Am. J. Hum. Genet. 61:1424–1430, 1997

Genomewide Transmission/Disequilibrium Testing—Consideration of the Genotypic Relative Risks at Disease Loci

Nicola J. Camp

Section of Molecular Medicine, University of Sheffield, Sheffield

Summary

Genomewide association studies are set to become the tool of the future for detection of small-effect genes in complex diseases. It will therefore be necessary to calculate sufficient sample sizes with which to perform them. In this paper I illustrate how to calculate the required number of families for general genotypic relative risks (GRRs). I show the superior sensitivity of the genomewide association study over the standard genomewide affected-sib-pair linkage analysis, for a range of different underlying GRR patterns. I also illustrate the extent of change in the sample sizes that is necessary for a genomewide association analysis depending on the pattern of the GRRs at the disease locus. In many cases, the comparative numbers of families required under different genetic mechanisms vary by several orders of magnitude. These sometimes dramatic differences have important implications for the determination of the size of the collection of samples prior to analysis and for the types of effects that are likely—and unlikely—to be detected by such an analysis.

Introduction

It has been proposed (Risch and Merikangas 1996) that, pending the existence of a map consisting of several biallelic markers in each gene (~1/kb), genomewide association studies will become not only a possibility but a superior alternative to the standard genomewide affected-sib-pair (ASP) linkage analyses. In their work, Risch and Merikangas compared ASP linkage analysis with association analysis (using the transmission/disequilibrium test [TDT]; Spielman et al. 1993), for sensitivity in genomewide screens, but only under certain simplifications. Specifically, they assumed that the set of biallelic markers screened actually contained any disease loci and, also, that a multiplicative relationship for the genotypic relative risk (GRR) existed at those disease loci. In a response to the Risch and Merikangas work, Müller-Myhsok and Abel (1997) addressed the former of these two assumptions and commented that, if less than maximum linkage disequilibrium existed between the biallelic marker and true disease locus, and if the allele frequencies at the marker and disease locus differed greatly, the sample sizes necessary for the TDT would be greatly increased. The work undertaken here addresses the latter assumption, that of multiplicative GRRs. The existence of multiplicative GRRs is a special case, since these conditions provide stability in the sense that Hardy-Weinberg equilibrium (HWE) is maintained in the disease population. In reality, however, we know that a disease population can differ greatly from HWE; for example, in insulin-dependent diabetes mellitus the distribution of genotypes at the HLA-DRB1 locus differs significantly from that expected under HWE (Field et al. 1986). I have therefore derived the necessary equations to compare genomewide ASP linkage analysis with TDT, for general GRRs. I also compare ASP linkage and TDT analyses for sample sizes necessary under four standard and commonly assumed modes of inheritance (MOIs): multiplicative, additive, common recessive, and common dominant.

Methods

I here derive the equations of Risch and Merikangas (1996), for general GRRs. I define the single-locus-specific penetrances for the disease to be f_0, f_1, and f_2, for the probability of disease, given genotypes aa, Aa, and AA, respectively, where A is the putative disease allele. I will consider two specific GRRs—GRR_1 and GRR_2—such that GRR_i indicates the relative risk to an individual carrying i A alleles, compared with that of an individual carrying none. These GRRs, defined for $f_0 \neq 0$, can be written as $GRR_1 = f_1/f_0$ and $GRR_2 = f_2/f_0$.

Received July 28, 1997; accepted for publication October 7, 1997; electronically published December 12, 1997.

Address for correspondence and reprints: Dr. Nicola J. Camp, Section of Molecular Medicine, M Floor, Royal Hallamshire Hospital, University of Sheffield, Sheffield S10 2JF, United Kingdom. E-mail: njc@mendel.shef.ac.uk

0002-9297/97/6106-0028$02.00

Table 1

Case-Control-Type Table for a Candidate Locus

	Genotype		
Population	AA	Aa	aa
Affected	*a*	*b*	*c*
Control	*d*	*e*	*f*

These quantities can be estimated at candidate loci by the heterozygous and homozygous odds ratios (ORs), which can be calculated simply by use of a case-control-study table, such as table 1. The heterozygous OR (OR_1) and homozygous OR (OR_2) are found as follows: $OR_1 = bf/ce$; and $OR_2 = af/cd$, with OR_1 providing an estimate for GRR_1 and with OR_2 providing an estimate for GRR_2.

