

Transmission/disequilibrium tests for quantitative traits

F. Z. SUN¹, W. D. FLANDERS², Q. H. YANG³ AND H. Y. ZHAO⁴

¹*Department of Mathematics, University of Southern California, Los Angeles, CA, USA*

²*Department of Epidemiology, Emory University, Atlanta, GA, USA*

³*Birth Defects and Genetic Diseases Branch, Centers for Disease Control and Prevention, Atlanta, GA, USA*

⁴*Department of Epidemiology and Public Health, Yale School of Medicine, New Haven, CT, USA*

(Received 1.3.00. Accepted 19.7.00)

SUMMARY

The transmission/disequilibrium test (TDT) is a powerful method of locating disease genes. The TDT was originally proposed for use in studies of qualitative traits in families with both parents available. Recently, the TDT has been extended to studies of qualitative traits in sibships without parents available and in families with one parent available. It has also been extended for use in studies of quantitative traits in families with both parents available and in sibships with multiple offspring. In this paper, we first propose a new class of TDT-type tests for linkage in the presence of linkage disequilibrium for use in studies of families with both parents available. The TDT of Spielman *et al.* (1993) for qualitative traits and the TDT of Rabinowitz (1997) for quantitative traits are special cases of the new tests. Second, we propose a new class of TDT-type tests for linkage for use in studies of families with one parent available. Third, we study the validity and the power of the tests using simulations. Finally, we propose a method of combining data from different types of families. The combined test is valuable and allows researchers full use of the available data in detecting linkage between a marker locus and an unobservable quantitative trait locus. An important feature of the tests proposed in this paper is that no assumptions on the distribution of the quantitative traits are needed.

INTRODUCTION

There has been considerable interest in locating disease genes or quantitative trait loci using the transmission/disequilibrium test (TDT). Spielman *et al.* (1993) first developed the transmission/disequilibrium test (TDT) for qualitative traits for use in studies of families with at least one affected offspring and two parents. The basic idea of the TDT is to compare the number of times an allele of interest is transmitted from heterozygous parents to affected offspring with the number of times some other alleles are transmitted from these same heterozygous parents to affected offspring. The TDT has the desired property of not giving spurious significant results even if there is population stratification (Ewens & Spielman, 1995; Spielman & Ewens, 1996). It has also been shown to be more powerful than allele-sharing methods for genomic screening if highly dense marker maps are available (Risch & Merikangas, 1996). The TDT has been extended for use in studies of genetic markers with multiple alleles (Sham & Curtis, 1995; Schaid, 1996; Lazzeroni & Lange, 1998), in studies of sibships with at least one affected and one unaffected individual (Curtis, 1997; Boehnke & Langefeld, 1998; Horvath & Laird, 1998; Spielman & Ewens, 1998; Teng & Risch, 1999), and in studies of families with one parent available (Sun *et al.* 1998, 1999; Weinberg, 1999). Knapp (1999)

Correspondence: Fengzhu Sun, PhD, Department of Mathematics, University of Southern California, 1042 W 36th Place, DRB155, Los Angeles, CA 90089, USA. Tel: (213) 740-2413; Fax: (213) 740-2424
E-mail: fsun@hnto.usc.edu

extended the TDT to studies of families in which the parental genotypes can be inferred. Several investigators proposed TDT-type tests for linkage and association studies of quantitative traits in families with both parents available (Allison, 1997; Rabinowitz, 1997; Xiong *et al.* 1998; George *et al.* 1999; Abecasis *et al.* 2000; Monks & Kaplan, 2000; Lunetta *et al.* 2000). Allison *et al.* (1999) proposed two TDT-type tests for studies of quantitative trait in sibships with at least two offspring: the mixed-effects model test and the permutation test.

In this paper, we first propose a new class of TDTs to test for linkage in the presence of linkage disequilibrium based on the idea of Rabinowitz (1997) for use in studies of families with both parents available. The TDT for qualitative traits proposed by Spielman *et al.* (1993) and the TDT for quantitative traits proposed by Rabinowitz (1997) are special cases of the general tests. Secondly, we propose a new class of tests for linkage based on the idea of Sun *et al.* (1998, 1999) for use in studies of quantitative traits applicable to families with only one parent available. The 1-TDT of Sun *et al.* (1999) is a special case of this new class of tests. Thirdly, we study the validity and the power of the different tests using simulations. Finally, we propose a method of combining data from different types of families.

