

SAMPLE SIZES FOR THE TRANSMISSION DISEQUILIBRIUM TESTS: TDT, S-TDT AND 1-TDT

Dai Wang¹ and Fengzhu Sun^{1,2}

¹Department of Genetics
Emory University
Atlanta, GA 30322

²Department of Mathematics
University of Southern California
Los Angeles, CA 90089

Key Words: gene; association; linkage disequilibrium; power.

ABSTRACT

The transmission/disequilibrium test (TDT) is widely used to detect the linkage disequilibrium between a candidate locus (a marker) and a disease locus. The TDT is a family-based design and has the advantage that it is a valid test when population stratification exists. The TDT requires the marker genotypes of affected individuals and their parents. For diseases with late age of onset, it is difficult or impossible to obtain the marker genotypes of the parents. Therefore, when both parents' marker genotypes are unavailable, Ewens and Spielman extended the TDT to the S-TDT for use in sibships with at least one affected individual and one unaffected individual. When only one of the parents' genotype is available, Sun et al. proposed a test, the 1-TDT, for use with marker genotypes of affected individuals and only one available parent. Here, we study the sample sizes of TDT, S-TDT, and 1-TDT. We show that the sample size needed for the 1-TDT is roughly the same as the sample size needed for the S-TDT with two sibs and is about twice the sample size needed for the TDT.

1. INTRODUCTION

The transmission/disequilibrium test (TDT) was introduced by Spielman et al. (1993) to test the linkage disequilibrium between a candidate marker and a susceptible disease gene. If linkage disequilibrium is detected, the marker

may be within or linked to the susceptible disease gene. Risch and Merikangas (1996) showed that the TDT is more powerful than allele sharing methods to detect linkage if the disease is common and the relative risks are low. The TDT tests the unequal transmission of alleles from the heterozygous parents to the affected offspring. It does not produce spurious association due to population stratification (Ewens and Spielman 1995). The TDT requires the marker genotypes of affected individuals and their parents. For diseases with late age of onset or short survival time, it is difficult or impossible to obtain the marker genotypes of the parents. Recently, several new tests have been proposed to detect linkage disequilibrium when the parental genotypes are not all available.

Several methods have been proposed to test the linkage between a marker locus and the disease locus for sibships with at least one affected and one unaffected individual, when neither parents' genotypes are available (Curtis 1997, Spielman and Ewens 1998, Boehnke and Langefeld 1998, Horvath and Laird 1998). The basic idea is to compare the high risk allele frequency in affected individuals with the high risk allele frequency in unaffected individuals. Simulations have been carried out to compare the power of the different tests (Monks et al. 1998). It was concluded that the S-TDT developed by Spielman and Ewens (1998) is generally the most powerful of all the proposed tests. No sample size formula was available for the S-TDT.

Recently, Knapp (1999b) proposed the RC-TDT that can be used to analyze families conditional on that the parental genotypes can be inferred. For some families, the parental genotypes cannot be inferred, for example when there is one affected offspring and one available parent. Sun et al. (1998, 1999) developed a test, the 1-TDT, that can be used to analyze families with one affected offspring and one parent. Weinberg (1999) proposed a unified framework for the different transmission disequilibrium tests using the EM algorithm. In Sun et al. (1998, 1999) and Weinberg (1999), it was assumed that the availability of the parents does not depend on the marker genotype under study.

In the above studies, it was assumed that the genotype of each individual is known. For complex diseases, a large number of families are needed to detect the association of the disease with certain genetic markers. DNA pooling can

potentially greatly reduce the genotyping cost in association studies. Shaw et al. (1998) showed the experimental feasibility of DNA pooling. Risch and Teng (1998) studied the power of different tests for association using data from DNA pooling experiments.

The sample size for the TDT has been extensively studied (Risch and Merikangas 1996, Camp 1997, Knapp 1999a, Tu and Whittemore 1999) under a variety of different conditions. The power of the S-TDT was only studied using simulations (Monks et al. 1998). No sample size formulas were available for the S-TDT and the 1-TDT. In this paper, we develop sample size formulas for the S-TDT and the 1-TDT. We also compare the efficiency of the different tests under a variety of different conditions.

This paper is organized as follows. In section 2, we explain the genetic model and present the general method to obtain the sample sizes. In section 3, we give the sample size required to detect linkage disequilibrium and compare the relative power of the TDT, S-TDT and 1-TDT. A real data example is also given in this section. In section 4, we show the limitations of our study and point out directions for future research.