Special relationships may exist between GRR_1 and GRR_2. Table 2 indicates these, in terms of a background penetrance parameter $\alpha(\neq 0)$ and a genetic penetrance parameter $\gamma(>1)$, for some standard MOIs.

ASP Linkage Analysis

Similarly to Risch and Merikangas (1996), I have assumed the best-case scenario for ASP linkage analysis—that is, a fully informative (heterozygosity = 1) closely linked ($\theta = 0$) marker. Now, from Suarez et al. (1978), with $\theta = 0$,

$$P(0 \text{ IBD genes/ASP}) = \frac{1}{4} - \left[\frac{\frac{1}{2}V_A + \frac{1}{4}V_D}{4(K^2 + \frac{1}{2}V_A + \frac{1}{4}V_D)}\right]$$

$$P(1 \text{ IBD gene/ASP}) = \frac{1}{2} - \left[\frac{\frac{1}{2}V_D}{4(K^2 + \frac{1}{2}V_A + \frac{1}{4}V_D)}\right],$$

$$P(2 \text{ IBD genes/ASP}) = \frac{1}{4} + \left[\frac{\frac{1}{2}V_A + \frac{3}{4}V_D}{4(K^2 + \frac{1}{2}V_A + \frac{1}{4}V_D)}\right]$$

where $K = p^2f_2 + 2pqf_1 + q^2f_0$ is the population prevalence, p is the allele frequency for allele A and $q = 1 - p$, $V_A = 2pq[p(f_2 - f_1) + q(f_1 - f_0)]^2$ is the additive variance, and $V_D = p^2q^2(f_2 - 2f_1 + f_0)^2$ is the dominance variance.

The expected proportion of alleles shared IBD by an ASP, Y, can therefore be written as

$$Y = \frac{1}{2}\left[1 + \frac{V_A + V_D}{4(K^2 + \frac{1}{2}V_A + \frac{1}{4}V_D)}\right].$$

To calculate the number of ASP families necessary, Risch and Merikangas (1996) defined the random variable (RV) B_i, such that $B_i = 1$ if the ASP shares an allele from the ith parent and $B_i = -1$ if it does not. Now, $P(B_i = 1) = Y$ and $P(B_i = -1) = 1 - Y$ (see appendix A). This simplifies under the null hypothesis (H_0) of no linkage, since $V_A = V_D = 0$, and hence $P(B_i = 1) = P(B_i = -1) = .5$. The mean and variance of B_i under H_0 are therefore given by 0 and 1, respectively. In a sample of N families (two parents and two offspring) there are $2N$ independent observations of B_i, giving a total mean and variance, over all families, of 0 and $2N$, respectively. Under the alternative hypothesis, the mean and variance over all families, in terms of Y, are given by $2N(2Y - 1)$ and $8NY(1 - Y)$, respectively. With these quantities and standard normal-distribution theory (see appendix B), the number of families necessary to achieve 80% power under any genetic mechanism can be calculated:

$$N \geqslant \frac{[z_{1-\alpha} - 2z_\beta\sqrt{Y(1-Y)}]^2}{2(2Y-1)^2},$$

where $z_x = \Phi^{-1}(x)$, with $\Phi()$ as the cumulative distribution function (CDF) of the standard normal distribution, and where α is the probability of a type I error and β is the probability of a type II error (i.e., power = $1 - \beta$).

Table 3 shows the sample sizes needed, for various MOI and various values of p. A critical value of 2.2×10^{-5} has been used to determine genomewide significance, as suggested by Lander and Kruglyak (1995).

TDT Association and Linkage Analysis Using Singletons

I assume that N families consisting of a single affected offspring (SAO) plus both parents have been collected. The unit for use in the TDT is a heterozygous parent plus SAO, and therefore only heterozygous parents will ultimately be included in the analysis. The probability that a parent is heterozygous, given that he or she has an SAO, is

$$h_s = P(\text{heterozygous/SAO}) = \frac{pq(pf_2 + f_1 + qf_0)}{K},$$

Table 2

Penetrance Functions and GRRs for Selected MOIs

	Penetrance Function			GRR	
MOI	f_0	f_1	f_2	GRR_1	GRR_2
Multiplicative	α	$\alpha\gamma$	$\alpha\gamma^2$	γ	γ^2
Additive	α	$\alpha\gamma$	$2\alpha\gamma$	γ	2γ
Common recessive	α	α	$\alpha\gamma$	1	γ
Common dominant	α	$\alpha\gamma$	$\alpha\gamma$	γ	γ

Note.—$\alpha \neq 0$; and $\gamma > 1$.