METHODS

General assumptions and notation

Suppose we have a quantitative trait of interest, with n_i , $i = 0, 1, 2$, families having no, one, and two parents available for study, respectively. Here, the subscript ‘ i ’ in n_i denotes the number of parents in a family. Let $l_i(j)$, $i = 0, 1, 2$; $j = 1, 2, \dots, n_i$ be the number of offspring in the j -th family among families with i parents. The trait value of the k -th offspring in the j -th family is $Q_{j,k}^{(i)}$, $k = 1, 2, \dots, l_i(j)$. We also assume that the marker locus of interest has two alleles ‘M’ and ‘N’. The parents and the offspring are genotyped at a marker locus of interest. For families with both parents, we define the following index function: $Y_{j,k}^{(m)} = 1$ (or -1) if the mother in the j -th family is heterozygous and transmits the M (or N) allele to the k -th offspring, and $Y_{j,k}^{(m)} = 0$ if the mother is homozygous. We similarly define $Y_{j,k}^{(f)}$ for the father. For families with one parent, we only consider offspring–parent pairs with genotypes (NM, NN) or (MM, NM), and offspring–parent pairs with genotypes (NN, NM) or (NM, MM). The first genotype in the bracket is the offspring’s genotype and the second genotype is the available parent’s genotype (Sun *et al.* 1999). We define the index function $Y_{j,k}^{(1)}$ as $Y_{j,k}^{(1)} = 1$ if the offspring–parent genotypes are (NM, NN) or (MM, NM), $Y_{j,k}^{(1)} = -1$ if the offspring–parent genotypes are (NN, NM) or (NM, MM), and $Y_{j,k}^{(1)} = 0$ otherwise. Note that the index functions for families with two parents and for families with one parent differ by the superscript. N_m and N_f are the numbers of families with the mother and the father available, respectively.

TDTs for quantitative traits when both parents are available

Several investigators proposed various TDT-type tests for quantitative traits (Allison, 1997; Rabinowitz, 1997; Xiong *et al.* 1998; George *et al.* 1999; Abecasis *et al.* 2000). Here, we propose a new class of TDT-type tests for quantitative traits. The new class of tests includes the TDT for qualitative traits of Spielman *et al.* (1993) and the TDT for quantitative traits of Rabinowitz (1997) as special cases.

Rabinowitz (1997) first noted that, under the null hypothesis of no linkage between the marker locus and the quantitative loci, the trait value and the index functions $Y_{j,k}^{(f)}$ and $Y_{j,k}^{(m)}$ are conditionally independent, given the parental alleles. Thus, for any constant c , conditional on the trait values and the parental genotypes,

$$s_2(c) = \sum_{j=1}^{n_2} \sum_{k=1}^{l_2(j)} (Q_{j,k}^{(2)} - c)(Y_{j,k}^{(f)} + Y_{j,k}^{(m)})$$

has mean 0. Conditional on the trait values and the parental genotypes, $Y_{j,k}^{(f)}$ and $Y_{j,k}^{(m)}$ are independent under the null hypothesis. Thus the conditional variance of $s_2(c)$ can be estimated by

$$\sigma_2^2(c) = \sum_{j=1}^{n_2} \sum_{k=1}^{l_2(j)} (Q_{j,k}^{(2)} - c)^2 (|Y_{j,k}^{(f)}| + |Y_{j,k}^{(m)}|).$$

The class of statistics is given by

$$T_2(c) = s_2(c)/\sigma_2(c),$$

which has an approximate normal distribution when the number of heterozygous parents is large. When the number of families is small, simulations can be used to determine the p -value. In the simulations, we keep the trait values of the individuals under study at the observed values and determine the offspring's genotypes using Mendelian inheritance. After simulating the process for many runs, we can approximate the p -value by the fraction of times the simulated absolute values of $T_2(c)$ exceed the observed absolute value of $T_2(c)$.

In previous studies of TDTs for quantitative traits, investigators compared the trait values of individuals who inherited the M allele from heterozygous parents with the trait values of individuals who inherited the N allele from these same heterozygous parents. Thus, investigators usually assumed that the trait values were normally distributed or that the sample size was large so that normal approximation can be applied (Xiong *et al.* 1998). Here, we condition on the trait values of the individuals and study the transmission of the M allele versus that of the N allele from heterozygous parents. We do not make any assumptions about the distribution of the trait values, and the tests are valid for any types of sampling schemes based on phenotypes of the individuals under study. Like the TDT for qualitative traits, the class of tests is also applicable to studies of multiple generation families and families with multiple offspring.

As shown by the above arguments, $T_2(c)$ is a valid test for the null hypothesis of no linkage between the marker locus and the quantitative trait loci, that is, the type I error equals the pre-specified type I error rate. Certainly, the power of $T_2(c)$ is different for different values of c under the alternative hypothesis of linkage. If we let c be the average trait values, \bar{Q} , of all the offspring in the sample, $T_2(\bar{Q})$ is the same as the TDT of Rabinowitz (1997). We can also let c be the average trait value of interest, AQ , among the general population if it is known.

For qualitative traits, we can assign a trait value 1 if a person is affected and 0 if the person is unaffected. Then, with families having at least one affected offspring, $s_2(c) = (1-c)(b-b')$, $\sigma_2^2(c) = (1-c)^2(b+b')$, and $T_2(c) = (b-b')/\sqrt{b+b'}$, where $b(b')$ is the number of times M (N) is transmitted from heterozygous parents to the affected offspring. Thus, our test reduces to the TDT of Spielman *et al.* (1993) for qualitative traits.

