Am. J. Hum. Genet. 61:200–210, 1997

A Linkage Strategy for Detection of Human Quantitative-Trait Loci. I. Generalized Relative Risk Ratios and Power of Sib Pairs with Extreme Trait Values

Chi Gu[1] and D. C. Rao[1,2]

[1]Division of Biostatistics and [2]Departments of Psychiatry and Genetics, Washington University School of Medicine, St. Louis

Summary

We generalize the concept of the relative risk ratio (λ) to the case of quantitative traits, to take into account the various trait outcomes of a relative pair. Formulas are derived to express the expected proportion of genes shared identical by descent by a sib pair, in terms of the generalized λ's for sib pairs (λ_S), parent-offspring pairs (λ_O), and monozygotic twins (λ_M) and in terms of the recombination fraction, with the assumption of no residual correlations. If residual correlations are nonzero among relative pairs, we assume that they are the same among sib pairs, parent-offspring pairs, and monozygotic twins, and we employ a slightly different definition for the generalized λ so that the same set of formulas still hold. The power (or, the sample size necessary) to detect quantitative-trait loci (QTLs) by use of extreme sib pairs (ESPs) is shown to be a function of the three generalized λ's. Since λ_M can be derived by use of values of λ_S and λ_O, estimates of the latter two λ's will suffice for the analysis of power and the necessary sample sizes of ESPs, for a QTL linkage study.

Introduction

For dichotomous qualitative traits, Risch (1990*a*) introduced the concept of the relative risk ratio (λ) and demonstrated how it determined the statistical power to detect linkage by use of affected relative pairs (including sib pairs). Many important human traits (such as blood pressure and body-mass index) are, however, of a quantitative nature, and the power to detect linkage to quantitative-trait loci (QTLs) is rather low in unselected samples. The enhanced power of sib pairs with extreme phenotypic values has been discussed recently by several authors (Carey and Williamson 1991; Fulker et al. 1991; Eaves and Meyer 1994; Risch and Zhang 1995; Gu et al. 1996). As demonstrated by Risch and Zhang (1995), the gain in power is dramatic when extremely discordant (ED) sib pairs are used. In a previous article (Gu et al. 1996), we discussed some practical issues of applying Risch and Zhang's method, especially the difficulty of finding ED pairs, and suggested that it may be more cost effective to combine both the ED pairs and the extremely concordant (EC) sib pairs that are available in the sampling pool used for identifying the ED pairs. In the current series of two articles (also see Gu and Rao 1997 [in this issue]), we aim to answer the question of when and how to combine ED and EC sib pairs, by generalizing the concept of λ to quantitative traits.

When there is no residual correlation between relatives, in addition to that due to the trait locus under consideration, simple formulas for expected identical (identity)-by-descent (IBD) proportions are derived in terms of the generalized λ's and the recombination fraction (θ). When there is nonzero residual correlation between relatives, we assume that the residual correlations among sib pairs, parent-offspring pairs, and monozygotic twins are the same, and we obtain the same set of formulas by modifying slightly the definition of λ. Using estimated values of the λ's, we can calculate either the power of extreme sib pairs (ESPs) for given numbers of ED and/or EC sib pairs or the necessary sample sizes of ED/EC pairs for a preset power. The optimization of combining ED and EC pairs is dealt with in the subsequent paper of this series (Gu and Rao 1997).

We begin with a definition of $\lambda_R(h,l)$, the generalized λ for a type-R relative pair with trait outcomes (h,l), and develop its relationship with the expected IBD proportions for sib pairs. The effect of θ is incorporated by the expression of the probability of IBD at the linked trait locus, conditioned on IBD at the marker locus. Basic properties of $\lambda_R(h,l)$ are explored by the study of its distribution among sib pairs with different trait outcomes and by comparison of its values under different genetic models. We then give formulas to calculate the power and the necessary sample sizes of ESPs, using (estimated) values of $\lambda_S(h,l)$ (for sib pairs) and $\lambda_O(h,l)$

Received June 5, 1996; accepted for publication April 17, 1997.

Address for correspondence and reprints: Dr. Chi Gu, Division of Biostatistics, Washington University School of Medicine, Box 8067, 660 South Euclid Avenue, St. Louis, MO 63110. E-mail: gc@wubios.wustl.edu

0002-9297/97/6101-0027$02.00

(for parent-offspring pairs). Both Risch and Zhang's ED sib-pair test and our combined ED and EC sib-pair test, called "the EDAC test" (Gu et al. 1996), are discussed. Note that with this methodology, the estimation of power (or sample sizes) depends on the values of the λ's, not directly on the underlying model. The estimated power and necessary sample sizes of ESPs are given for a grid of (λ_O,λ_S).

Methods

Let us consider the following model for a quantitative phenotype X:

$$X = \mu + g + e, \tag{1}$$

where μ is the overall phenotypic mean and the major genetic effect g is generated by a biallelic locus, while the residual term e, which is uncorrelated with g, encompasses any multifactorial inheritance and pure error. The alleles A_1 (the so-called risk allele) and A_2 at the major locus are assumed to have frequencies p and $q = 1 - p$, respectively, with the genotypic effects given by

$$g = \begin{cases} -a \text{ for } A_2A_2 \\ d \text{ for } A_1A_2 \\ a \text{ for } A_1A_1 \end{cases}.$$

For simplicity, we assume that e has variance $\sigma_e^2 = 1$. The e is allowed to be correlated among relatives, but we make the assumption that sib pairs, parent-offspring pairs, and monozygotic twins all share the same residual correlation ρ. Relaxation of this assumption will be discussed later. Without loss of generality, we assume that high trait values are associated with an increased risk of disease and divide the trait values into 10 consecutive subintervals with equal probabilities (or deciles), unless otherwise noted. Sib pairs with one member sampled from each extreme decile are called "ED sib pairs" and (10, 1) is used to denote their trait values. Sib pairs with both members from the same extreme decile are called "EC sib pairs." If both siblings have trait values in the highest decile, the sib pair is called an "extremely high–concordant (HC) sib pair," and, similarly, if both siblings have trait values in the lowest decile, the sib pair is called an "extremely low–concordant (LC) sib pair." Their trait values are denoted by (10,10) and (1,1), respectively. Generally, we use $(h,l)_R$ to denote a pair of relatives of type R, with its members' trait values in the hth and lth deciles. $P[(h,l)_R]$, also denoted as $P_R(h,l)$ when there is no ambiguity, is the probability of this trait outcome. The subscript "$_R$" denotes the type of relative. We will use "$_S$" for sib pairs, "$_O$" for parent-offspring pairs, and "$_M$" for monozygotic twins. Moreover, for the power calculations in this paper, we assume that the selective sampling method through probands at high risk is used, whereby one member of each sib pair comes from the highest decile.