Table 3

Sample Sizes Necessary to Gain 80% Power in ASP Linkage Analysis, in TDT with Singletons, and in TDT with ASPs

	Sample Size Necessary to Gain 80% Power											
					TDT							
	ASP Linkage Analysis				With Singletons				With ASPs			
γ and p	Multiplicative	Additive	Recessive	Dominant	Multiplicative	Additive	Recessive	Dominant	Multiplicative	Additive	Recessive	Dominant
4.0:												
.01	7,912	9,399	2.5×10^8	9,778	523	549	4.3×10^6	562	59	67	7.0×10^5	70
.10	342	813	30,617	1,055	86	123	5,056	153	20	36	866	49
.50	551	2,399	1,022	9,163	103	222	205	712	71	143	104	516
.80	3,737	18,264	4,320	4.0×10^5	291	663	337	9,873	293	518	312	7,810
2.0:												
.01	5.5×10^5	5.5×10^5	$>10^{10}$	6.0×10^5	4,154	4,154	3.8×10^7	4,317	1,018	1,018	1.2×10^7	1,083
.10	9,995	9,995	2.2×10^6	19,300	533	533	43,331	766	161	161	13,762	268
.50	4,638	4,638	15,743	57,693	340	340	949	1,861	188	188	431	1,133
.80	22,134	22,134	30,335	2.1×10^6	750	750	976	22,728	553	553	661	15,149
1.5:												
.01	8.6×10^6	7.8×10^6	$>10^{10}$	9.0×10^6	16,008	15,550	1.5×10^8	16,487	5,366	5,121	6.1×10^7	5,579
.10	1.3×10^5	62,068	3.4×10^7	2.1×10^5	1,908	1,485	1.7×10^5	2,554	704	490	67,104	1,012
.50	33,428	7,197	1.5×10^5	3.4×10^5	949	464	3,078	4,599	492	237	1,423	2,575
.80	1.3×10^5	24,881	1.9×10^5	1.1×10^7	1,833	816	2,553	51,914	1,156	583	1,509	31,206

which can be written in terms of GRR_1 and GRR_2 by dividing both the numerator and denominator by $f_0(\neq 0)$:

$$h_s = \frac{pq(p\mathrm{GRR}_2 + \mathrm{GRR}_1 + q)}{p^2\mathrm{GRR}_2 + 2pq\mathrm{GRR}_1 + q^2} .$$

The probability that the parent transmits allele A to the SAO is

$$\begin{aligned}\tau_s &= P(\text{transmit A/heterozygous and SAO}) \\ &= \frac{pf_2 + qf_1}{(pf_2 + f_1 + qf_0)} .\end{aligned}$$

I define the following RV, B_i, such that $B_i = (2pq)^{-1/2}$ if the parent is heterozygous and A is transmitted, $B_i = -(2pq)^{-1/2}$ if the parent is heterozygous and a is transmitted, and $B_i = 0$ if the parent is homozygous.

This definition for B_i differs from that described by Risch and Merikangas (1996), but I believe this to be a more appropriate definition (M. Iles and D. T. Bishop, personal communication), since the values for the RV should not depend on the hypothesis assumed. Under the null hypothesis of no linkage and no association (the TDT is a valid test for both, under these conditions), $P[B_i = (2pq)^{-1/2}] = pq$, $P[B_i = -(2pq)^{-1/2}] = pq$, and $P(B_i = 0) = p^2 + q^2$. The mean and the variance for B_i under H_0 are therefore 0 and 1, respectively. In a sample of N families (two parents and an SAO), there are $2N$ independent observations of B_i, giving a total mean and variance, over all families, of 0 and $2N$, respectively. Under the alternative hypothesis, $P[B_i = (2pq)^{-1/2}] = h_s\tau_s$, $P[B_i = -(2pq)^{-1/2}] = h_s(1 - \tau_s)$ and $P(B_i = 0) = 1 - h_s$, with h_s and τ_s as defined above. The mean and variance under H_1 are given by $2N(2pq)^{-1/2}h_s(2\tau_s - 1)$ and $2N(2pq)^{-1}h_s[1 - h_s(2\tau_s - 1)^2]$, respectively, summed over all N families. With these quantities, the number of families necessary to achieve 80% power under any genetic mechanism can be calculated.

