Am. J. Hum. Genet. 58:836–843, 1996

Mapping Quantitative Trait Loci with Extreme Discordant Sib Pairs: Sampling Considerations

Neil J. Risch[1] and Heping Zhang[2]

[1]Department of Genetics, Stanford University School of Medicine, Stanford; and [2]Department of Epidemiology and Public Health, Yale University School of Medicine, New Haven

Summary

Elsewhere we have proposed the use of extreme discordant sib pairs (EDSPs) for mapping quantitative trait loci in humans. Here we present sample sizes necessary to achieve a given level of power with this study design, as well as the number of sibs that need to be screened to obtain the required sample. Further, we present simple formulas for adjusting sample sizes to account for variable significance levels and power, as well as the density and informativeness of linkage markers in a multipoint sib-pair analysis. We conclude that with EDSPs, the most powerful study design, the smallest genetic effect detectable with a realistic sample size is ~10% of the variance of the trait.

Introduction

Recent years have witnessed tremendous advances in mapping and identifying the mutations that cause numerous Mendelian syndromes. The general paradigm has included initial linkage studies with multiplex families to map a locus to a specific chromosomal segment, followed by positional cloning. This endeavor is often expedited by observing linkage disequilibrium in the region of interest, identifying chromosomal aberrations, such as deletions or translocations in some affected individuals, and searching for candidate genes known to lie in the area. The picture is different for non-Mendelian or complex diseases. The lack of a simple correspondence between genotype and phenotype and the involvement of multiple loci make identification of single contributing loci difficult. One class of complex trait are quantitative, i.e., those measured on a continuous, rather than discrete, scale. Sometimes a disease phenotype is defined by thresholds applied to a continuous variable, such as weight and obesity or blood pressure and hypertension. For these cases, understanding the genetic basis for the continuous trait can lead to understanding of the disease (defined by the threshold) as well. One design commonly employed to map loci underlying a quantitative trait (quantitative trait loci, or QTLs) is sib pairs. In conventional analysis, the squared difference in trait values for a sib pair is regressed on identity by descent (IBD) at a marker locus or loci. However, for sib pairs selected at random, it has been shown that the power of such an approach is quite low unless the proportion of variance (heritability) due to a single contributing locus is large (50%) (Blackwelder and Elston 1982). When sib pairs are selected through probands with extreme values and the second, unselected sib's trait value is regressed on IBD with the proband, the power is increased but still remains low at small values of heritability (Carey and Williamson 1991).

We have recently shown that only three types of sib pairs have substantial power to detect linkage for a QTL, namely, those concordant for high or low values and those discordant for high and low values (extreme discordant sib pairs, or EDSPs) (Risch and Zhang 1995). In fact, pairs involving sibs with intermediate values provide little to no power to detect linkage. Further, we showed that the EDSPs provide the greatest power across most plausible genetic models and increase in power when there is a residual correlation among sibs (as is likely for multifactorial traits), as well as when allele frequencies are high. Thus, we concluded that EDSPs are the design of choice for mapping QTLs in humans. The power of EDSPs has also been noted by Eaves and Meyer using simulations (1994).

In our prior analysis, we only described the limiting case of no recombination between trait and marker loci and complete marker polymorphism. Here we provide basic power tables (i.e., sample-size requirements) for varying degrees of heritability, gene frequencies, residual correlation, and mode of inheritance, as well as outline adjustments to these numbers to account for recombination between marker and trait loci (in a multipoint analysis) as well as incomplete marker polymorphism.

Received October 9, 1995; accepted for publication January 23, 1996.

Address for correspondence and reprints: Dr. Neil Risch, Department of Genetics, M322, Stanford University School of Medicine, Stanford, CA 94305-5120.

0002-9297/96/5804-0022$02.00

Calculation of Power

We consider two types of discordant sib pairs: (1) one sib in the top 10% of the distribution and the other in the bottom 10% (defined as *T1B1*); and (2) one sib in the top 10% and the other in the bottom 30% (defined as *T1B3*). The first group is symmetric, while the second is not. For the latter, we assume that high values of the trait are associated with disease, and hence a collection of individuals in the top 10% is easy to obtain from clinical samples. Then, to expand the number of sib pairs obtained from this proband group, we consider up to the 30th percentile instead of only the 10th.