Generalized λ's and IBD Sharing

For a qualitative trait, James (1971) introduced a simple relationship between the relative recurrence risk K_R and the population prevalence K. Risch (1990*a*) simplified the formula by introducing the concept of relative risk ratio, $\lambda_R = K_R/K$:

$$\lambda_R = 1 + \frac{1}{K^2}\,\mathrm{Cov}(X_1,X_2)\,, \tag{2}$$

where $\mathrm{Cov}(X_1,X_2)$ is the covariance between the trait values (0 or 1) of a relative pair. From this formula, he explored the relationship between different types of relative pairs and used it for predicting genetic models. Furthermore, the expected IBD proportions of various types of relatives were expressed in terms of λ_S and λ_O, the λ's for sib pairs and parent-offspring pairs, and thus power calculations for sib-pair tests became tractable.

For continuous quantitative traits, equation (2) is not defined readily, but, as we show here, the concept of λ can be generalized easily. Although the relationship between the λ's for different types of relatives is less clear, the role of λ in determining the power of sib-pair tests still holds and becomes more useful in light of the increasing need for combining different types of ESPs (Gu et al. 1996).

$\rho = 0$.—Let us first consider the case when there is no residual correlation. Using the notation described earlier, we define the generalized λ for a type-R relative pair as a function of the trait outcomes of the relative pair:

$$\lambda_R(h,l) = \frac{K_R(l\,|\,h)}{K(l)}\,, \tag{3}$$

where $K(l)$ is the probability that a randomly selected person has a trait value in the lth decile, and $K_R(l\,|\,h)$ is the probability that a person has a trait value in the lth decile given that the trait value of his/her type-R relative is in the hth decile. That is, $K(l) = P(X_2 \in l) = P(l)$ and $K_R(l\,|\,h) = P(X_2 \in l\,|\,X_1 \in h) \equiv P_R(l\,|\,h)$, where X_1 and X_2 are the trait values of the relative pair. Thus, if we divide the trait values into 10 deciles with equal probability, then $K_R(l\,|\,h) = 10 \times P[(h,l)_R]$ (since $P(X_1 \in h) = .1$ for any $h = 1, \ldots, 10$). Depending on our interest, a lower or a higher trait value may be associated with the risk of some disease, but we still use the term "risk

ratio," for apparent reasons. Notice that λ now is a function of the relative pair's trait outcomes and that it also depends (implicitly) on the way the trait values are divided into intervals.

Theoretically, if the model parameters are known, the generalized λ's can be calculated easily. We derive the following formula in terms of D_i, the probability that a sib pair has outcome (h,l) and shares i alleles IBD: $D_i = P[(h,l)_S \text{ and } \pi = i]$. Given a particular genetic model, D_i may be calculated as a sum of the products of conditional probabilities of the trait outcomes, given the genotypes of a sib pair, and of probabilities of the genotypes conditioned on IBD at the trait locus (e.g., see Risch and Zhang 1995).

Since $\rho = 0$, it is easy to see that $P[(h,l)_S|\pi = 0] = P(h)P(l)$; thus we have

$$\lambda_S(h,l) = \frac{P_S(h,l)}{P(h)P(l)} = \frac{D_0 + D_1 + D_2}{4D_0}. \quad (4)$$

Similarly, for the same reason, we have $P_O(h,l) = P[(h,l)_S|\pi = 1]$ and $P_M(h,l) = P[(h,l)_S|\pi = 2)]$. Therefore,

$$\begin{aligned}\lambda_O(h,l) &= \frac{K_O(l|h)}{K(l)} = \frac{P_O(h,l)}{P(h)P(l)} \\ &= \frac{P[(h,l)_S|\pi = 1]}{P[(h,l)_S|\pi = 0]} = \frac{D_1/1/2}{D_0/1/4} = \frac{D_1}{2D_0},\end{aligned} \quad (5)$$

and

$$\lambda_M(h,l) = \frac{D_2}{D_0}. \quad (6)$$

Then, it is easy to verify that

$$\lambda_M(h,l) = 4\lambda_S(h,l) - 2\lambda_O(h,l) - 1 \quad (7)$$

holds for all valid h and l.

Let us now express the expected IBD sharing between relative pairs in terms of the generalized λ's. First, let us consider the case in which the marker is completely linked to the trait locus ($\theta = 0$). We denote the probability that a type-R relative pair shares i alleles IBD, given the trait outcome $(h,l)_R$, by $Z_{R,i}$. We have

$$\begin{aligned}Z_{R,0}(h,l) &= P[\pi = 0|(h,l)_R] \\ &= P(\pi = 0)\frac{P[(h,l)_R|\pi = 0]}{P_R(h,l)} \\ &= P(\pi = 0)\frac{P(h)P(l)}{P(h)P_R(l|h)} \\ &= P(\pi = 0)\frac{P(l)}{P_R(l|h)} \\ &= P(\pi = 0)/\lambda_R(h,l).\end{aligned} \quad (8)$$

Similarly,

$$Z_{R,1}(h,l) = P(\pi = 1)\frac{\lambda_O(h,l)}{\lambda_R(h,l)}, \quad (9)$$

and

$$Z_{R,2}(h,l) = P(\pi = 2)\frac{\lambda_M(h,l)}{\lambda_R(h,l)}. \quad (10)$$

To accommodate the effect of recombination ($\theta > 0$), let $\Psi = \theta^2 + (1-\theta)^2$. We apply the conditional IBD probability for the linked trait locus, given IBD at the marker locus, as displayed in table 1 (Suarez et al. 1978; Risch 1990*b*) where π_t and π_m denote the IBDs at the trait locus and at the marker locus, respectively. Therefore, we have

$$\begin{aligned}Z_{S,0}(h,l) &= \frac{P(\pi_m = 0)}{P[(h,l)_S]}P[(h,l)_S|\pi_m = 0] \\ &= \frac{P(\pi_m = 0)}{P(h)P_S(l|h)}\sum_{j=0}^{2}P[(h,l)_S|\pi_t = j] \\ &\quad \times P(\pi_t = j|\pi_m = 0) \\ &= \frac{P(\pi_m = 0)}{P(h)P_S(l|h)}[P(h)P(l)\Psi^2 + 2\Psi(1-\Psi) \\ &\quad \times P(h)P_O(l|h) + (1-\Psi)^2P(h)P_M(l|h)] \\ &= \frac{1}{4}\left[\Psi^2\frac{1}{\lambda_S(h,l)} + 2\Psi(1-\Psi)\frac{\lambda_O(h,l)}{\lambda_S(h,l)}\right.\end{aligned}$$