$$N \geqslant \frac{\left\{z_{1-\alpha}\sqrt{2pq} - z_\beta\sqrt{h_s\,[1 - h_s\,(2\tau_s - 1)^2]}\right\}^2}{2h_s^2(2\tau_s - 1)^2} .$$

Table 3 illustrates sample sizes needed for a range of MOIs and values of p. A critical value of 5×10^{-8} has been used to determine genomewide significance for the TDT. This constitutes a Bonferroni correction accounting for both alleles at five biallelic markers being tested (one-sided test) in each of 100,000 genes, with each test being independent.

TDT Linkage Analysis Using ASPs

In this case I assume that N families consisting of an ASP plus both parents have been collected. The probability that a parent is heterozygous, given that it has an ASP, is

$$\begin{aligned}h_p &= P(\text{heterozygous/ASP}) \\ &= \{pq[(pf_2 + f_1 + qf_0)^2 + p(f_2 + f_1)^2 + q(f_1 + f_0)^2]\}/ \\ &\quad [4p^4f_2^2 + 4p^3q(f_2 + f_1)^2 + p^2q^2(f_2 + 2f_1 + f_0)^2 \\ &\quad + 8p^2q^2f_1^2 + 4pq^3(f_1 + f_0)^2 + 4q^4f_0^2] ,\end{aligned}$$

and the probability that the parent transmits allele A to an affected offspring is

$$\begin{aligned}\tau_p &= P(\text{transmit A/heterozygous and ASP}) \\ &= [(pf_2 + qf_1)(pf_2 + f_1 + qf_0) + pf_2(f_2 + f_1) \\ &\quad + qf_1(f_1 + f_0)]/[(pf_2 + f_1 + qf_0)^2 + p(f_2 + f_1)^2 \\ &\quad + q(f_1 + f_0)^2] .\end{aligned}$$

The same definition for the RV B_i as for the singleton case was used. Under the null hypothesis of no linkage (the TDT is valid only as a test for linkage when multiple affected offspring are used), $P[B_i = (2pq)^{-1/2}] = pq$, $P[B_i = -(2pq)^{-1/2}] = pq$, and $P(B_i = 0) = p^2 + q^2$. The mean and variance under H_0 are therefore given by 0 and 1, respectively. In a sample of N families (two parents and two offspring) there are $4N$ independent observations of B_i, giving a total mean and variance, over all families, of 0 and $4N$, respectively. Under the alternative hypothesis, $P[B_i = (2pq)^{-1/2}] = h_p\tau_p$, $P[B_i = -(2pq)^{-1/2}] = h_p(1 - \tau_p)$, and $P(B_i = 0) = 1 - h_p$, with h_p and τ_p as defined above. The mean and variance under H_1 are given by $4N(2pq)^{-1/2}h_p(2\tau_p - 1)$ and $4N(2pq)^{-1}h_p[1 - h_p(2\tau_p - 1)^2]$, respectively, summed over the N families. With these quantities the number of families necessary to achieve 80% power under any genetic mechanism can be calculated:

$$N \geqslant \frac{\left\{z_{1-\alpha}\sqrt{2pq} - z_\beta\sqrt{h_p\,[1 - h_p(2\tau_p - 1)^2]}\right\}^2}{4h_p^2(2\tau_p - 1)^2} .$$

Table 3 illustrates sample sizes needed for a range of MOIs and values of p. A critical value of 5×10^{-8} has been used to determine genomewide significance.

