 1985) for the proportion of alleles shared at a disease locus, in a sample of 200 ASPs, conclusive evidence ($p \approx 10^{-4}$) for linkage requires allele sharing of ~59%, whereas a sample of 400 ASPs requires allele sharing of ~56%. In table 4, the expected allele sharing at the disease locus (or loci) is calculated for the nine different genetic models investigated. The susceptibility allele frequency is set at a value such that the population prevalence of disease is ~1/1,000 at low values of γ. For a disease with a single susceptibility allele, a λ_S of either 5.0 for a dominant allele with a

Table 2

Comparison of λ_S and λ^*_S, at Different Levels of γ and p

Genetic Model, γ, and p[a]	λ_S	λ^*_S[b]	λ^*_S[c]
Single-locus dominant:			
γ = 5:			
p = .05	1.27	2.28	2.95
p = .20	1.22	1.50	1.79
γ = 20:			
p = .05	2.82	3.97	5.51
p = .20	1.60	1.73	2.16
Single-locus recessive:			
γ = 5:			
p = .05	1.01	1.16	2.08
p = .20	1.15	1.47	2.10
γ = 20:			
p = .05	1.22	2.83	5.95
p = .20	2.49	3.53	4.45
Epistatic model 1:			
γ = 5:			
p = .05	1.04	1.43	1.63
p = .20	1.28	1.60	1.85
γ = 20:			
p = .05	1.70	4.23	5.67
p = .20	2.21	2.56	3.16
Epistatic model 2:			
γ = 5:			
p = .05	1.00	1.21	1.32
p = .20	1.00	1.97	2.42
γ = 20:			
p = .05	1.00	2.02	2.89
p = .20	1.07	6.04	9.76
Epistatic model 3:			
γ = 5:			
p = .05	1.00	1.02	1.26
p = .20	1.09	1.29	1.72
γ = 20:			
p = .05	1.03	1.30	3.13
p = .20	2.04	2.96	3.99
Two-locus dominant additive:			
γ = 5:			
p = .05	1.44	2.07	2.59
p = .20	1.23	1.38	1.56
γ = 20:			
p = .05	2.42	2.95	3.89
p = .20	1.36	1.49	1.72
Two-locus dominant multiplicative:			
γ = 5:			
p = .05	1.84	3.10	4.01
p = .20	1.67	1.93	2.31
γ = 20:			
p = .05	8.57	11.61	16.12
p = .20	2.69	2.84	3.55
Two-locus recessive additive:			
γ = 5:			
p = .05	1.02	1.17	2.07
p = .20	1.23	1.48	2.00
γ = 20:			
p = .05	1.41	2.86	5.75
p = .20	2.46	2.97	3.50
Two-locus recessive multiplicative:			
γ = 5:			
p = .05	1.02	1.18	2.10
p = .20	1.33	1.69	2.42
γ = 20:			
p = .05	1.49	3.47	7.29
p = .20	6.21	8.79	11.10

[a] Epistatic models are as defined in table 1.
[b] Proband has at least one susceptibility allele at putative disease locus.
[c] Proband has both susceptibility alleles at putative disease locus.

frequency of 1% or a λ_S of ~2.0 for a recessive allele with a frequency of 10% would result in allele sharing of 59% (table 4). In general, both single-locus and two-locus models under a recessive mode of inheritance show a substantial increase in allele sharing at the disease locus, with modest increases in λ_S. For all two-locus models, since the γ for each locus was held equal to that of the other, the allele sharing shown in table 4 is the percent of the disease allele shared at either susceptibility locus. If the underlying genetic model includes two or more recessive alleles acting either additively or multiplicatively, then λ_S values of 3–5 should indicate disease loci that can be mapped on the basis of the expected proportion of allele sharing. In the case of epistasis, allele sharing at a disease locus depends on the effect each disease allele has on the other. For instance, epistatic model 1 requires at least one allele at both loci to be present for disease risk to be increased. In this model, a λ_S of 10 will not, under most circumstances, be large enough to allow mapping of either locus; on the other hand, in epistatic models 2 and 3, a λ_S of 3 will result in an ~.56 allele-sharing proportion at the disease locus, which, provided that both a tightly linked, highly polymorphic marker and 400 ASPs are available, would allow both susceptibility loci to be mapped.