Table 1

Conditional Probability of π_t, Given π_m

	Conditional Probability for π_t = [a]		
π_m	0	1	2
0	Ψ^2	$2\Psi(1-\Psi)$	$(1-\Psi)^2$
1	$\Psi(1-\Psi)$	$\Psi^2 + (1-\Psi)^2$	$\Psi(1-\Psi)$
2	$(1-\Psi)^2$	$2\Psi(1-\Psi)$	Ψ^2

[a] $\Psi = \theta^2 + (1-\theta)^2$.

$$+ (1-\Psi)^2 \frac{\lambda_M(h,l)}{\lambda_S(h,l)}\Bigg]$$
$$= \frac{1}{4} - \frac{1}{4}\left[\frac{1}{\lambda_S(h,l)}\right](2\Psi-1)\{[\lambda_S(h,l)-1] + 2(1-\Psi)[\lambda_S(h,l)-\lambda_O(h,l)]\}. \quad (11)$$

Similarly,

$$Z_{S,1} = \frac{1}{2} - \frac{1}{2}(2\Psi-1)^2 \frac{1}{\lambda_S(h,l)} \times [\lambda_S(h,l) - \lambda_O(h,l)], \quad (12)$$

and

$$Z_{S,2} = \frac{1}{4} + \frac{1}{4}(2\Psi-1)\frac{1}{\lambda_S(h,l)}\{[\lambda_S(h,l)-1] + 2\Psi[\lambda_S(h,l)-\lambda_O(h,l)]\}. \quad (13)$$

When θ is small, $\Psi \approx 1$; then, $\Psi(\lambda_S - 1) \approx \lambda_S - 1$, and one may approximate equation (13) by the following:

$$\begin{aligned} Z_{S,2} &\approx \frac{1}{4} + \frac{1}{4}(2\Psi-1)\frac{1}{\lambda_S(h,l)}\{\Psi[\lambda_S(h,l)-1] + 2\Psi[\lambda_S(h,l)-\lambda_O(h,l)]\} \\ &= \frac{1}{4} + \frac{1}{4}\Psi(2\Psi-1)\frac{1}{\lambda_S(h,l)} \times [3\lambda_S(h,l) - 2\lambda_O(h,l) - 1] \\ &= \frac{1}{4} + \frac{1}{4}\Psi(2\Psi-1)\frac{1}{\lambda_S(h,l)} \times [\lambda_M(h,l) - \lambda_S(h,l)]. \end{aligned} \quad (14)$$

$\rho > 0$.—If $\rho > 0$, we assume that sib pairs, parent-offspring pairs, and monozygotic twins all share the same ρ, as mentioned previously. Under this assumption, we see, for example, that the probability of a parent-offspring pair having some particular trait outcome is equal to that of a sib pair having the same trait outcome, given that they share exactly one allele IBD at the trait locus. And the probability that any type of relative pair has trait outcome $(h,l)_R$, given that the relatives share no allele IBD at the trait locus, can be written as $P[(h,l)_R \mid \pi = 0] = P(h)P_r(l \mid h)$, where $P_r(l \mid h)$ ("r" denotes residual) is the probability that a person's trait value is in the lth decile, given that a person who only is residually correlated with him/her has a trait value in the hth decile. In view of this derivation, we modify the definition of the generalized λ as

$$\lambda_R(h,l) = \frac{K_R(l \mid h)}{P_r(l \mid h)}. \quad (15)$$

Note that $P_r(l \mid h) = P(l)$ when $\rho = 0$. It is easy to check that all the formulas derived previously still hold in this case simply by the replacement of $P(l)$ with $P_r(l \mid h)$. In particular, those formulas that calculate the values for $\lambda_R(h,l)$, when the parameters of the underlying genetic models are given, still hold.

In practice, the estimation of $P_r(l \mid h)$ may be a far more complicated task than the estimation of the simple population prevalence $P(l)$. However, as far as this article is concerned, we are interested only in the estimation of the probabilities of a sib pair sharing one or two genes IBD, given various trait outcomes, which in turn will determine the power of the linkage test, as we will show below. Therefore, from equations (9) and (10), only the estimation of the *ratios* of the λ's is necessary when $\theta = 0$. If $\theta > 0$ is small, it is still true, by use of the approximation in equation (13). These ratios—more specifically, $\lambda_O(h,l)/\lambda_S(h,l)$ and $\lambda_M(h,l)/\lambda_S(h,l)$—can be written as ratios of the recurrence risks $K_S(l \mid h)$, $K_O(l \mid h)$, and $K_M(l \mid h)$. This means that estimation of $K_S(l \mid h)$, $K_O(l \mid h)$, and $K_M(l \mid h)$ would be practically sufficient for power and sample-size calculations. Hence, the direct estimation of $P_r(l \mid h)$ could be avoided.

Power and Sample Size for Sib-Pair Tests

For sib pairs with any type of phenotypic configuration (h,l), the above results express the expected IBD sharing in terms of two parameters, $\lambda_S(h,l)$ and $\lambda_O(h,l)$. Thus, we may calculate the power of a sib-pair test, for a given number of sib pairs, or estimate the number of various types of sib pairs necessary to achieve a desired statistical power.

Let us first consider Risch and Zhang's (1995) ESP test $\bar{X}$: $\bar{X} = (1/n_1)\sum_{i=1}^{n_1} \pi_{1i}$, where π_{ki} is IBD of the ith sib pair, with k sibs having extremely high trait values (the other sibs having extremely low trait values) and where n_k is the number of such sib pairs. The power of n sib pairs of type ED (i.e., [10,1]), for example, can be calculated by use of $\lambda_S(10,1)$ and $\lambda_O(10,1)$, by $\Phi(Z_{1-\beta})$, where Φ is the cumulative distribution function of the standard normal and where

$$Z_{1-\beta} = \frac{|2\lambda_S(10,1) - \lambda_O(10,1) - 1|\sqrt{n} + Z_\alpha \lambda_S(10,1)\sqrt{2}}{\sqrt{[\lambda_O(10,1)+2][2\lambda_S(10,1) - \lambda_O(10,1)] - 1}}. \quad (16)$$

The necessary sample size of ED sib pairs for a desired power of $\Phi(Z_{1-\beta}) = 1 - \beta$ can be estimated via the λ's, by