Results

The results in table 3 show that, in line with results from Risch and Merikangas (1996), the number of families necessary to perform a genomewide screen using the

TDT (singletons or sib-pairs) is smaller, in all cases, than that necessary for standard ASP linkage analysis. This observation holds true independently of the MOI or the relative values of GRR_1 and GRR_2 that are assumed. Similarly, it also has been found that the number of families necessary to perform the genomewide TDT using affected pairs is fewer, in all cases, than that necessary for singletons. For low values of p, the relative difference between using families with ASPs rather than those with an affected singleton is as large as fourfold (increasing as the value for γ increases). With the existence of a rare disease allele, therefore, the collection of ASPs would perhaps be more advantageous, since less genotyping would be necessary to gain the same power of study. For larger values (>.5) of p, the relative difference is much less, and therefore it could be argued that the advantage of reduced genotyping is offset by the increased difficulty in collecting ASPs versus collecting singletons. Another consideration is that, with singletons, the TDT is valid as a test for both association and linkage but that, with ASPs, it is valid only as a test for linkage (Spielman and Ewens 1996).

General observations on the effect of MOI on sample size include the intuitive relationship between additive and multiplicative GRRs. Specifically, for $\gamma > 2$ the sample sizes necessary under a multiplicative MOI are less than those under an additive model. For $\gamma = 2$ additive and multiplicative models are the same—and hence so are the sample sizes that are necessary—and for $\gamma < 2$ the sample sizes necessary under the additive MOI are smaller. A relationship also exists between the common recessive and common dominant MOIs based on p. For small values of p, the sample sizes needed under a dominant model are smaller than those needed under a recessive model, and the reverse is true for large values of p.

The most significant result illustrated in table 3 is the indication that the pattern of GRR at the disease locus may change—by several orders of magnitude, in some cases—the sample size necessary to perform a genomewide TDT (or ASP linkage). This result has important consequences for sample collection and in determining what type of effects could realistically be expected to be identified in such a study. For example, if $p = .01$ and $\gamma = 4.0$, the number of families necessary to perform a genomewide TDT with singletons varies from ~550, in the multiplicative, additive, and common-dominant cases, to 4.3×10^6 if the MOI is common recessive. As another illustration, table 4 shows the γ values for which loci could be detected with 80% power when 1,000 SAO-parent trios are used. It can be seen that a sample size of 1,000 would be sufficient to detect loci with γ's of fairly small magnitude (~1.00–3.15) under multiplicative and additive models—and that this number would be similarly sufficient under a common-recessive model if $p \geqslant .5$ and under a common-dominant model if $p \leqslant .5$. If, however, small-effect loci existed either with a recessive-type inheritance with $p < .5$ or with a dominant model with $p > .5$, they would almost definitely be missed.

Table 4

Values of γ for Which Loci Can Be Detected with 80% Power in a Genomewide TDT Using Singletons (N = 1, 000)

	Value of γ for Which Locus Can Be Detected with 80% Power			
p	Multiplicative	Additive	Recessive	Dominant
.01	3.114	3.150	212.4	3.182
.10	1.707	1.653	8.072	1.853
.50	1.484	~1.000	1.968	2.875
.80	1.785	~1.000	1.981	$>10^{10}$

Discussion

The four specific inheritance models considered here are obviously not exhaustive, but they clearly illustrate that patterns of GRRs at disease loci have substantial implications for the necessary sample size. We will, of course, rarely know all the loci involved and their inheritance patterns, and therefore a sample-size estimate based on a known MOI is of little value, although one application may be to use estimates for GRRs at any candidate-gene loci, to ensure that these at least could be detected/replicated. By considering a sensible range of inheritance patterns, however, and creating a table such as table 3, we can calculate highly informative guidelines on necessary sample size for unknown loci. Once a sample size has been chosen and collected, a table such as table 4 would indicate the magnitudes of gene effects that could be detected with 80% power. Such information, indicating what a study is likely (and unlikely) to detect, is of great value and should lead to a reduction in the number of false-negative conclusions.

As mentioned above, Müller-Myhsok and Abel (1997) showed that, if less than maximum linkage disequilibrium exists between marker and true disease locus, and if differences between the relative allele frequencies exist, then the sample size necessary increases. These possibilities must also, therefore, be taken into account when one is determining a sufficient sample size. As Risch and Merikangas (1997) point out in their response, however, the outlook may be optimistic. They mention two particular disease-locus regions, the apoE region (involved in Alzheimer disease) and the insulin VNTR region on chromosome 11p, in which both strong linkage disequilibrium and comparable allele frequencies exist (Ju-

lier et al. 1991; Chartier-Harlin et al. 1994; Bennett et al. 1995). The issue, nevertheless, should not be dismissed, and any knowledge of the degree of linkage disequilibrium across the genome should be considered, especially within any candidate genes.