2. METHODS

2.1. A Genetic Model

We consider a marker locus with alleles “M” and “m” and a disease locus with alleles “D” and “d”. “D” is the high risk susceptible disease allele. As in Risch and Teng (1998), we assume a general disease model for penetrances at the disease locus. Let the penetrances for genotypes “DD”, “Dd” and “dd” be f_2 , f_1 and f_0 , respectively. We also assume that the penetrances, f_2 , f_1 and f_0 , are low, so unaffected individuals can be treated as random. Under this assumption, we can reduce the penetrance parameters to relative penetrance parameters. Fix $f_0 = 1$, then f_1 is the relative risk of genotype “Dd” versus “dd”, and f_2 is the relative risk of genotype “DD” versus “dd”. For a dominant model, we have $f_2 = f_1$; for a recessive model, we have $f_1 = 1$; for an additive model, $f_2 = 2f_1 - 1$; and for a multiplicative model, $f_2 = f_1^2$.

For simplicity, we assume that the marker locus and the disease locus are in complete disequilibrium. Let the allele frequency of “M” be p and the allele frequency of allele “m” be $q = 1 - p$, respectively. In the following, we denote

the genotype “MM”=2, “Mm”=1 and “mm”=0, where the number indicates the number of “M” alleles a genotype has. Throughout this paper, we assume random mating.

2.2. The General Formulation

To obtain the sample size required for the S-TDT and the 1-TDT, we use the general idea in Knapp (1999a) for the sample size formulas of the TDT. Suppose we have the marker genotype data from n families of the same structure. For the TDT, we have the genotype data of n affected individuals together with the genotypes of their parents. For the S-TDT, we have the genotype data of individuals in n sibships in which a of them are affected and u of them are unaffected. For the 1-TDT, we have the genotype data of n affected offspring and the genotype of their one available parent. The sample size issue for a mixture of families with different structures will be addressed in the Discussion section.

The general statistic to test the association of a genetic marker with the disease can be written as

$$X_n = \frac{\sum_{i=1}^n U(i)}{\sqrt{\sum_{i=1}^n V(i)}},$$

where $(U(i), V(i))$, $i = 1, 2, \dots, n$ are independent identically distributed random vectors. $(U(i), V(i))$ are usually chosen so that X_n has an approximate standard normal distribution under the null hypothesis of no association. The definition of $(U(i), V(i))$ for the S-TDT and the 1-TDT will be given in the following sections.

Suppose $(U(1), V(1))^T$ is a random vector with expectation $(\mu, \nu)^T$ and covariance matrix $\begin{pmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{pmatrix}$, and $\nu > 0$. From the general statistical theory (Rao 1973), let $g(x, y) = x/\sqrt{y}$, we have

$$\frac{\sum_{i=1}^n U(i)}{\sqrt{\sum_{i=1}^n V(i)}} - \sqrt{n} \frac{\mu}{\sqrt{\nu}} = \sqrt{n} \left(g \left(\frac{1}{n} \sum_{i=1}^n U(i), \frac{1}{n} \sum_{i=1}^n V(i) \right) - g(\mu, \nu) \right) \\ \xrightarrow{D} Normal(0, \sigma^2),$$

where $\sigma^2 = \frac{1}{\nu} \sigma_1^2 - \frac{\mu}{\nu^2} \rho\sigma_1\sigma_2 + \frac{\mu^2}{4\nu^3} \sigma_2^2$.

As in Risch and Merikangas (1996), we consider one sided test. For a given point-wise type I error rate α , the null hypothesis of no association is rejected if $X_n > z_{1-\alpha}$ (or $X_n < z_\alpha$).