Discussion

Since it was first proposed by Risch (1987, 1990*a*), λ_S has been widely used as a measure of genetic effect to estimate the power of a proposed ASP linkage study (Risch 1990*b*; Holmans 1993; Gu and Rao 1997*a*, 1997*b*). Despite the widespread use of λ_S, the relationship between it and γ has not been clearly defined. In the analyses presented in this report, we have shown that, when one or two loci contribute to an increased disease risk, λ_S will usually be less than the genotype relative risk, γ. In fact, without prior knowledge of the genetic model and allele frequency, it becomes difficult, if not impossible, to use λ_S to infer the magnitude of γ and the proportion of alleles shared, at the disease locus, by affected sibs.

In a response to the article by Risch and Merikangas

Table 3

Comparison of λ_S and λ^*_S, of Sarcoidosis, for Three Putative Susceptibility Alleles in 86 Randomly Ascertained Probands

Proband Sample	λ_S	$\lambda^*_S(1)$[a]	$\lambda^*_S(2)$[a,b]
Entire sample	3.37	...	...
Restricted by *ACE D* allele	...	3.54	2.24
Restricted by *IL-1A*137* allele	...	4.03	NE
Restricted by *F13a*188* allele	...	2.60	NE

[a] As defined in table 2.
[b] NE = Not estimable.

Table 4

Relationship between λ_S, γ, and Allele Sharing at Disease Locus, under Different Genetic Models

Genetic Model, Disease-Allele Frequency, and λ_S[a]	γ	Proportion of Alleles Shared at Disease Locus
Single-locus dominant, frequency .01:		
λ_S = 1.5	10.0	.530
λ_S = 3.0	21.8	.566
λ_S = 5.0	35.8	.595
λ_S = 10.0	79.3	.646
Single-locus recessive, frequency .1:		
λ_S = 1.5	16.0	.552
λ_S = 3.0	36.4	.611
λ_S = 5.0	59.6	.661
λ_S = 10.0	125.6	.746
Epistatic model 1, frequency .01:		
λ_S = 1.5	72.9	.506
λ_S = 3.0	149.0	.513
λ_S = 5.0	215.5	.519
λ_S = 10.0	337.0	.529
Epistatic model 2, frequency .2:		
λ_S = 1.5	54.7	.531
λ_S = 3.0	118.4	.562
λ_S = 5.0	181.0	.588
λ_S = 10.0	316.4	.633
Epistatic model 3, frequency .1:		
λ_S = 1.5	30.7	.530
λ_S = 3.0	67.8	.563
λ_S = 5.0	106.2	.590
λ_S = 10.0	196.9	.637
Two-locus dominant additive, frequency .01:		
λ_S = 1.5	7.4	.519
λ_S = 3.0	18.0	.544
λ_S = 5.0	34.4	.566
λ_S = 10.0	150.4	.604
Two-locus dominant multiplicative, frequency .01:		
λ_S = 1.5	6.3	.517
λ_S = 3.0	11.5	.538
λ_S = 5.0	15.5	.549
λ_S = 10.0	22.2	.567
Two-locus recessive additive, frequency .1:		
λ_S = 1.5	12.3	.535
λ_S = 3.0	30.3	.578
λ_S = 5.0	55.8	.614
λ_S = 10.0	183.0	.675
Two-locus recessive multiplicative, frequency .2:		
λ_S = 1.5	6.0	.548
λ_S = 3.0	11.8	.602
λ_S = 5.0	17.2	.640
λ_S = 10.0	28.1	.694

[a] Epistatic models are as defined in table 1.