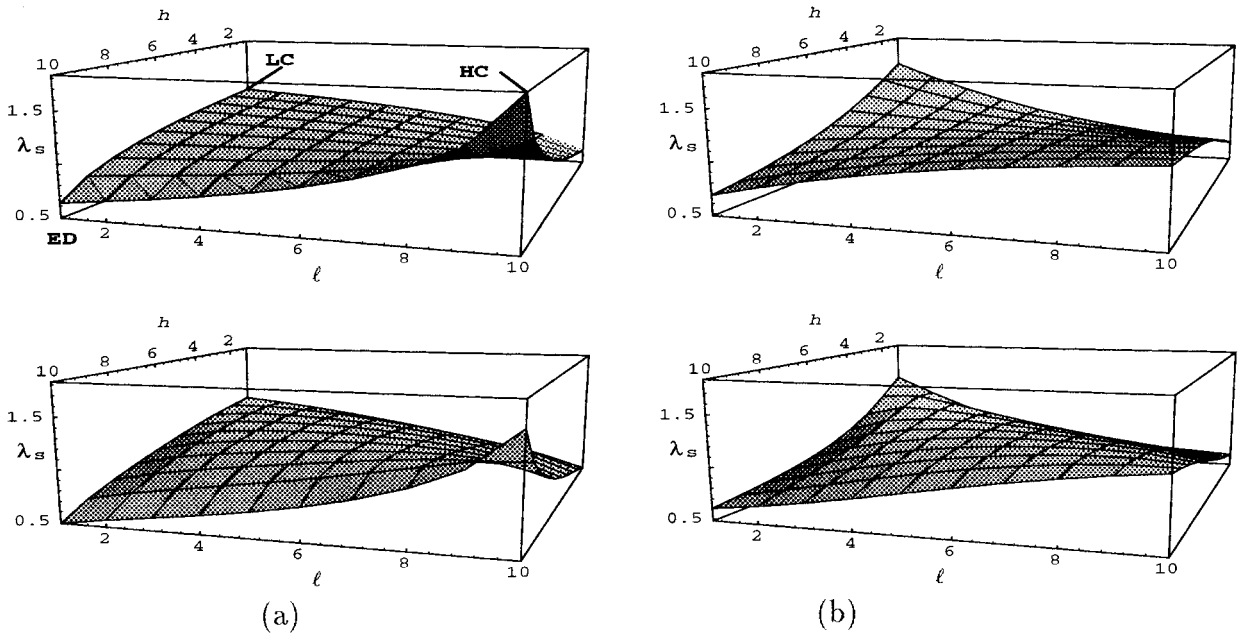

Figure 1 Distribution of λ_S over sib-pair trait outcomes (h,l), under different genetic models. *a*, Additive models with $p = .2$. *b*, Dominant models with $p = .4$. All models have heritability equal to .3, whereas the top row has $\rho = 0$ and the bottom row has $\rho = .4$.

$$n_1 = \left\{\frac{Z_{1-\beta}\sqrt{[\lambda_O(10,1) + 2][2\lambda_S(10,1) - \lambda_O(10,1)] - 1} - Z_\alpha\lambda_S(10,1)\sqrt{2}}{2\lambda_S(10,1) - \lambda_O(10,1) - 1}\right\}^2 . \quad (17)$$

When there is substantial linkage information carried by the EC sib pairs in the sample, we have proposed to combine them (the HC or LC sib pairs or both) with the ED sib pairs, using the EDAC test of Gu et al. (1996). For instance, if one would combine all the available HC (i.e., [10,10]) and LC (i.e., [1,1]) pairs with the ED pairs, the EDAC statistic would be

$$T_{\text{EDAC}} = \frac{1}{n_2 + n_0}\left(\sum_{i=1}^{n_2}\pi_{2i} + \sum_{i=1}^{n_0}\pi_{0i}\right) - \frac{1}{n_1}\sum_{i=1}^{n_1}\pi_{1i} .$$

The power of a sample with n_2 HC pairs, n_1 ED pairs, and n_0 LC pairs is given by $\Phi(Z_{1-\beta})$, with

$$Z_{1-\beta} = \frac{\tau' - \sqrt{\frac{n_1 + n_2 + n_0}{2n_1(n_2 + n_0)}}(Z_\alpha/2)}{\sigma'} , \quad (18)$$

where Φ is the cumulative distribution function of the standard normal and where

$$\tau' = \frac{n_2}{n_2 + n_0}\left[\frac{4\lambda_S(10,10) - \lambda_O(10,10) - 1}{\lambda_S(10,10)}\right] + \frac{n_0}{n_2 + n_0}\left[\frac{4\lambda_S(1,1) - \lambda_O(1,1) - 1}{\lambda_S(1,1)}\right] - \frac{4\lambda_S(10,1) - \lambda_O(10,1) - 1}{\lambda_S(10,1)} ,$$

$$f_0 = \frac{[\lambda_O(1,1) + 2][2\lambda_S(1,1) - \lambda_O(1,1) - 1]}{8\lambda_S(1,1)^2} ,$$

$$f_1 = \frac{[\lambda_O(10,1) + 2][2\lambda_S(10,1) - \lambda_O(10,1) - 1]}{8\lambda_S(10,1)^2} ,$$

$$f_2 = \frac{[\lambda_O(10,10) + 2][2\lambda_S(10,10) - \lambda_O(10,10) - 1]}{8\lambda_S(10,10)^2} ,$$

and

$$\sigma'^2 = \frac{1}{2(n_2 + n_0)}\left(\frac{n_2}{n_2 + n_0}f_2 + \frac{n_0}{n_2 + n_0}f_0\right) + \frac{1}{2n_1}f_1 .$$

To achieve a desired power $1 - \beta$, the necessary sample sizes for each type of ESP (HC, ED, and LC) are derived by the solving of equation (18) after the ratios between the different types of sib pairs to be used in the analysis are determined. An optimization procedure may be used to obtain the ratios for the most cost-effective design (see Gu and Rao 1997).

Results

The distributions of λ_S for sib pairs with various trait outcomes, under different additive and dominant mod-

els, are displayed in figure 1. Positive ρ's generally decrease the value of λ_S and, with regard to the detection of linkage, result in better power for ED pairs and less power for HC pairs. This is consistent with the power analysis done by Risch and Zhang (1995) and that done by Gu et al. (1996). The values of λ_S for LC sib pairs are close to 1 for both additive models, which indicates that the LC pairs have little power in the detection of genes with additive effects. As a comparison, under the dominant models (fig. 1*b*), we see much larger values of λ_S for LC pairs and smaller values for HC pairs. This implies that, under some dominant models, LC sib pairs may add more power.