Under the alternative hypothesis of association, the power of the test is given by

$$\begin{aligned} 1 - \beta &= P(X_n > z_{1-\alpha}) \\ &\approx P\left(Z > \frac{z_{1-\alpha} - \sqrt{n}\mu/\sqrt{\nu}}{\sigma}\right) \\ &= 1 - \Phi\left(\frac{c - \sqrt{n}\mu/\sqrt{\nu}}{\sigma}\right). \end{aligned} \quad (1)$$

For a given point-wise type I error rate α , the sample size needed to have power $1 - \beta$ is given by

$$S = \left(\frac{z_{1-\alpha} + z_{1-\beta}\sigma}{\mu/\sqrt{\nu}}\right)^2. \quad (2)$$

(We can obtain the same formula in the $X_n < z_\alpha$ case.) In the rest of this paper, we set the genome wide type I error rate at 0.05 and approximately 10^6 bio-allelic markers are considered (Risch and Merikangas 1996, Tu and Whittemore 1999, Risch and Teng 1998). Using Bonferroni correction, the adjusted point-wise type I error rate is 5×10^{-8} . Then $z_{1-\alpha} = 5.33$. We also let the power be 80% corresponding to $z_{1-\beta} = 0.84$.

Using the above general approach, Knapp (1999a) gave the power and sample size formulas for the TDT. In the following sections, we study the sample size requirement for the S-TDT and the 1-TDT separately.

2.3. The S-TDT

The S-TDT compares the high risk allele frequency in affected individuals with the high risk allele frequency in unaffected individuals. Let a and u be the numbers of affected sibs and unaffected sibs within a family, respectively and $t = a + u$. Let r and s be the numbers of sibs with genotypes 2 and 1, respectively. The S-TDT statistic is given by (Spielman and Ewens 1998)

$$T_S = \frac{Y_S - A_S}{\sqrt{V_S}},$$

where Y_S is the number of ‘‘M’’ alleles among the affected sibs.

$$A_S = \sum (2r + s)a/t$$

and

$$V_S = \sum au[4r(t - r - s) + s(t - s)]/[t^2(t - 1)],$$

are the mean and variance of Y_S , respectively, under the null hypothesis of no linkage disequilibrium.

Let $Y_S(i)$, $A_S(i)$ and $V_S(i)$ be the corresponding values of Y_S , A_S and V_S within the i -th family, respectively. Then

$$Y_S = \sum_{i=1}^n Y_S(i), \quad A_S = \sum_{i=1}^n A_S(i), \quad V_S = \sum_{i=1}^n V_S(i).$$

Let $U_S(i) = Y_S(i) - A_S(i)$, then

$$T_S = \frac{\sum_{i=1}^n U_S(i)}{\sqrt{\sum_{i=1}^n V_S(i)}}.$$

Here, we consider sibships with one affected and one unaffected first. According to the numbers of sibs with genotypes 2 and 1, there are 6 types of sibships: 1) $r = 2, s = 0$; 2) $r = 1, s = 1$; 3) $r = 1, s = 0$; 4) $r = 0, s = 2$; 5) $r = 0, s = 1$; 6) $r = 0, s = 0$. Denote the conditional probabilities of the six types of sibships given one affected and one unaffected by $p_i, i = 1, 2, \dots, 6$, respectively. We have

$$\begin{aligned} p_1 &= p^2(1 + p)^2 f_2 / (4K), \\ p_2 &= p^2 q(1 + p)(f_1 + f_2) / (2K), \\ p_3 &= p^2 q^2 (f_2 + 1) / (4K), \\ p_4 &= pq(1 + pq) f_1 / K, \\ p_5 &= pq^2(1 + q)(f_1 + 1) / (2K), \\ p_6 &= q^2(1 + q)^2 / (4K). \end{aligned} \tag{3}$$

Here K is the probability that there is one affected and one unaffected in a sibship. Since we are assuming low penetrance, we have $K = p^2 f_2 + 2pqf_1 + q^2$. The procedure for the calculation of $p_i, i = 1, 2, \dots, 6$, is given in the appendix.

Table I gives the possible values of $V_S(i), A_S(i), Y_S(i)$ and $U_S(i)$ according to different types of sibships and their corresponding probabilities.

TABLE I.

Possible values of $V_S(i)$, $A_S(i)$, $Y_S(i)$ and $U_S(i)$
and the corresponding probabilities.