Power for these sib pairs can be calculated as we have described previously. Define a single locus A with two alleles A_1 and A_2. Let p equal the frequency of allele A_1 and $q = 1 - p$ the frequency of allele A_2. Let a be the mean value of individuals with genotype A_1A_1, d the mean for genotype A_1A_2, and $-a$ for genotype A_2A_2. We assume that the residual variance within each genotype is σ_E^2 and that there is a residual correlation between sibs of value ρ. The additive genetic variance due to locus A is $\sigma_A^2 = 2pq[a + (q - p)d]^2$, and dominance variance is $\sigma_D^2 = (2pqd)^2$; hence, the total variance due to locus A is $\sigma_G^2 = \sigma_A^2 + \sigma_D^2$. The heritability H due to locus A is $\sigma_G^2/(\sigma_G^2 + \sigma_E^2)$. Without loss of generality, we assume $\sigma_E^2 = 1$, so that $H = \sigma_G^2/(\sigma_G^2 + 1)$.

To calculate power, we need to determine the probability that a sib pair of given phenotypes shares 2, 1, or 0 alleles IBD at locus A. Let π denote the number of alleles (0, 1, or 2) shared IBD by the sib pair. Then, for example, for the top 10%–bottom 10% strategy, we calculate $P(\pi = i/\text{T1B1})$ for $i = 0$, 1, and 2. To do so, we apply Bayes's theorem, so that

$$Z_i = P(\pi = i/\text{T1B1}) = \frac{1}{D} P(\pi = i)P(\text{T1B1}/\pi = i)$$

and

$$D = \sum_{i=0}^{2} P(\pi = i)P(\text{T1B1}/\pi = i) .$$

We note that

$$P(\pi = 2) = P(\pi = 0) = 1/4 \quad \text{and} \quad P(\pi = 1) = 1/2 .$$

To calculate $P(\text{T1B1}/\pi = i)$, we use the formula

$$P(\text{T1B1}/\pi = i) = \sum_k P(G_k/\pi = i)P(\text{T1B1}/G_k) ,$$

where G_k denotes the kth possible pair of genotypes at locus A for a sib pair (of nine possible ordered genotype pairings). $P(\text{T1B1}/G_k)$ is obtained by integrating a bivariate normal density function with mean values specified by the genotypes of G_k, a variance of 1 within genotype, and correlation ρ. For more details on this calculation, see Risch and Zhang (1995). Calculations for other designs, such as T1B3, are performed similarly.

We initially assume a sample of n fully informative sib pairs (i.e., both parents typed, marker PIC value of 1). Define X_1 as the number of alleles (1 or 0) shared IBD from the father and X_2 the number of alleles shared IBD from the mother. Then, $Z_2 = P(X_1 = 1, X_2 = 1)$, $Z_1 = P(X_1 = 1, X_2 = 0) + P(X_1 = 0, X_2 = 1)$, $Z_0 = P(X_1 = 0, X_2 = 0)$, and $\tau = Z_2 + 1/2 Z_1$. For the ith sib pair, let X_{1i} be the IBD outcome with respect to the father and X_{2i} be the IBD outcome with respect to the mother. Then we use as our outcome statistic the mean number of alleles shared, namely, $\bar{X} = 1/2n \sum_{i=1}^{n} (X_{1i} + X_{2i})$. We assume that $\bar{X}$ is approximately normal with mean τ and variance $[\tau(1 - 2\tau) + Z_2]/2n$. We are testing the null hypothesis $H_0{:}\tau = 1/2$ against the alternative $H_1{:}\tau < 1/2$; because we are focusing on discordant pairs, we expect under linkage that allele sharing will be less than the null expectation. Hence, we employ the power for a one-sided test of a normal random variable, namely,

$$\Phi\left[\frac{z_\alpha/2 + (\tau - 1/2)\sqrt{2n}}{\sqrt{\tau(1 - 2\tau) + Z_2}}\right] , \tag{1}$$

where Φ is the cumulative standard normal distribution function and z is the normal deviate corresponding to a type 1 error probability of α (i.e., significance level). Then, the number of sib pairs required to obtain a power of $1 - \beta$ (i.e., the probability of rejecting the null hypothesis *when it is false*) is

$$\frac{1}{2}\left[\frac{z_{1-\beta}\sqrt{\tau(1 - 2\tau) + Z_2} - z_\alpha/2}{\tau - 1/2}\right]^2 . \tag{2}$$