From equations (8)–(14), we see that the power of ESP tests actually depends on the ratios λ_O/λ_S and λ_M/λ_S. In table 2, we present the ratios for various additive and dominant models, together with τ, the expected IBD sharing for sib pairs (i.e., $\frac{1}{2}Z_{S,1} + Z_{S,2}$). We see that, for example, as ρ changes from 0 to .4, τ for ED pairs moves further away from $\frac{1}{2}$ (the value expected under the null hypothesis of no linkage), whereas τ for HC pairs moves closer to $\frac{1}{2}$. For traditional λ's (defined only for affected sib pairs that are, in many cases, close to what we call "HC pairs," in this article), one would have $\lambda_O \leq \lambda_S$ (i.e., $\lambda_O/\lambda_S \leq 1$), and their values would be nearly equal under the dominant models (James 1971; Risch 1990*a*). The same relationship holds for HC pairs, by use of the generalized λ's, but it is not surprising that it no longer holds for other types of sib pairs (see fig. 2). For ED sib pairs, one might expect the inequality to be in the opposite direction, but we see in figure 2*b* that this is not always true either.

Table 2

Patterns of Expected Values of τ for Sib Pairs, with Corresponding Ratios of λ's, Corresponding to Different Genetic Models

	ED Pairs			HC Pairs		
p	λ_O/λ_S	λ_M/λ_S	τ	λ_O/λ_S	λ_M/λ_S	τ
	Additive Model with ρ = 0					
.20	.969	.466	.359	.978	1.496	.618
.40	.943	.467	.352	.971	1.417	.597
.60	.943	.467	.352	.980	1.315	.574
.80	.969	.466	.359	.992	1.187	.545
	Additive Model with ρ = .4					
.20	.866	.194	.265	.981	1.407	.597
.40	.811	.201	.253	.980	1.334	.578
.60	.811	.201	.253	.986	1.261	.562
.80	.866	.194	.265	.994	1.167	.540
	Dominant Model with ρ = 0					
.10	1.014	.454	.367	.986	1.544	.633
.30	1.049	.390	.360	.976	1.295	.568
.50	1.086	.400	.371	.977	1.158	.534
.70	1.103	.515	.405	.987	1.060	.512
	Dominant Model with ρ = .4					
.10	1.021	.163	.296	.988	1.469	.614
.30	1.068	.161	.307	.979	1.263	.561
.50	1.122	.147	.317	.979	1.150	.532
.70	1.180	.152	.333	.987	1.060	.512

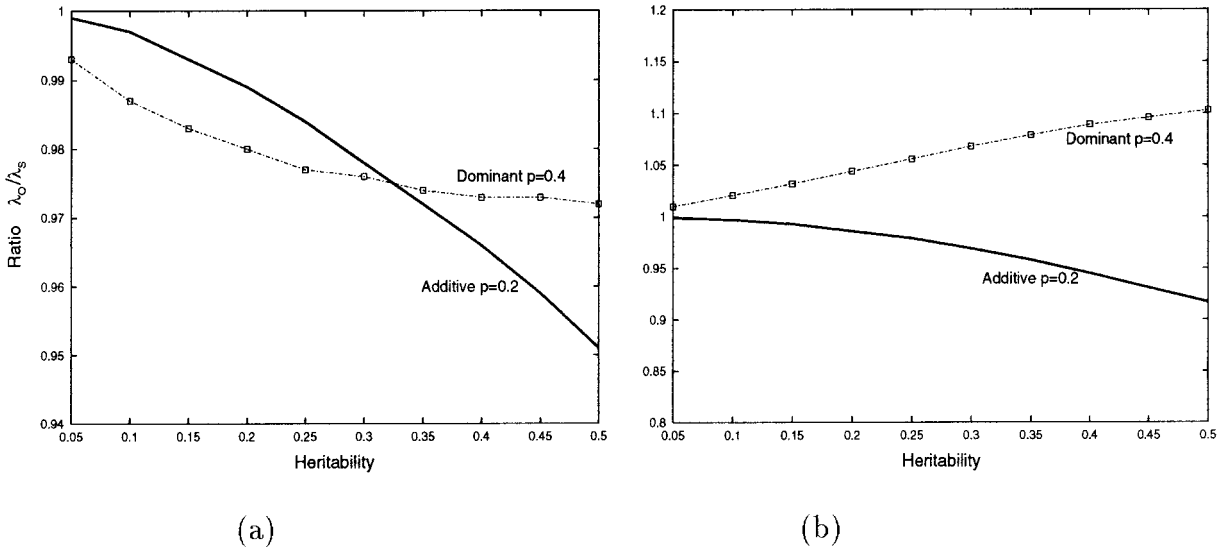

Figure 2 Typical values of λ_O/λ_S, plotted as functions of heritability, for HC [(10,10)] sib pairs (*a*) and for ED [(10,1)] sib pairs (*b*), when ρ = 0 is assumed. The graphs show the results under two models, an additive model with $p = .2$ and a dominant model with $p = .4$.