	Sib genotype (r, s)	Genotype of affected	Probability	$V_S(i)$	$A_S(i)$	$Y_S(i)$	$U_S(i)$
1	(2, 0)	2	p_1	0	2	2	0
2	(1, 1)	2	$\frac{f_2}{f_1 + f_2} p_2$	$\frac{1}{4}$	$\frac{2}{3}$	2	$\frac{1}{2}$
		1	$\frac{f_1}{f_1 + f_2} p_2$	$\frac{1}{4}$	$\frac{2}{3}$	1	$-\frac{1}{2}$
3	(1, 0)	2	$\frac{f_2}{f_2 + 1} p_3$	1	1	2	1
		0	$\frac{1}{f_2 + 1} p_3$	1	1	0	-1
4	(0, 2)	1	p_4	0	1	1	0
5	(0, 1)	1	$\frac{f_1}{f_1 + 1} p_5$	$\frac{1}{4}$	$\frac{1}{2}$	1	$\frac{1}{2}$
		0	$\frac{1}{f_1 + 1} p_5$	$\frac{1}{4}$	$\frac{1}{2}$	0	$-\frac{1}{2}$
6	(0, 0)	0	p_6	0	0	0	0

The conditional distribution of $(U_S(i), V_S(i))^T$ given one affected sib and one unaffected sib is

$$\begin{aligned}
 (U_S(i), V_S(i)) : & \text{Probability} \\
 (0, 0) : & p_1 + p_4 + p_6 \\
 \left(-\frac{1}{2}, \frac{1}{4}\right) : & \frac{f_1}{f_2 + f_1} p_2 + \frac{1}{f_1 + 1} p_5 \\
 \left(\frac{1}{2}, \frac{1}{4}\right) : & \frac{f_2}{f_2 + f_1} p_2 + \frac{f_1}{f_1 + 1} p_5 \\
 (-1, 1) : & \frac{1}{f_2 + 1} p_3 \\
 (1, 1) : & \frac{f_2}{f_2 + 1} p_3
 \end{aligned} \tag{4}$$

Then for $i = 1, 2, \dots, n$, we have

$$EU_S(i) = \frac{f_2 - f_1}{2(f_2 + f_1)} p_2 + \frac{f_2 - 1}{f_2 + 1} p_3 + \frac{f_1 - 1}{2(f_1 + 1)} p_5;$$

$$EV_S(i) = E(U_S(i))^2 = \frac{1}{4} p_2 + p_3 + \frac{1}{4} p_5;$$

$$E(V_S(i))^2 = \frac{1}{16} p_2 + p_3 + \frac{1}{16} p_5;$$

$$EU_S(i)V_S(i) = \frac{f_2 - f_1}{8(f_2 + f_1)}p_2 + \frac{f_2 - 1}{f_2 + 1}p_3 + \frac{f_1 - 1}{8(f_1 + 1)}p_5. \quad (5)$$

From (5), we can calculate the mean and variance of T_S . The sample size needed for S-TDT can then be calculated from equation (2).

For sibships with more than one affected sib and more than one unaffected sib, we can use the same method to calculate the number of families needed to detect linkage disequilibrium.

2.4. The 1-TDT

When the genotype of an affected individual and the genotype of only one parent are available, we use the 1-TDT. The 1-TDT statistic is (Sun et al. 1999)

$$T_1 = \frac{A_{01} + A_{12} - A_{10} - A_{21}}{\sqrt{V_1}} = \frac{b_1 - c_1}{\sqrt{V_1}},$$

where A_{ij} , $i, j = 0, 1, 2$ is the number of affected individuals with genotype i whose one available parent has genotype j and

$$b_1 = A_{01} + A_{12}, \quad c_1 = A_{10} + A_{21}.$$

Here we assume that only the genotypes of one affected individual and one available parent are considered for each family. The variance of $b_1 - c_1$ can be estimated by $V_1 = b_1 + c_1$. T_1 has an approximate standard normal distribution when $b_1 + c_1$ is large under the null hypothesis.

Let $b_1(i)$ and $c_1(i)$ be the corresponding values of b_1 and c_1 , respectively within the i -th family. Then $b_1 = \sum_{i=1}^n b_1(i)$ and $c_1 = \sum_{i=1}^n c_1(i)$. Let $U_1(i) = b_1(i) - c_1(i)$ and $V_1(i) = b_1(i) + c_1(i)$, we have

$$T_1 = \frac{\sum_{i=1}^n U_1(i)}{\sqrt{\sum_{i=1}^n V_1(i)}}.$$

Table II gives the values of $b_1(i)$, $c_1(i)$, $U_1(i)$ and $V_1(i)$ for each family type and the conditional probability of each family type given one affected child. In the table, family type $i \rightarrow j(i, j = 0, 1, 2)$ indicates that the genotype of the affected offspring is j and the genotype of the available parent is j . $K_1 = p^2 f_2 + 2pqf_1 + q^2$ is the probability that there is one affected offspring.