Sample-Size Considerations

In tables 1–3, we provide required sample sizes to detect linkage with the T1B1 and T1B3 strategies for additive, dominant, and recessive models, respectively. We assume a significance level $\alpha = .0001$ (corresponding approximately to a lod score of 3) and power $1 - \beta = .80$. Numbers corresponding to gene frequencies ranging from .1 to .9 and heritabilities H ranging from .05 to .3 are given, as well as for the cases $\rho = 0$ and $\rho = .4$. As can be seen in the tables, residual correlation always reduces the necessary sample size. This is because, when there is a residual correlation causing sibs to be similar, phenotypically discordant sib pairs will be more likely to be genetically discordant at the locus of

Table 1

Required Number of Sib Pairs to Detect Linkage for EDSPs for an Additive Model, $\alpha = .0001$ and $1 - \beta = .8$

	H ($\rho = 0$)				H ($\rho = .4$)			
p	.05	.1	.2	.3	.05	.1	.2	.3
	A. Top 10% and Bottom 10%							
.1	6,827	1,647	378	155	1,367	342	94	52
.3	6,449	1,482	314	120	1,394	346	88	42
.5	6,405	1,464	308	116	1,397	347	87	40
.7	6,449	1,482	314	120	1,394	346	88	42
.9	6,827	1,647	378	155	1,367	342	94	52
	B. Top 10% and Bottom 30%							
.1	11,706	2,544	507	185	2,894	673	161	75
.3	13,410	2,967	596	217	3,263	775	181	78
.5	14,750	3,377	707	263	3,501	852	204	88
.7	16,527	3,982	894	348	3,782	952	239	108
.9	22,079	6,247	1,845	930	4,523	1,248	366	189

interest. This is especially true at low values of heritability, where the necessary sample size is often reduced by at least threefold. This is important because it is for loci with low heritability that there is likely to be residual sib correlation due to other genetic effects. At high heritability, most of the sib correlation is probably due to that locus, and hence there is unlikely to be a large residual correlation.

The sample sizes presented in tables 1–3, especially for the T1B1 strategy, are within experimental limits. This is true even for a low heritability of .1, especially when there is a significant residual correlation. The one exception for which EDSPs will not be practical is a rare recessive gene. For this situation, concordant pairs are more appropriate. This case can be identified a priori by evaluating the role of dominance variance for the

Table 2

Required Number of Sib Pairs to Detect Linkage for EDSPs for a Dominant Model, $\alpha = .0001$ and $1 - \beta = .8$

	H ($\rho = 0$)				H ($\rho = .4$)			
p	.05	.1	.2	.3	.05	.1	.2	.3
	A. Top 10% and Bottom 10%							
.1	6,625	1,567	353	143	1,404	359	103	59
.3	6,369	1,454	312	127	1,460	384	114	65
.5	6,610	1,565	358	151	1,480	396	122	72
.7	7,623	2,049	573	276	1,457	398	134	86
.9	28,765	19,984	19,001	18,996	1,610	849	753	753
	B. Top 10% and Bottom 30%							
.1	11,713	2,476	471	168	2,962	689	163	77
.3	14,713	3,352	698	263	3,569	887	225	106
.5	18,685	4,879	1,289	596	4,190	1,128	322	164
.7	29,814	10,384	4,446	3,158	5,523	1,733	641	415
.9	332,421	297,274	295,120	295,117	25,126	19,277	18,797	18,795

Table 3

Required Number of Sib Pairs to Detect Linkage for EDSPs for a Recessive Model, $\alpha = .0001$ and $1 - \beta = .8$