(1996) that first described the relationship between λ_S and γ, Scott et al. (1997) briefly expanded on the difference between λ_S and γ, but they gave only one example and did not explore different genetic models and/or allele frequencies. Others have partially described the relationship between genotype relative risk and the recurrence risk in relatives. Neuman and Rice (1992) presented the necessary formulas for calculation of recurrence risk in different classes of relatives, under various two-locus models, on the basis of allele frequency and penetrance parameters. However, most of their models assumed complete penetrance at one locus, and they did not parameterize their results in terms of γ and λ_S. Beaty et al. (1987) described a class of two-locus epistatic models with regard to recurrence risk in relatives, but they also did not explore the relationship between γ and λ_S.

The results that we have presented are supported by examples from complex diseases for which reported genotype-specific relative risks are much larger than sib recurrence-risk ratios. For instance, the λ_S generally reported for rheumatoid arthritis is 2–5 (Rigby et al. 1993), yet γ values as high as 30 have been reported for specific alleles of the HLA-DR (Dieye et al. 1997) and HLA-DRB1 (Hall et al. 1996) loci. One general misconception is that modest levels of λ_S ($\leqslant$10) imply that, if genetic factors play a role, the susceptibility allele(s) must have a reduced penetrance. Whereas this is usually true for recessive inheritance and rare disease alleles, we have shown that a λ_S of <10 can also be the result of single-locus or two-locus additive inheritance when penetrance is virtually complete (i.e., $\gamma \approx \infty$), provided that the susceptibility allele(s) is common ($\geqslant$20%). In such a situation, the value of λ_S is limited—not because the underlying genotypic risk is low, but because of the relatively large proportion of the population affected, which, in turn, inflates the denominator of λ_S (the population prevalence). In table 4, we have shown that allele sharing between affected sibs varies greatly for similar values of λ_S, depending on the underlying genetic model. For example, between affected sibs, disease-allele sharing at proportions $\geqslant$55% is possible for λ_S values as low as 1.5, even in the single-locus dominant model, if the disease allele has a population frequency of $\geqslant$20% (data not shown).

Both the way in which a disease allele is inherited and its joint action with other disease alleles can have a profound effect on the amount of evidence that it generates for linkage. For instance, a considerably higher allele-sharing proportion will occur for a disease under the two-locus recessive-multiplicative-inheritance model, compared with epistatic model 2 (in which two copies of both alleles must be present to increase risk), at comparable levels of λ_S (see table 4). This is true even though the genotypic relative risk of the latter is approximately an order of magnitude higher. As we move from linkage studies to studies designed to identify disease-susceptibility alleles, the genetic effect expected for a susceptibility allele may not be easily predicted from just the original evidence for linkage, even within a general class of inheritance models.

In summary, λ_S and γ have a characteristic relationship that depends on both allele frequency and mode of inheritance. Whereas, theoretically, γ has no upper limit, the value of λ_S is constrained by the disease-susceptibility-allele frequency and the underlying genetic model. If nothing is known about the putative disease allele(s) and its effects, an epidemiologically derived estimate of λ_S cannot by itself justify an ASP linkage study. For instance, we have shown that λ_S values as high as 10 are not always indicative of a disease gene that can be easily localized by linkage. Perhaps either association studies that directly measure γ or segregation analyses that can provide both an indirect estimate of γ and a possible inheritance model may provide better preliminary results for planning an ASP linkage study.

Acknowledgments

We gratefully acknowledge Michael Iannuzzi and his research team, who provided the data used in table 3, and Gary Chase, who provided helpful comments. This work was supported, in part, by National Institutes of Health grants R01 HL54306, U01 HL60263, R01 GM28356, and P41 RR03655.