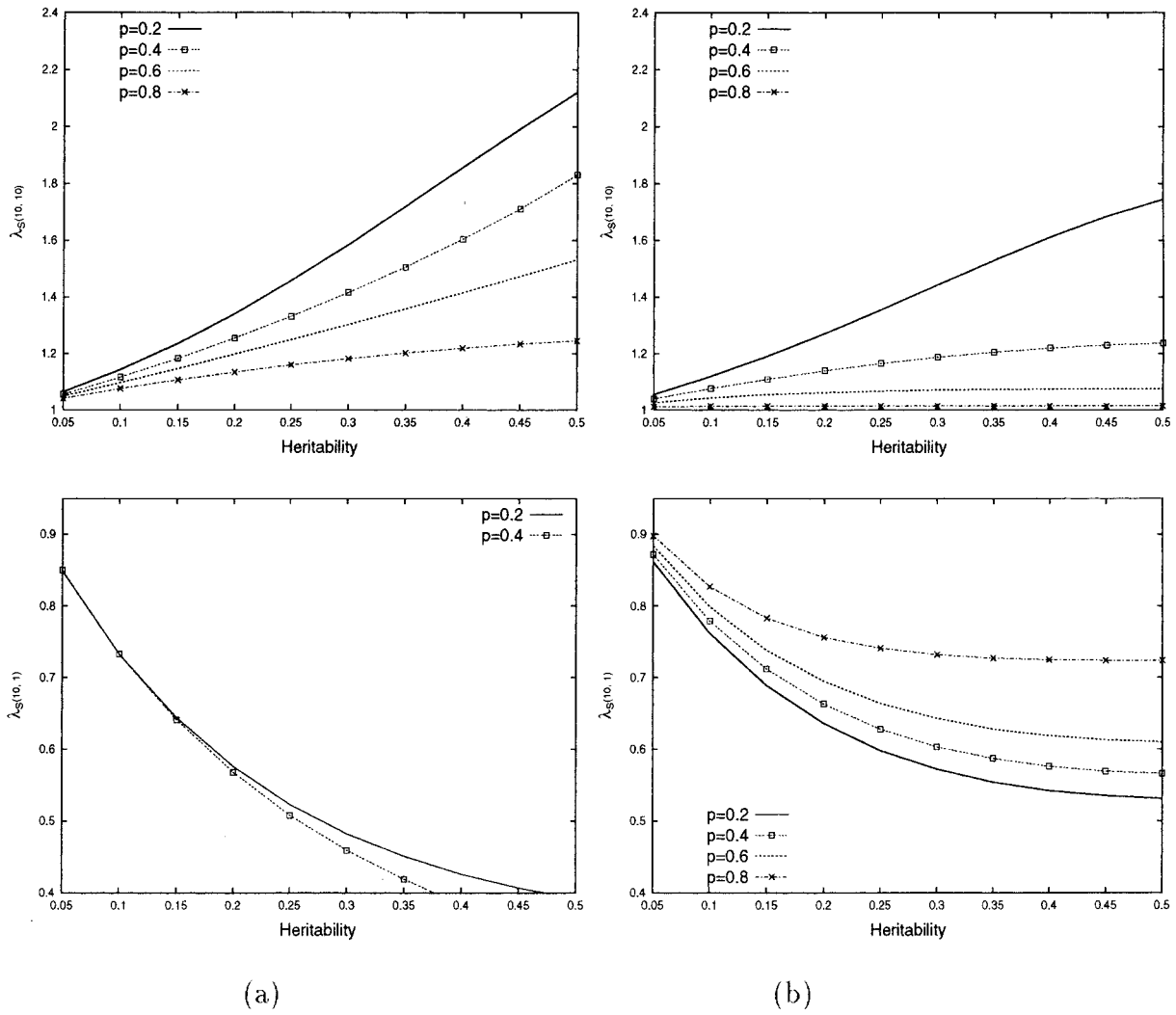

Figure 3 Values of λ_S for HC (*top*) and ED (*bottom*) sib pairs, as functions of heritability, for additive (*a*) and dominant (*b*) models. $\rho = .4$ is assumed for all the models displayed. $p = .2, .4, .6$, or .8. Under the additive models, since λ_S, for ED sib pairs, for $p = .6$ and .8 takes the same values as $p = .4$ and .2, respectively, these values are not plotted. Smaller values of λ_S for ED pairs and larger values of λ_S for HC pairs both indicate better power to detect linkage, for the respective ESPs.

In figure 3, we demonstrate how the values of λ_S for ED and HC sib pairs change with the underlying genetic models, by plotting λ_S against the heritability. We fix $\rho = .4$ and let the genotypic values vary to get the right value for heritability. The values of the gene frequencies also vary ($p = .2, .4, .6$, or .8). Note that under an additive model, ED sib pairs have the same λ_S, for frequencies p and $1 - p$. Also, under recessive models with gene frequency p, λ_S is the same as that under the dominant model with frequency $1 - p$, and with the positions of LC and HC sib pairs reversed; hence, we skip displaying here the results for recessive models. When p is high, the curves of λ_S for HC pairs are fairly flat, indicating that the sole HC design is not efficient. The curves of λ_S for ED pairs are closer to each other, compared with those for HC pairs, especially under the additive models, showing that the ED design is less sensitive to the underlying gene frequencies.

Remember that all the calculations done so far are based on the division of trait values into 10 deciles (with equal probabilities). Different types of division certainly will yield different values for the λ's. As an example, let us divide the trait values into three intervals, with the middle interval having a fixed proportion of 40% and an arbitrary probability assigned to the lower interval. Using the (upper) threshold T of the lower interval as the indicator of the division, we plot the values of λ_S as functions of T, in figure 4, for $\rho = 0$ and $\rho = .4$. As the threshold increases, λ_S for HC pairs also increases (i.e., becomes more powerful) under both the additive and