	H ($\rho = 0$)				H ($\rho = .4$)			
p	.05	.1	.2	.3	.05	.1	.2	.3
	A. Top 10% and Bottom 10%							
.1	28,765	19,984	19,001	18,996	1,610	849	753	753
.3	7,623	2,049	573	276	1,457	398	134	86
.5	6,610	1,565	358	151	1,480	396	122	72
.7	6,369	1,454	312	127	1,460	384	114	65
.9	6,625	1,567	353	143	1,404	359	103	59
	B. Top 10% and Bottom 30%							
.1	29,138	19,991	19,001	18,996	3,918	2,257	2,063	2,062
.3	11,211	2,591	622	282	2,836	702	199	111
.5	12,425	2,679	524	193	3,177	768	193	95
.7	14,505	3,282	676	252	3,534	875	221	104
.9	20,490	5,640	1,632	823	4,343	1,179	339	174

trait. Such a locus would produce substantial dominance variance, for example as demonstrated by a higher correlation between sibs than between parent and offspring. Specifically, consider the formulas for additive variance (σ_A^2) and dominance variance (σ_D^2) given above. The ratio of sibling correlation to parent-offspring correlation, for a recessive locus, is $1 + (V_D/2V_A) = 1 + (q/4p)$. When $p = .1$, this ratio is 3.25, or quite large. When $p = .33$, the ratio is only 1.5, or when $p = .50$, the ratio is 1.25. Thus, concordant pairs will generally be more useful than discordant ones only when the ratio of sibling to parent-offspring correlations is large.

Population Screening

An important question remains as to how many sib pairs need to be screened to obtain such a selected sample. Here we assume that only one tail of the distribution is of primary clinical interest, so that a large population of individuals at the high end (top 10%) of the distribution is readily available. The question, then, is how many sibs of these individuals need to be screened to obtain the requisite sample size (either bottom 10% or bottom 30%). These numbers are provided in tables 4–6 for both the T1B1 and T1B3 strategies, for the same models (heritabilities and gene frequencies) as shown in tables 1–3. Here the numbers range from the low thousands at high heritability to the tens of thousands at low heritability.

Adjustments to Sample Sizes

The numbers given in tables 1–3 are idealized, in that they assume no recombination ($\theta = 0$), parents are typed, and the marker is completely polymorphic. To obtain equivalent sample sizes under other conditions is straightforward.

First, we consider significance level and power. Suppose instead of a power of 80%, we would like a value $1 - \beta$. From formula 2, the sample sizes in tables 1–3 would need to be multiplied by

$$\left[\frac{2z_{1-\beta}\sqrt{\tau(1-2\tau)+Z_2}-z_\alpha}{2z_{.80}\sqrt{\tau(1-2\tau)+Z_2}-z_\alpha}\right]^2.$$

Unless τ deviates greatly from $\frac{1}{2}$, the above ratio can be well approximated by

$$\left[\frac{z_{1-\beta}-z_\alpha}{z_{.80}-z_\alpha}\right]^2. \tag{8}$$

For example, if a power of 90% is desired, the numbers in tables 1–3 can be multiplied by 1.2. A similar formula can be derived for a different significance level α.

Adjustment for recombination is also straightforward. Under recombination between a marker and the trait locus, the IBD probability τ changes to a value τ' closer to the null value of $\frac{1}{2}$. Again, resorting to formula 2, the sample sizes would need to be multiplied by

$$\left[\frac{2z_{1-\beta}\sqrt{\tau'(1-2\tau')+Z_2}-z_\alpha}{2z_{1-\beta}\sqrt{\tau(1-2\tau)+Z_2}-z_\alpha}\right]^2\times\left[\frac{\tau-\frac{1}{2}}{\tau'-\frac{1}{2}}\right]^2.$$

Table 4

Number of Sib Pairs to Be Screened to Obtain the Required EDSP Families for an Additive Model

	H ($\rho = 0$)				H ($\rho = .4$)			
p	.05	.1	.2	.3	.05	.1	.2	.3
	A. Top 10% and Bottom 10%							
.1	73,642	19,120	5,048	2,356	85,355	21,441	5,635	2,833
.3	69,725	17,357	4,335	1,958	87,698	22,336	5,873	2,829
.5	69,268	17,166	4,273	1,918	87,968	22,484	5,894	2,784
.7	69,725	17,357	4,335	1,958	87,698	22,336	5,873	2,829
.9	73,642	19,120	5,048	2,356	85,355	21,441	5,635	2,833
	B. Top 10% and Bottom 30%							
.1	41,359	9,575	2,180	910	31,577	7,481	1,825	843
.3	47,203	11,075	2,526	1,056	35,601	8,647	2,107	942
.5	51,790	12,518	2,941	1,239	38,155	9,479	2,360	1,053
.7	57,864	14,638	3,629	1,565	41,144	10,535	2,718	1,249
.9	76,794	22,521	7,066	3,745	48,954	13,587	3,951	1,976