TABLE II.

Possible values of $b_1(i)$, $c_1(i)$, $U_1(i)$ and $V_1(i)$
and their corresponding probabilities.

Family type	Probability given one affected child	$b_1(i)$	$c_1(i)$	$U_1(i)$	$V_1(i)$
$2 \rightarrow 2$	$p^3 f_2 / K_1$	0	0	0	0
$2 \rightarrow 1$	$p^2 q f_1 / K_1$	1	0	1	1
$1 \rightarrow 2$	$p^2 q f_2 / K_1$	0	1	-1	1
$1 \rightarrow 1$	$p q f_1 / K_1$	0	0	0	0
$1 \rightarrow 0$	$p q^2 / K_1$	1	0	1	1
$0 \rightarrow 1$	$p q^2 f_1 / K_1$	0	1	-1	1
$0 \rightarrow 0$	q^3 / K_1	0	0	0	0

So the distribution of $(U_1(i), V_1(i))^T$ is

$$\begin{aligned}
 (U_1(i), V_1(i)) : & \text{Probability} \\
 (0, 0) : & (p^3 f_2 + p q f_1 + q^3) / K_1 \\
 (1, 1) : & (p^2 q f_1 + p q^2) / K_1 \\
 (-1, 1) : & (p^2 q f_2 + p q^2 f_1) / K_1
 \end{aligned} \tag{6}$$

Then for $i = 1, 2, \dots, n$, we have

$$\begin{aligned}
 E U_1(i) &= E U_1(i) V_1(i) = p q ((2p - 1) f_1 - p f_2 + q) / K_1; \\
 E V_1(i) &= E (U_1(i))^2 = E (V_1(i))^2 = p q (p f_2 + f_1 + q) / K_1.
 \end{aligned} \tag{7}$$

From (7), we can calculate the mean and variance of T_1 . The sample size needed for S-TDT can then be calculated from equation (2).

3. RESULTS

We compare the numbers of families needed to detect linkage disequilibrium using the TDT, the S-TDT and the 1-TDT. For the TDT, we consider families with the genotypes of one affected offspring and the parents. For the S-TDT, we consider families with genotypes of one affected offspring and one unaffected offspring. For the 1-TDT, we consider families with the genotypes of one affected offspring and only one parent. Because the population frequency of the disease gene will influence the number of families needed to detect linkage disequilibrium, we consider three different frequencies: 1) $p = 0.05$, 2) $p = 0.20$, and 3) $p = 0.70$. The relative penetrance f_2 is fixed

to be 4. For different genetic models, we have 1) dominant model: $f_1 = 4$, 2) recessive model: $f_1 = 1$, 3) multiplicative model: $f_1 = 2$, and 4) additive model: $f_1 = 2.5$.

TABLE III.

Number of families needed to detect linkage equilibrium for four genetic models using the TDT, S-TDT and 1-TDT. $f_2 = 4$.

	TDT	S-TDT	1-TDT
Dominant			
$p = 0.05$	315	623	634
$p = 0.20$	231	447	471
$p = 0.70$	2,917	5,640	5,916
Recessive			
$p = 0.05$	38,880	79,610	77,118
$p = 0.20$	975	2,079	1,913
$p = 0.70$	204	432	404
Multiplicative			
$p = 0.05$	1,248	2,508	2,495
$p = 0.20$	423	859	846
$p = 0.70$	457	926	913
Additive			
$p = 0.05$	732	1,464	1,468
$p = 0.20$	340	679	683
$p = 0.70$	691	1,381	1,387

From Table III, we can see that the number of families needed for the 1-TDT is roughly the same as the number of families needed for the S-TDT with two sibs and is about twice the number of families needed for the TDT.

Example.

The warfarin-aspirin symptomatic intracranial disease study (WASID) (Chimowitz et al. 1995) is a multicenter study comparing the efficacy of warfarin with aspirin for the prevention of vascular events in intracranial disease patients funded by the US National Institute of Health. During the tenure of this award, 806 patients with intracranial athrosclerosis will be recruited and randomized to warfarin and aspirin. It was proposed (Dr. B. Stern, Personal communication) to find genetic mutations associated with intracranial athrosclerosis based on this cohort of 806 patients and their family based con-

trols, such as parents or sibs. Since the ages of the patients in this cohort are about 45-65, their parents are usually not available. It is estimated that about 10% of the patients, or 80 patients, have one parent available. About half of the other $806-80=726$ patients, or 363 patients, will have one unaffected sib available for study. We are interested in knowing the power to detect the association of the genetic mutations with intracranial atherosclerosis.