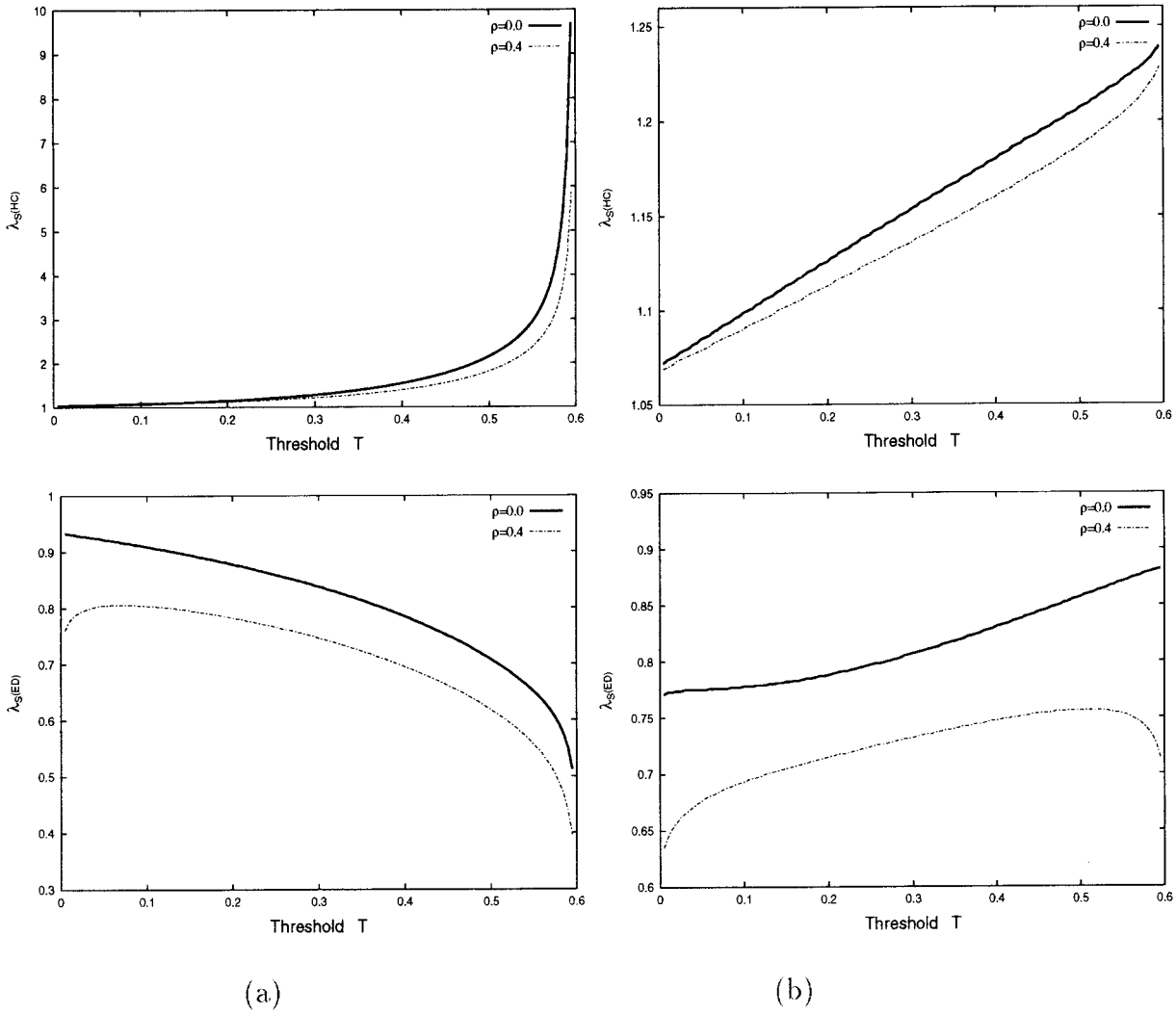

Figure 4 Values of λ_S plotted against T, for HC sib pairs (*top*) and ED sib pairs (*bottom*), under additive models with $p = .1$ (*a*) and dominant models with $p = .4$ (*b*). T is the upper threshold of the lower interval, which results from a trichotomization of trait values, with the middle interval having a fixed probability of .4. T is allowed to vary within the range of .005–.595. For each type of model, both cases of $\rho = 0$ and $\rho = .4$ are plotted in the same graph, for comparison.

the dominant models, whereas λ_S for ED pairs will decrease or increase (i.e., become more or less powerful), depending on the model. Also, compared with ED sib pairs, the values of λ_S for HC pairs are less sensitive to the presence of positive ρ's, when the threshold defining extreme-trait values varies. This suggests that, to achieve an optimum design, one must choose carefully the thresholds defining the extreme trait values, as well as the types of ESPs used for analysis.

In tables 3 and 4, we give the power on a rough grid of $(0 < \lambda_S, \lambda_O \leqslant 1)$ for 50 ED sib pairs and $(1 \leqslant \lambda_S, \lambda_O \leqslant 3)$ for 50 HC pairs. The necessary sample sizes for a power of 80% on the same grids of (λ_S,λ_O) are shown in tables 5 and 6. All the calculations were done by use of a computer program (LAMBDA), which is available from the authors and which will do the calculations for any values of the λ's, at any level of significance. We have used the significance level of .001 in tables 5 and 6. The power and sample sizes at points where the combination (λ_S,λ_O) is apparently invalid for any genetic model are not shown in the tables, even though they are mathematically computable.

We see that for ED sib pairs ($\lambda_S < 1.0$), for a fixed λ_O, the smaller the values of λ_S, the more powerful the ED sib pairs; for a fixed λ_S, the closer λ_O is to 1, the more powerful are the ED sib pairs. For HC sib pairs ($\lambda_S \geq 1.0$), for a fixed λ_O, the larger the values of λ_S, the more powerful the HC sib pairs; again, for a fixed λ_S, the closer λ_O is to 1, the more powerful are the HC sib pairs. We interpret the pattern as follows: Since

Table 3

Power of 50 ED Sib Pairs, at a Significance Level of .001, Calculated on a Grid of [$0 < \lambda_S(10,1), \lambda_O(10,1) \leq 1.0$]

	Power, for λ_O =[a]									
λ_S	.10	.20	.30	.40	.50	.60	.70	.80	.90	1.00
.10	...	...	...	...	...	...	...	...	...	...
.20	...	...	...	...	...	...	...	...	...	...
.30	1.00	...	...	...	...	...	...	...	...	...
.40	.97	1.00	1.00	...	...	...	...	...	...	...
.50	.10	.42	.85	1.00	1.00	...	...	...	...	...
.60	...	.01	.06	.26	.67	.97	1.00	...	...	...
.70	...	...	...	.00	.03	.15	.47	.87	1.00	...
.80	...	...	...	...	...	.00	.02	.09	.31	.71
.90	...	...	...	...	...	...	...	.00	.01	.05
1.00	...	...	...	...	...	...	...	...	...	.00

[a] An ellipsis indicates that the (λ_S,λ_O) combination is not valid for a genetic model.

smaller values of λ_S indicate a higher chance that an ED sib pair shares zero genes IBD and since larger values of λ_S indicate a higher chance that an HC sib pair shares two genes IBD, they both lead to more-powerful tests. When λ_O approaches 1, from equation (7) we see that values of λ_M increase for HC sib pairs and decrease for ED sib pairs. In either case, the expected IBD sharing for the HC or the ED sib pairs moves further away from $^1/_2$ (the value expected under the null hypothesis of no linkage), resulting in more-powerful tests (also see table 2).

Discussion

Although sib pairs with EC phenotypes and, even more so, sib pairs with ED phenotypes are more powerful than other types of sib pairs, for the detection of QTLs, there are a lot of practical issues that need to be addressed before one can fully take advantage of this fact. In a previous article (Gu et al. 1996), we explored the benefits of combining ED and EC sib pairs, to compensate for the fact that ED sib pairs alone are hard to find. This investigation takes that step further by the provision of a method for the estimation of the necessary ED or EC sample sizes, even when little information about the underlying genetic model is available, through the introduction and application of the generalized λ's (the λ method). The expected IBD sharing of a sib pair is expressed in terms of $\lambda_O(h,l)$, $\lambda_S(h,l)$, and $\lambda_M(h,l)$, the generalized λ's for parent-offspring pairs, sib pairs, and monozygotic twins, respectively. By use of equation (7), the expected IBD sharing can be expressed in terms of

Table 4

Power of 50 HC Sib Pairs, at a Significance Level of .001, Calculated on a Grid of [$1.0 \leq \lambda_S(10,10), \lambda_O(10,10) < 3.0$]

	Power, for λ_O =[a]									
λ_S	1.00	1.20	1.40	1.60	1.80	2.00	2.20	2.40	2.60	2.80
1.00	.00	...	...	...	...	...	...	...	...	...
1.20	.24	.03	.00	...	...	...	...	...	...	...
1.40	.82	.48	.14	.01	.00	...	...	...	...	...
1.60	.98	.90	.67	.32	.07	.01	.00	...	...	...
1.80	1.00	.99	.95	.80	.52	.21	.04	.00	...	...
2.00	1.00	1.00	1.00	.97	.89	.68	.39	.13	.02	.00
2.20	1.00	1.00	1.00	1.00	.98	.93	.80	.55	.27	.08
2.40	1.00	1.00	1.00	1.00	1.00	.99	.96	.87	.69	.43
2.60	1.00	1.00	1.00	1.00	1.00	1.00	.99	.98	.92	.79
2.80	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	.98	.95

[a] An ellipsis indicates that the (λ_S,λ_O) combination is not valid for a genetic model.