The equations presented here involve the assumption that, conditional on the genotype at the trait locus in question, disease occurrences in sib pairs are independent. This is unlikely to be the case for complex traits, which are influenced by several genetic loci and environmental factors. Clearly, this assumption has no influence on sample sizes calculated for TDT singleton analysis, since sib pairs are not involved. For ASP linkage analysis using allele sharing or TDT, however, violation of the assumption could influence the calculations, and this may be worth considering when one is drawing comparisons.

The γ values considered here are small (4.0, 2.0, and 1.5) and correspond to a range for λ_s, [$= 1 + (\frac{1}{2} V_A + \frac{1}{4} V_D)/K^2$], from just over 1.0 to 1.5, depending on the value of p and on the inheritance model. These values are small but are likely to be reasonable for the disease loci of truly polygenic diseases. Such complex—and invariably common—diseases are where standard ASP genomewide screening is currently producing fewer answers—and where, arguably, genomewide TDT will be most useful. On a positive note, these diseases do have the added advantage of being easy to collect in large numbers, especially among SAO families, because of their relatively high prevalence. The possibility of a new, more sensitive genomewide screening method should not, however, be viewed as the means to detect all disease loci. The genomewide approach bypasses the need for any knowledge of the biological system or environmental interactions and, as such, will miss any disease loci that interact and are only of consequence together.

The future will provide the technology to perform genomewide TDT studies, and in the meanwhile we can prepare by initiating family collections or by expanding current ones. With information both on GRR patterns, as outlined in this paper, and on linkage disequilibrium in the genome as it becomes known, guidelines on the number of families necessary can be calculated. The genomewide TDT screen will be an extremely useful tool in the future genetic study of polygenic diseases—but it should be viewed as one approach among the many that we have already.

Acknowledgments

The author would like to thank Mr. Mark Iles and Dr. Angela Cox for their useful comments. This work was supported by Arthritis and Rheumatism Council grant D0528.

Appendix A

Calculation of $P(B_i = 1)$ under H_1 for Standard ASP Linkage Analysis

I seek to show that $P(B_i = 1) = Y$ and $P(B_i = -1) = 1 - Y$, where $B_i = 1$ if the ASP shares an allele from the ith parent and $B_i = -1$ if it does not and where Y is the expected proportion of alleles shared IBD by an ASP.

In an ASP family, let i and j be the two parents. Then,

$$Y = \sum_{n=0}^{2} \frac{n}{2} P(n_i + n_j = n) \ ,$$

where n is the number of alleles shared IBD by the ASP and where n_i and n_j are the number of alleles (0 or 1) shared from parent i and j, respectively.

Each parent transmits alleles independently and identically, so we can write

$$\begin{aligned} Y &= \frac{0}{2} P(n_i + n_j = 0) + \frac{1}{2} P(n_i + n_j = 1) \\ &\quad + \frac{2}{2} P(n_i + n_j = 2) \\ &= \frac{1}{2} [P(n_i = 0)P(n_j = 1) + P(n_i = 1)P(n_j = 0)] \\ &\quad + P(n_i = 1)P(n_j = 1) \\ &= \frac{1}{2} [P(n_i = 0)P(n_i = 1) + P(n_i = 1)P(n_i = 0)] \\ &\quad + P(n_i = 1)P(n_i = 1) \ . \end{aligned}$$

Now, $P(n_i = 0) = P(B_i = -1)$ and $P(n_i = 1) = P(B_i = 1)$, such that

$$\begin{aligned} Y &= P(B_i = -1)P(B_i = 1) + P^2(B_i = 1) \\ &= P(B_i = 1) [P(B_i = -1) + P(B_i = 1)] = P(B_i = 1) \ . \end{aligned}$$

Hence, $P(B_i = 1) = Y$ and $P(B_i = -1) = 1 - Y$, since $P(B_i = 1) + P(B_i = -1) = 1$.