Appendix A

Single-Locus Model

To illustrate how λ_S is derived for a single-locus dominant model, we must consider first the probability that an affected proband is homozygous for the disease allele. This probability would be $K_G \times p^2$ divided by the population prevalence; label this proband genotypic probability "f_1." The probability that a sib of an affected proband also has the disease is the sum of two separate probabilities: the probability of disease if the sib has the susceptibility genotype and the probability of disease if the sib does not have the susceptibility genotype. If the proband is homozygous for the susceptibility allele, then the probability that a sib of the proband does *not* inherit at least one copy of the allele is $q^2/4$, or, in terms of the frequency of the susceptibility allele, $(p^2 - 2p + 1)/4$. Therefore, the probability that such a sib is affected is $K_0 \times (p^2\text{-}2p + 1)/4$. Alternatively, the probability that a sib of the proband inherits one or two copies of the disease allele is $1 - (p^2 - 2p + 1)/4$, and the probability that this sib with the high-risk genotype is affected is $K_G \times [1 - (p^2 - 2p + 1)/4]$. Summing the two probabilities and multiplying by the proband genotypic probability results in the joint probability that a proband is homozygous for the disease allele and the sib is also affected: $f_1\{K_0(p^2 - 2p + 1)/4 + K_G[1 - (p^2 - 2p + 1)/4]\}$. Corresponding joint probabilities can be calculated for a proband who is heterozygous for the susceptibility allele (label this marginal probability "f_2") and homozygous for the normal allele (label this marginal probability "f_3," so that $f_1 + f_2 + f_3 = 1$). The summation of these three probabilities,

$$f_1\left[K_0\left(\frac{p^2-2p+1}{4}\right)+K_G\left(1-\frac{p^2-2p+1}{4}\right)\right]+f_2\left[K_0\left(\frac{p^2-3p+2}{4}\right)+K_G\left(1-\frac{2-p^2+3p}{4}\right)\right]$$
$$+f_3\left[K_0\left(1-p+\frac{p^2}{4}\right)+K_G\left(p-\frac{p^2}{4}\right)\right]\ , \tag{A1}$$

is the overall risk of disease to a sib of an individual with disease. It follows, then, that λ_S for a disease for which only one dominant susceptibility allele that increases disease risk exists is expression (A1) divided by the population prevalence. The overall risk of disease for a sib of an individual with a disease influenced by a recessively acting allele with frequency p can be computed in the same manner, and the resulting expression is as follows:

$$f_1\left[K_0\left(\frac{3-p^2-2p}{4}\right)+K_G\left(\frac{1+-2p+p^2}{4}\right)\right]+f_2\left[K_0\left(1-\frac{p^2-p}{4}\right)+K_G\left(1-\frac{p^2+p}{4}\right)\right]$$
$$+f_3\left[K_0\left(1-\frac{p^2}{4}\right)+K_G\left(\frac{p^2}{4}\right)\right]\ . \tag{A2}$$

Important to note, with regard to the relationship between λ_S and γ, is that the probability that any randomly sampled person with disease has the susceptibility genotype depends on both the gene frequency and γ. For example, in the case of a single dominant gene, if $p = .1$, $K_0 = 0.001$, and $\gamma = 100$—that is, $K_G = 0.1$—then the probability that an individual with disease does not have the susceptibility genotype is only 4%. If γ decreases to 10, then the probability that an individual with disease does not have the susceptibility genotype increases to 30%.

Two-Locus Model

To explore the relationship between λ_S and γ under a two-locus model, we can expand the expressions for disease prevalence and sib risk. Consider two unlinked diallelic loci with alleles *A,a,* and *B,b,* respectively, where *A* and *B* act dominantly over *a* and *b* to increase risk of disease. If, as before, K_0 is the risk when neither susceptibility allele is present (i.e., the genotype *aabb*), then three risk parameters need to be considered: K_A, the risk of disease when only *A* is present; K_B, the risk of disease when only *B* is present; and K_{AB}, the risk of disease when both *A* and *B* are present. If *A* and *a* have frequencies *p* and *q* and if *B* and *b* have frequencies *r* and *s,* and if, in addition to Hardy-Weinberg equilibrium, there are no allelic associations (and, hence, no linkage disequilibrium between the loci), then the disease prevalence expressed in terms of genotype risks and frequencies is

$$K_0q^2s^2+K_A(p^2+2pq)s^2+K_Bq^2(r^2+2rs)+K_{AB}(p^2+2pq)(r^2+2rs)\ . \tag{A3}$$