Since we showed above that the power of the S-TDT is roughly the same as the power of the 1-TDT, we treat the cases with one parent available the same as cases with one unaffected sib. Using equation (1), we are able to calculate the power of the S-TDT with $363+80=443$ cases for given high allele frequency and relative risk under dominant and multiplicative models (Table IV). Here the relative risk is the relative penetrance of genotype “Dd”.

TABLE IV.

Power of the S-TDT with 443 cases for given allele frequency and relative risk under dominant and multiplicative models.

	$f_1 = 2$	$f_1 = 4$	$f_1 = 6$
Dominant			
$p = 0.01$	7.48×10^{-6}	0.0010	0.0207
$p = 0.1$	0.0072	0.7757	0.9970
$p = 0.25$	0.0143	0.6858	0.9624
$p = 0.5$	0.0006	0.0303	0.0870
$p = 0.75$	3.50×10^{-6}	2.59×10^{-5}	4.93×10^{-5}
Multiplicative			
$p = 0.01$	7.76×10^{-6}	0.0012	0.0255
$p = 0.1$	0.0229	0.9915	1
$p = 0.25$	0.2308	1	1
$p = 0.5$	0.3179	1	1
$p = 0.75$	0.0688	0.9462	0.9994

4. DISCUSSION

In this paper, we used the general framework developed by Knapp (1999) to derive the sample size formulas for the S-TDT and the 1-TDT. We had done simulation to test the appropriateness of the approximation and found that the simulated power with the given sample size is very close to theoretical

power (data not shown). Using the sample size formulas, we calculated the sample sizes needed for the TDT, S-TDT, and 1-TDT under a variety of different disease models. From the results, we see that in most situations the sample size needed for the 1-TDT is roughly the same as the sample size needed for the S-TDT with one affected and one unaffected sibs and is about twice the sample size needed for the TDT. In addition, with a real example on intracranial atherosclerosis, we calculated the power of detecting the association of genetic markers with the disease under study with the available data set.

In this paper, we assumed that there is one affected and one unaffected in a sibship for the S-TDT. In reality, we may have different types of sibships available. With any given fixed structure of sibships, such as a affected and u unaffected in a sibship with a and u fixed, the general framework described in this paper can still be used to derive the power and the sample size formulas for the S-TDT. If the number of sibships for each type of structure is large, normal approximation can be used to approximate the power of the S-TDT. Generally, if the number of sibships is small, it might be best to use the permutation method developed by Spielman and Ewens (1998) to study the power. Similarly, we assumed that we have one affected offspring for the 1-TDT. A similar method can be used to study the power and sample size for the 1-TDT when multiple affected offspring are included.

APPENDIX

Here, we calculate the p_i , $i = 1, 2, \dots, 6$ in Section 2.3. Denote the event that there is one affected offspring and one unaffected offspring by ‘A’. Let

$$\begin{aligned} B_1 &= \{r = 2, s = 0\}; \\ B_2 &= \{r = 1, s = 1\}; \\ B_3 &= \{r = 1, s = 0\}; \\ B_4 &= \{r = 0, s = 2\}; \\ B_5 &= \{r = 0, s = 1\}; \\ B_6 &= \{r = 0, s = 2\}. \end{aligned}$$

r and s be the numbers of offspring with genotype 2 and 1 within the family, respectively. Let (i, j) denote the mating type, that is, the genotypes of the

two parents are i and j , respectively. We have

$$\begin{aligned}
 p_1 &= P(B_1 | A) \\
 &= \sum P((i, j) | A) P(B_1 | A \cap (i, j)) \\
 &= \sum \frac{P(A \cap (i, j))}{P(A)} \cdot \frac{P(B_1 \cap A \cap i, j)}{P(A \cap (i, j))} \\
 &= (P(B_1 \cap A \cap (2, 2)) + P(B_1 \cap A \cap (2, 1)) + P(B_1 \cap A \cap (1, 1))) / K \\
 &= \left(p^4 f_2 + 4p^3 q \cdot \frac{1}{4} f_2 + 4p^2 q^2 \cdot \frac{1}{16} f_2 \right) / K \\
 &= p^2(1 + p)^2 f_2 / (4K).
 \end{aligned}$$

We can use the same method to calculate p_i , $i = 2, 3, 4, 5, 6$.