Appendix B

Calculation of the Number of Families Necessary for Genomewide Screens

Let H_0 be

$$B = \sum_{i=0}^{xN} B_i \sim \text{norm}(0, xN) \ ,$$

and let H_1 be

$$B = \sum_{i=0}^{xN} B_i \sim \text{norm}(\mu xN, \sigma^2 xN) \ ,$$

where μ and σ^2 are the mean and variance, respectively, of a single observation of B_i under H_1 and where x is the number of observations of B_i per family.

We require that $P(\text{type I error}) = \alpha$ and that $P(\text{type II error}) = \beta$; that is, power $= (1 - \beta)$. First, we will therefore accept H_0 at the α level, in a one-sided test, if $[(B - 0)/\sqrt{xN}\,] < z_{1-\alpha}$; that is, $B < z_{1-\alpha}\sqrt{xN}$, where $z_{(1-a)} = \Phi^{-1}(1 - \alpha)$, with $\Phi()$ being the CDF of the standard normal distribution.

Second, to achieve a type II error of β, we require that $P(B < z_{1-\alpha}\sqrt{xN}/H_1 \text{ is true}) \leq \beta$, which, after being standardized, becomes

$$\Phi\left(\frac{z_{1-\alpha}\sqrt{xN} - \mu xN}{\sigma\sqrt{xN}}\right) \leq \beta \ ,$$

which, after being rearranged, can be written as $N \geqslant [(z_{1-\alpha} - \sigma z_{\beta})^2/x\mu^2]$. In the case of standard ASP linkage analysis, $x = 2$, $\mu = 2Y - 1$, $\sigma^2 = 4Y(1 - Y)$, $\alpha = 2.2 \times 10^{-5}$, and $\beta = .2$ (i.e., 80% power). For a genomewide screen using the TDT with singletons, $x = 2$, $\mu = (2pq)^{-1/2}h_s(2\tau_s - 1)$, $\sigma^2 = (2pq)^{-1}h_s[1 - h_s(2\tau_s - 1)^2]$, $\alpha = 5 \times 10^{-8}$, and $\beta = .2$. Similarly, for TDT with ASPs, $x = 4$, $\mu = (2pq)^{-1/2}h_p(2\tau_p - 1)$, $\sigma^2 = (2pq)^{-1}h_p[1 - h_p(2\tau_p - 1)^2]$, $\alpha = 5 \times 10^{-8}$, and $\beta = .2$.

References

Bennett ST, Lucassen AM, Gough SCL, Powell EE, Undlien DE, Pritchard LE, Merriman ME, et al (1995) Susceptibility to human type 1 diabetes at IDDM2 is determined by tandem repeat variation at the insulin gene minisatellite locus. Nat Genet 9:284–292

Chartier-Harlin M-C, Parfitt M, Legrain S, Pérez-Tur J, Brousseau T, Evans A, Berr C, et al (1994) Apolipoprotein E, ϵ4 allele as a major risk factor for sporadic early and late-onset forms of Alzheimer's disease: analysis of the 19q13.2 chromosomal region. Hum Mol Genet 3:569–574

Field LL, Fothergill-Payne C, Bertrams J, Baur MP (1986) HLA-DR effects in a large German IDDM dataset. Genet Epidemiol Suppl 1:323–328

Julier C, Hyer RN, Davies J, Merlin F, Soularue P, Briant L, Cathelineau G, et al (1991) Insulin-IGF2 region on chromosome 11p encodes a gene implicated in HLA-DR4-dependent diabetes susceptibility. Nature 354:155–159

Lander E, Kruglyak L (1995) Genetic dissection of complex traits: guidelines for interpreting and reporting linkage results. Nat Genet 11:241–247

Müller-Myhsok B, Abel L (1997) Genetic analysis of complex disease. Science 275:1328–1329

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

——— (1997) Genetic analysis of complex disease. Science 275:1329–1330

Spielman RS, Ewens WJ (1996) The TDT and other family-based tests for linkage disequilibrium and association. Am J Hum Genet 59:983–989

Spielman RS, McGinnis RE, Ewens WJ (1993) Transmission test for linkage disequilibrium: the insulin gene region and insulin-dependent diabetes mellitus (IDDM). Am J Hum Genet 52:506–516

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