In the above expression, in general the second term predominates, so to a very close approximation the multiplier is $[(\tau - \frac{1}{2})/(\tau' - \frac{1}{2})]^2$. For a recombination fraction θ, the IBD probability $\tau' = \tau\Psi + (1 - \tau)(1 - \Psi)$, where $\Psi = \theta^2 + (1 - \theta)^2$. Then,

$$\tau' - \tfrac{1}{2} = \tau(2\psi - 1) + 1 - \psi - \tfrac{1}{2} = (\tau - \tfrac{1}{2})(2\psi - 1) .$$

Thus, the ratio above reduces to

$$\left[\frac{(\tau - \frac{1}{2})}{(\tau' - \frac{1}{2})}\right]^2 = \left[\frac{(\tau - \frac{1}{2})}{(\tau - \frac{1}{2})(2\psi - 1)}\right]^2 = \frac{1}{(2\psi - 1)^2} = 1 + \frac{4\psi(1 - \psi)}{(2\psi - 1)^2} . \quad (4)$$

For example, for a recombination fraction θ of .05,

Table 5

Number of Sibs to Be Screened to Obtain the Required EDSP Families for a Dominant Model

	H ($\rho = 0$)				H ($\rho = .4$)			
p	.05	.1	.2	.3	.05	.1	.2	.3
	A. Top 10% and Bottom 10%							
.1	71,360	18,153	4,704	2,171	87,300	22,320	6,062	3,119
.3	68,379	16,777	4,150	1,920	90,381	23,957	7,066	3,915
.5	70,464	17,769	4,588	2,158	90,199	23,768	6,872	3,693
.7	80,337	22,615	6,847	3,531	85,781	21,777	5,902	2,943
.9	294,105	205,218	195,254	195,204	85,130	41,016	35,522	35,514
	B. Top 10% and Bottom 30%							
.1	41,309	9,294	2,023	828	32,215	7,611	1,823	846
.3	51,420	12,307	2,843	1,192	38,506	9,660	2,478	1,165
.5	64,731	17,513	4,938	2,421	44,717	11,980	3,333	1,618
.7	102,091	36,212	15,865	11,407	58,074	17,793	6,196	3,786
.9	1,115,392	997,837	990,632	990,622	261,307	199,189	194,085	104,064

Table 6

Number of Sibs to Be Screened to Obtain the Required EDSP Families for a Recessive Model

	H ($\rho = 0$)				H ($\rho = .4$)			
p	.05	.1	.2	.3	.05	.1	.2	.3
	A. Top 10% and Bottom 10%							
.1	294,105	205,218	195,254	195,204	85,130	41,016	35,522	35,514
.3	80,337	22,615	6,847	3,531	85,781	21,777	5,902	2,943
.5	70,464	17,769	4,588	2,158	90,119	23,768	6,872	3,693
.7	68,379	16,777	4,150	1,920	90,381	23,957	7,066	3,915
.9	71,360	18,153	4,704	2,171	87,300	22,320	6,062	3,119
	B. Top 10% and Bottom 30%							
.1	99,292	68,430	65,085	65,068	39,460	22,004	19,964	19,953
.3	39,031	9,434	2,462	1,200	29,860	7,195	1,863	919
.5	43,412	9,855	2,154	889	34,005	8,225	2,045	979
.7	50,710	12,062	2,761	1,148	38,134	9,533	2,437	1,146
.9	71,296	20,372	6,285	3,343	46,924	12,795	3,631	179

the sample sizes need to be multiplied by 1.52; for a recombination fraction of .10, the multiplier is 2.44.