The numerator for λ_S in the two-locus model consists of nine separate joint probabilities corresponding to the nine possible genotypes for two diallelic loci. If the nine proband genotypic probabilities are labeled "f_1"–"f_9," and sum to 1, as in the single locus model, then a similar probability for the risk of disease in the sib of a proband can be constructed. For an affected sib of a proband, nine separate probabilities exist for each of the nine possible genotypes of the proband. Let g_{ij} be the probability that an affected sib of a proband with genotype *i* possesses genotype *j,* where $i, j = 1, 2, \ldots, 9$. Then the joint probability for a proband with the *AABB* genotype and an affected sib is

$$f_1[K_0g_{19}+K_A(g_{17}+g_{18})+K_B(g_{13}+g_{16})+K_{AB}(g_{11}+g_{12}+g_{14}+g_{15})]\ .$$

The overall risk of disease for a sib of an affected person is the sum of the nine joint probabilities corresponding to each possible genotype:

$$\sum_{i=1}^{9} f_1[K_0g_{i9}+K_A(g_{i7}+g_{i8})+K_B(g_{i3}+g_{i6})+K_{AB}(g_{i1}+g_{i2}+g_{i4}+g_{i5})]\ . \tag{A4}$$

Dividing expression (A4) by expression (A3), one has the value of the overall λ_S for a disease with two unlinked

loci in which each has one dominantly acting susceptibility allele. As with the single-locus model, this expression can be easily modified for models with different genotypic risks (e.g., a two-locus recessive model).

Generalized Expression for λ_S

The expression for λ_S can be generalized for any situation in which susceptibility genotypes exist. As described earlier, let p_G be the probability of a susceptibility genotype and let $p_0 = 1 - p_G$ be the probability of a nonsusceptible genotype. Then in the single-locus models, for example, we can substitute $K_G = gK_0$ so that expressions (A1) and (A2), the numerators of λ_S under dominant and recessive inheritance, respectively, are both of the form

$$\sum_{i=1}^{3} \frac{\gamma K_0 a_i}{p_0 + \gamma p_G}(b_i + \gamma c_i) \ , \tag{A5}$$

In expression (A5), a_i, b_i, and c_i are defined as follows: a_i = prior probability of the ith genotype on the basis of allele frequencies; b_i = the probability of a sib having a nonsusceptible genotype, given that the proband has the ith genotype; c_i = the probability of a sib having a susceptibility genotype, given that the proband has the ith genotype.

By definition, the $\Sigma a_i = 1$, and, for any genotype i, $b_i + c_i = 1$. The denominator of λ_S is $K_0 p_0 + gK_0 p_G$, so that, canceling out K_0, we can write λ_S in the form

$$\frac{1}{(p_0 + \gamma p_G)^2} \sum_{i=1}^{3} \gamma a_i (b_i + \gamma c_i) \ . \tag{A6}$$

For the two-locus models that we have considered, three relative risks exist—γ_A, γ_B and γ_{AB}, which we rewrite as $\theta_A\gamma$, $\theta_B\gamma$, and γ, respectively. Dividing expression (A4) by expression (A3), we can then write λ_S for the two-locus models, in the form

$$\frac{1}{(p_0 + \theta_A\gamma p_A + \theta_B\gamma p_B + \gamma p_G)^2} \sum_{i=1}^{9} \gamma a_i (b_i + \theta_A\gamma c_i + \theta_B\gamma d_i + \gamma e_i) \ , \tag{A7}$$

where $p_0 + p_A + p_B + p_{AB} = 1$. In the expression for the two-locus model, a_i and b_i represent the same values as in the single-locus model, however, the c_i term is now expanded to three terms—c_i, d_i, and e_i—that are defined as follows: c_i = the probability of a sib having a genotype with only susceptibility allele A, given that the proband has the ith genotype; d_i = the probability of a sib having a genotype with only susceptibility allele B, given that the proband has the ith genotype; e_i = the probability of a sib having a genotype with both susceptibility alleles A and B, given that the proband has the ith genotype.