Table 5

Sample size of ED Sib Pairs Needed for 80% Power, at a Significance Level of .001, Calculated on a Grid of [$0 < \lambda_S(10,1), \lambda_O(10,1) \leq 1.0$]

	Sample Size for ED Sib Pairs, for λ_O =[a]									
λ_S	.10	.20	.30	.40	.50	.60	.70	.80	.90	1.00
.10	...	...	...	...	...	...	...	...	...	...
.20	...	...	...	...	...	...	...	...	...	...
.30	8	...	...	...	...	...	...	...	...	...
.40	60	30	17	...	...	...	...	...	...	...
.50	887	212	89	47	27	...	...	...	...	...
.60	...	–[b]	–[c]	294	124	66	38	...	...	...
.70	...	...	...	–[b]	–[c]	386	164	88	52	...
.80	...	...	...	...	...	–[b]	–[c]	491	210	112
.90	...	...	...	...	...	...	...	–[b]	–[c]	607
1.00	...	...	...	...	...	...	...	...	...	–[b]

[a] An ellipsis indicates that the (λ_S,λ_O) combination is not valid for a genetic model.
[b] >99,999.
[c] >999.

$\lambda_O(h,l)$ and $\lambda_S(h,l)$ only. The values of $\lambda_O(h,l)$ and $\lambda_S(h,l)$ can be estimated by use of the current data set or by use of results from previous studies. The estimation involves only the recurrence risks of parent-offspring and sib pairs and the population prevalence of various trait outcomes. More specifically, for example, it does not require the genotyping of parent-offspring pairs. The estimation of recurrence risks certainly will depend on the sampling method used in a study, a topic on which we did not elaborate in this paper.

The generalization of the concept of λ in the setting of quantitative traits is straightforward when there are no residual correlations among the relative pairs. When there *is* residual correlation, we simply assume that the residual correlations among relatives are all the same and derive the same formulas for expected IBD sharing, using λ's. Two observations are worth noting. First, direct estimation of λ's then becomes more difficult, since the probability $P_r(l|h)$ is difficult to estimate in practice. However, one may avoid this difficulty by using ratios of recurrence risks, which are relatively easy to estimate from the data. Second, if the residual correlations among different types of relatives are not identical, the concept of λ's still can be generalized, by use of so-called λ fac-

Table 6

Sample Size of HC Sib Pairs Needed for 80% Power, at a Significance Level of .001, Calculated on a Grid of [$1.0 \leq \lambda_S(10,10), \lambda_O(10,10) < 3.0$]

	Sample Size for HC Sib Pairs, for λ_O =[a]									
λ_S	1.00	1.20	1.40	1.60	1.80	2.00	2.20	2.40	2.60	2.80
1.00	–[b]	...	...	...	...	...	...	...	...	...
1.20	284	–[c]	–[b]	...	...	...	...	...	...	...
1.40	97	170	375	–[c]	–[b]	...	...	...	...	...
1.60	56	80	123	216	477	–[c]	–[b]	...	...	...
1.80	39	51	69	99	153	267	591	–[c]	...	...
2.00	30	38	48	62	84	120	185	324	716	–[c]
2.20	25	30	37	45	57	74	100	143	220	386
2.40	22	25	30	36	43	53	67	87	117	167
2.60	19	22	25	30	35	41	50	62	78	101
2.80	17	20	22	26	29	34	40	48	58	71

[a] An ellipsis indicates that the (λ_S,λ_O) combination is not valid for a genetic model.
[b] >99,999.
[c] >999.

tors for the trait locus and the residual effect, but decomposition of the ratio by trait locus and residual factors will be very difficult to achieve in practice. Its use may be limited to simulation studies.

The case in which the marker is not linked completely with the trait locus ($\theta > 0$) also is covered here. When a genomewide search for QTLs is performed, by use of dense maps, θ may be approximated by use of the genetic distance between adjacent markers. Therefore, the expected IBD sharing for both the so-called best-case scenario (the QTL on top of the marker) and the so-called worst-case scenario (the QTL in the middle of two flanking markers) can be estimated by use of λ's and the approximated θ. Hence, the method developed here can be applied in a genomewide scan for QTLs.

The range of λ's depends on how the trait values are divided, and the way that the trait values are divided should have effects on the power of ESPs. We have shown that, in the case of a trichotomy (fig. 4), by moving the thresholds toward the higher end or the lower end of the trait distribution, the λ's for ED and EC sib pairs will change in accordance with the underlying model, as does the power. In general, when the thresholds are pushed toward the two ends, at the same time, the extremeness of the ESPs will increase, thus resulting in more-powerful tests. But the increased extremeness also will reduce the availability of ESPs. This simple observation reveals the triangular relationship among the polychotomization of quantitative traits, the need for the combination of ED and EC pairs, and the cost effectiveness of a study design. How to balance this triangle to get the best out of a study is a very important practical challenge in the application of the selected sib-pair methodology. By applying the concept of the generalized λ's developed here, we address this issue in the second article of this series (Gu and Rao 1997).

Acknowledgments

This work was supported, in part, by NIH and NIMH grants GM28719 and MH31302, respectively. We thank two anonymous reviewers for their many helpful comments, which improved this article.

References

Carey G, Williamson J (1991) Linkage analysis of quantitative traits: increased power by using selected samples. Am J Hum Genet 49:786–796

Eaves L, Meyer J (1994) Locating human quantitative trait loci: guidelines for the selection of sibling pairs for genotyping. Behav Genet 24:443–455

Fulker DW, Cardon LR, DeFries JC, Kimberling WJ, Pennington BF, Smith SD (1991) Multiple regression analysis of sib-pair data on reading to detect quantitative trait loci. Reading Writing 3:299–313

Gu C, Rao DC (1997) 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 (in this issue)

Gu C, Todorov A, Rao DC (1996) Combining extremely concordant sibpairs with extremely discordant sibpairs provides a cost effective way to linkage analysis of quantitative trait loci. Genet Epidemiol 13:513–533

James JW (1971) Frequency in relatives for an all-or-not trait. Ann Hum Genet 35:47–49

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

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

Risch N, Zhang H (1995) Extreme discordant sib pairs for mapping quantitative trait loci in humans. Science 268: 1584–1589

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