In general, in mapping trait loci, we do not test just a single marker but a map of markers in a multipoint analysis. Consider a map of completely informative equally spaced markers. Power is greatest when the trait locus occurs at the same site as a marker (corresponding to the case $\theta = 0$). The greatest loss of power occurs when the trait locus is exactly midway between flanking markers. Assume the trait locus is recombination fraction θ away from each of the flanking marker loci. Then, instead of measuring IBD from a single marker locus, we have IBD at the two marker loci. Letting 1 represent IBD and 0 non-IBD, the probabilities for the four possible outcomes for the two marker loci, assuming no interference, are

$$P(1,1) = \psi^2\tau + (1-\psi)^2(1-\tau),$$

$$P(1,0) = \psi(1-\psi),$$

$$P(0,1) = \psi(1-\psi),$$

$$P(0,0) = (1-\psi)^2\tau + \psi^2(1-\tau).$$

Note that the outcomes (1,0) and (0,1) provide no information regarding τ. Thus, our effective sample size is only

$$1 - 2\psi(1-\psi) = \psi^2 + (1-\psi)^2.$$

We can estimate τ' by

$$\tau' = \frac{P(1,1)}{P(1,1) + P(0,0)} = \frac{1}{[\psi^2 + (1-\psi)^2]}[\psi^2\tau + (1-\psi)^2(1-\tau)].$$

After some algebra it is easy to show that

$$\tau' - \tfrac{1}{2} = \frac{1}{[\psi^2 + (1-\psi)^2]}(\tau - \tfrac{1}{2})[\psi^2 - (1-\psi)^2].$$

Thus,

$$\left[\frac{(\tau - \tfrac{1}{2})}{(\tau' - \tfrac{1}{2})}\right]^2 = \left[\frac{\psi^2 + (1-\psi)^2}{\psi^2 - (1-\psi)^2}\right]^2.$$

However, in estimating the required sample size, we also need to account for loss of information due to the two recombinant groups. Thus, the ratio above needs to be divided by $\psi^2 + (1-\psi)^2$, the proportionate reduction in effective sample size. Finally, we get a ratio of

$$\frac{\psi^2 + (1-\psi)^2}{[\psi^2 - (1-\psi)^2]^2} = 1 + \frac{2\psi(1-\psi)}{(2\psi - 1)^2}. \qquad (5)$$

Notice that the second term in this expression is exactly half that for the single-locus case given above. Thus, the required increase in sample size is exactly half that for

a single marker located a recombination fraction θ away. For example, for a 10-cM map, the required increase in sample size is 1.26; for a 20-cM map, the required increase is 1.72. In the latter case, this amounts to a 40% reduction in necessary sample size by using a multipoint analysis compared to single-locus analysis.

Next, we consider the effect of using incompletely polymorphic marker loci. The probability that a sib pair is informative for IBD from a given parent is simply the PIC value, which we denote by s. Again, we consider the least-informative case, when the trait locus lies midway between two marker loci. In a proportion s^2 of cases, both flanking markers are informative. In a proportion $2s(1 - s)$ of cases, only one of the two flanking markers is informative. In the remaining $(1 - s)^2$ proportion of cases, neither marker is informative. To estimate τ', we need to combine information from the fully informative and half informative cases. Using arguments similar to those given above, we calculate

$$\tau' - \tfrac{1}{2} = \frac{(\tau - \tfrac{1}{2})(2\psi - 1)(2 - s)}{2 + s(2 - s)^2(2\psi - 1)^2}.$$

Hence, the sample-size ratio, after accounting for the reduced effective sample size due to recombinants between flanking markers (for the fully informative cases) and the fully uninformative cases (for marker uninformativeness), is

$$\frac{2 + s(2\psi^2 - 2\psi - 1)}{s(2 - s)^2(2\psi - 1)^2}. \qquad (6)$$

For a 20-cM map, or $\theta \approx .10$, if $s = .8$, the necessary multiplier is 2.04; if $s = .7$, it is 2.26 (note, these are in contrast to the 1.72 given above for the $s = 1$ case). For a 10-cM map, or $\theta \approx .05$, if $s = .8$, the required multiplier is 1.41; if $s = .7$, it is 1.52. These numbers compare to 1.26 for the $s = 1$ case given above. Note that these numbers are somewhat conservative, because we have not considered more distant flanking markers when an immediate marker is uninformative. This is likely to have led to slightly inflated multipliers for denser (e.g., 10-cm) maps.