From this framework, the limit of λ_S as $\gamma \to \infty$ can be easily derived. For $\gamma \to \infty$ in expression (A6), for the single-locus model, we have

$$\lim_{\gamma \to \infty} \lambda_S = \frac{1}{p_G^2} \sum_{i=1}^{3} a_i c_i \ ,$$

and, similarly, from expression (A7), for the two-locus models, we have

$$\lim_{\gamma \to \infty} \lambda_S = \frac{1}{(\theta_A p_A + \theta_B p_B + p_{AB})^2} \sum_{i=1}^{9} a_i (\theta_A c_i + \theta_B d_i + e_i) \ .$$

References

Beaty TH, Maestri NE, Meyers DA, Murphy EA (1987) Predicting recurrence risks under epistatic models. Am J Med Genet 28:631–645

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

Dieye A, Diallo S, Diatta M, Thiam A, Ndiaye R, Bao O, Sarthou JL (1997) Identification of HLA-DR alleles for sus-

ceptibility to rheumatoid polyarthritis in Senegal. Dakar Med 42:111–113

Edmondstone WM, Wilson AG (1985) Sarcoidosis in Caucasians, Blacks and Asians in London. Br J Dis Chest 79: 27–36

Gu C, Rao DC (1997*a*) A linkage strategy for detection of human quantitative-trait loci. I. Generalized relative risk-ratios and power of sib pairs with extreme trait values. Am J Hum Genet 61:200–210

——— (1997*b*) A linkage strategy for detection of human quantitative-trait loci. II. Optimization of study designs based on extreme sib pairs and generalized relative risk ratios. Am J Hum Genet 61:211–222

Hall FC, Weeks DE, Camilleri JP, Williams LA, Amos N, Darke C, Gibson K, et al (1996) Influence of the HLA-DRB1 locus on susceptibility and severity in rheumatoid arthritis. QJM 89:821–829

Holmans P (1993) Asymptotic properties of affected-sib-pair linkage analysis. Am J Hum Genet 52:362–374

Maliarik MJ, Rybicki BA, Malvitz E, Sheffer RG, Major M, Popovich JJ, Iannuzzi MC (1998) Angiotensin-converting enzyme gene polymorphism and risk of sarcoidosis. Am J Respir Crit Care Med 158:1566–1570

Neuman RJ, Rice JP (1992) Two-locus models of disease. Genet Epidemiol 9:347–365

Rigby AS, Voelm L, Silman AJ (1993) Epistatic modeling in rheumatoid arthritis: an application of the Risch theory. Genet Epidemiol 10:311–320

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. II. The power of affected relative pairs. Am J Hum Genet 46:229–241

——— (1990*c*) 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

Risch N, Merikangas K (1996) The future of genetic studies of complex human diseases. Science 273:1516–1517

Rybicki BA, Major M, Popovich JJ, Maliarik MJ, Iannuzzi MC (1997*a*) Racial differences in sarcoidosis incidence: a 5-year study in a health maintenance organization. Am J Epidemiol 145:234–241

Rybicki BA, Maliarik MJ, Major M, Popovich JJ, Iannuzzi MC (1997*b*) Genetics of sarcoidosis. Clin Chest Med 18: 707–717

——— (1998) Epidemiology, demographics, and genetics of sarcoidosis. Semin Respir Infect 13:166–173

Rybicki BA, Maliarik MJ, Malvitz E, Sheffer RG, Major M, Popovich J Jr, Iannuzzi MC (1999) The influence of T cell receptor and cytokine genes on sarcoidosis susceptibility in African Americans. Hum Immunol 60:867–874

Scott WK, Pericak-Vance MA, Haines JL (1997) Genetic analysis of complex diseases. Science 275:1327–1330