ACKNOWLEDGEMENTS

This research is supported by NIDDK R29DK53392.

BIBLIOGRAPHY

- Boehnke, M. and Langefeld, C. D.(1998). “Genetic association mapping based on discordant sib pairs: The discordant-alleles test”, *Am. J. Hum. Genet.* **62**, 950–961.
- Camp, N. J.(1997). “Genomewide transmission/disequilibrium testing: Consideration of the genotypic relative risks at disease loci”, *Am. J. Hum. Genet.* **61**, 1424–1430.
- Chimowitz, M. I., Kokkinos, J., Strong, J., Brown, M. B., Levine, S. R., Silliman, S., Pessin, M. S., Weichel, E., Sila, C. A., Furlan, A. J., Kargman, D. E., Sacco, R. L., Wityk, R. J., Ford, G. and Fayad, P. B.(1995). “The Warfarin-Aspirin symptomatic intracranial disease study”, *Neurology* **45**, 1488-1493.
- Curtis, D.(1997). “Use of siblings as controls in case-control association studies”, *Ann. Hum. Genet.* **61**, 319–333.
- Ewens, W. J. and Spielman, R. S.(1995). “The transmission disequilibrium test: History, subdivision and admixture”, *Am. J. Hum. Genet.* **57**, 455–464.
- Horvath, S. and Laird, N. M.(1998). “A discordant-sibship test for disequilibrium and linkage: No need for parental data”, *Am. J. Hum. Genet.* **63**, 1886–1897.
- Knapp, M.(1999a). “A note on power approximations for the transmission/disequilibrium test”, *Am. J. Hum. Genet.* **64**, 1177–1185.
- Knapp, M.(1999b). “The transmission/disequilibrium test and parental-

- genotype reconstruction: The reconstruction-combined transmission/disequilibrium test”, *Am. J. Hum. Genet.* **64**, 861–870.
- Monks, S. A., Kaplan, N. L and Weir, B. S.(1998). “A comparative study of sibship tests of linkage and/or association” *Am. J. Hum. Genet.* **63**, 1507–1516.
- Rao, C. R.(1973). *Linear statistical inference and its applications*, 2nd ed. John Wiley, New York.
- Risch, N. and Merikangas, K.(1996). “The future of genetic studies of complex human diseases”, *Science* **273**, 1516–1517.
- Risch, N. and Teng, J.(1998). “The relative power of family-based and case-control designs for linkage disequilibrium studies of complex human diseases I. DNA pooling”, *Genome Res.* **8**, 1273–1288.
- Shaw, S. H., Carrasquillo, M. M., Kashuk, C., Puffenberger, E. G. and Chakravarti, A.(1998). “Allele frequency distributions in pooled DNA samples: Applications to mapping complex disease genes”, *Genome Res.* **8**, 111–123.
- Spielman, R. S., McGinnis, R. E. and Ewens, W. J.(1993). “Transmission test for linkage disequilibrium: The insulin gene region and insulin-dependent diabetes mellitus (IDDM)”, *Am. J. Hum. Genet.* **52**, 506–516.
- Spielman, R. S. and Ewens, W. J.(1998). “A sibship test for linkage in the presence of association: The sib transmission/disequilibrium test”, *Am. J. Hum. Genet.* **62**, 450–458.
- Sun, F. Z., Flanders, W. D., Yang, Q. H. and Khoury, M. J.(1998). “A new method for estimating the risk ratio in studies using case-parental control design”, *Am. J. Epidemiol.* **148**, 902–909.
- Sun, F. Z., Flanders, W. D., Yang, Q. H. and Khoury, M. J.(1999). “The transmission disequilibrium test (TDT) when only one parent is available: the 1-TDT”, *Am. J. Epidemiol.* **150**, 97–104.
- Tu, I-Ping and Whittemore, A. S.(1999). “Power of association and linkage tests when the disease alleles are unobserved”, *Am. J. Hum. Genet.* **64**, 641–649.
- Weinberg, C. R.(1999). “Allowing for missing parents in genetic studies of case-parent triads”, *Am. J. Hum. Genet.* **64**, 1186–1193.