Finally, these calculations assume parents are available and typed. When such is not the case, resort to identity by state replaces IBD, and some information is lost. General guidelines for this case have been given elsewhere (Risch 1990, 1992; Holmans 1993; Hauser et al., in press). When other sibs are available for typing to help reconstruct missing parents, the situation becomes more complicated and is dealt with elsewhere (Hauser et al, in press).

Discussion

Tables 1–3, along with formulas 3–6, should be useful for designing studies to map loci for quantitative traits using extreme discordant sib pairs. The power in any particular case will depend on the heritability of the locus to be mapped, as well as the density and PIC value of the markers.

The feasibility of screening large numbers of individuals will depend on whether information on the trait is readily available (e.g., weight or height) or whether expensive and/or invasive testing is required. In any event, tables 1–3, in conjunction with tables 4–6, demonstrate the trade-offs in designing a linkage study. If the limiting factor is the expense of genotyping, but the sibling material is easily obtained, a T1B1 strategy is appropriate; on the other hand, if it is the sibling material that is the limiting resource, whereas the genotyping is inexpensive, the T1B3 strategy is preferred. Perhaps a reasonable trade-off is to collect all the T1B3 sibships, first type the T1B1 pairs, and then confirm the initial positive findings with the remaining pairs.

Tables 1–6 also illustrate that even using EDSPs, generally the most powerful sib-pair design for detecting linkage for quantitative trait loci (Risch and Zhang 1995), the power to detect loci of low heritability, i.e., <10%, is still limited, and in this region screening of very large numbers of individuals is necessary.

The example we have given uses the upper 10% of the trait distribution as representative of individuals of extreme phenotype likely to be classified as "affected." Similar tables can be generated for higher (e.g., upper 5%) or lower (e.g., upper 20%) thresholds. As we have shown elsewhere (Risch and Zhang 1995), using a higher threshold will increase power per pair (but leads to reduced sample sizes), while using a lower threshold decreases power per pair (but leads to an increased sample size). For some diseases, it is not possible to directly categorize affected individuals as being in the upper x% of a continuous distribution. For example, although hypertensives are in the upper tail of the blood-pressure distribution, it would not be correct to say that they precisely represent the upper x% of the blood pressure distribution (systolic or diastolic). However, if someone is given a diagnosis of hypertension and placed on antihypertensive medication, it is likely that he or she has extreme blood-pressure values, systolic and/or diastolic. The advantage of the approach presented here is that it does not depend on the actual values of the quantitative trait, only on the fact that sibs have extreme values. Sibs with low blood pressure can be measured and evaluated directly (for example, for being in the bottom 10% or 30% of the

distribution), while it is probably safe to assume that those considered affected are *approximately* in the upper x% of the distribution, where x is the population prevalence of the disease.

In general, association studies with candidate loci can be far more powerful for detecting weak gene effects than linkage studies (Risch 1987; Greenberg 1993; Eaves and Meyer 1994; Risch and Zhang 1995). The limitation of that approach, however, is that it requires either prior identification of the causative genetic variant itself, or another variant in linkage disequilibrium with it. In general, linkage disequilibrium spans only short genomic regions, limiting its utility in a global genome screen at this time. However, for candidate loci, the approach can be extremely powerful if variation at the locus contributes to trait variation. In any event, the EDSP design we have discussed is a powerful and robust design for such studies, as has also been shown by Eaves and Meyer (1994).

Acknowledgments

This work was supported by National Institutes of Health grants HG00348 (to N.R.) and HD30712 (to H.Z.). We are indebted to Cathy Halper for technical assistance.

References

Blackwelder WC, Elston RC (1982) Power and robustness of sib-pair linkage tests and extension to larger sibships. Commun Stat Theory Methods 11:449–484

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

Greenberg DA (1993) Linkage analysis of "necessary" disease loci versus "susceptibility" loci. Am J Hum Genet 52:135–143

Hauser ER, Boehnke M, Guo S-W, Risch N. Affected sib-pair interval mapping and exclusion for complex genetic traits: sampling considerations. Genet Epidemiol (in press)

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

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

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

——— (1992) Corrections to "Linkage strategies for genetically complex traits. III. The effect of marker polymorphism on analysis of affected relative pairs." Am J Hum Genet 51:673–675

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