Xiong *et al.* (1998) showed that their test is generally more powerful than Allison's basic test (Allison, 1997). We will show, in the Simulations section, that $T_2(AQ)$ and $T_2(\bar{Q})$ have roughly the same power as the test proposed by Xiong *et al.* (1998) under random sampling. For non-random sampling, $T_2(AQ)$ can be more powerful than $T_2(\bar{Q})$ and the test of Xiong *et al.* (1998).

TDTs for quantitative traits when only one parent is available

Sun *et al.* (1999) developed two test statistics for qualitative traits applicable to families with one parent. In this paper, we extend those test statistics to quantitative traits. As in Sun *et al.* (1999), we assume that the probability of a missing parent is the same for all genotypes at the marker locus of interest. The first class of tests is applicable when either of the following two assumptions, A1 or A2, holds. The second class of tests is applicable even if both assumption A1 and assumption A2 are violated. However, the second class of tests is generally less powerful than the first class of tests.

Assumption A1: Males and females with the same genotype at the marker locus have the same mating preference.

Assumption A2: The father and mother in each nuclear family are missing with the same probability $\frac{1}{2}$ given one of them is missing.

The 1-TDT when assumption A1 or assumption A2 holds

As in the above section, we condition on the trait values of the offspring. Under the null hypothesis of no linkage between the marker locus and the quantitative trait loci, $Y_{j,k}^{(1)}$ are independent of the trait values of the offspring. In Sun *et al.* (1999), we showed that $P(Y_{j,k}^{(1)} = 1) = P(Y_{j,k}^{(1)} = -1)$ under the null hypothesis of no linkage if Assumption A1 or A2 holds. Thus, for any constant c , the conditional mean of

$$s_1(c) = \sum_{j=1}^{n_1} \sum_{k=1}^{l_1(j)} (Q_{j,k}^{(1)} - c) Y_{j,k}^{(1)},$$

is zero. Unlike the situation when both parents are available, here $(Y_{j,1}^{(1)}, Y_{j,2}^{(1)}, \dots, Y_{j,l_1(j)}^{(1)})$ are not independent. To estimate the variance of $s_1(c)$, we note that the conditional mean of $\sum_{k=1}^{l_1(j)} (Q_{j,k}^{(1)} - c) Y_{j,k}^{(1)}$ is zero. Thus, its conditional variance can be estimated by $(\sum_{k=1}^{l_1(j)} (Q_{j,k}^{(1)} - c) Y_{j,k}^{(1)})^2$. The conditional variance of $s_1(c)$ can be estimated by $\sigma_1^2(c) = \sum_{j=1}^{n_1} (\sum_{k=1}^{l_1(j)} (Q_{j,k}^{(1)} - c) Y_{j,k}^{(1)})^2$. The class of statistical tests for the null hypothesis is given by

$$T_1(c) = s_1(c) / \sigma_1(c).$$

For qualitative traits, $T_1(c)$ is the same as the first test statistic in Sun *et al.* (1999) for any $c \neq 1$. We can also let c be the mean trait value among the general population or the sample mean trait value of all the offspring.

The 1-TDT when both assumption A1 and assumption A2 are violated

When both assumption A1 and assumption A2 are violated, $P(Y_{j,k}^{(1)} = 1) = P(Y_{j,k}^{(1)} = -1)$ no longer holds in general even under the null hypothesis of no linkage between the marker locus and the quantitative trait loci. $T_1(c)$ is no longer a valid test statistic for the null hypothesis. As in Sun *et al.* (1999), we modify $T_1(c)$ as follows. First, we calculate $s_1(c)$ and $\sigma_1^2(c)$ using families with fathers available. The corresponding values are denoted by $s_{1f}(c)$ and $\sigma_{1f}^2(c)$, respectively. Similarly, we calculate $s_1(c)$ and $\sigma_1^2(c)$ using families with mothers available and the corresponding values are denoted by $s_{1m}(c)$ and $\sigma_{1m}^2(c)$, respectively. Let N_f and N_m be the numbers of families with fathers and mothers available, respectively. Define

$$s_1^*(c) = N_m \times s_{1f}(c) + N_f \times s_{1m}(c), \quad \sigma_1^{*2}(c) = N_m^2 \times \sigma_{1f}^2(c) + N_f^2 \times \sigma_{1m}^2(c).$$

The new class of test statistics is then given by

$$T_1^*(c) = s_1^*(c) / \sigma_1^*(c).$$

The weight N_m for $s_{1f}(c)$ and the weight N_f for $s_{1m}(c)$ in the definition of $s_1^*(c)$ correct the potential biases of $s_1(c)$ due to the different missing probabilities for the father and the mother.

$T_1(c)$ as tests of linkage

Next we show that the expectation of $s_1(c)$ is zero if there is no linkage or no association between the marker locus and the quantitative trait locus. We prove this claim under the hypothesis of random mating and the Hardy–Weinberg equilibrium.

Let p and $1-p$ be the allele frequency of N and M, respectively. Let D and d , the two alleles at the trait locus, have frequencies q and $1-q$, respectively. θ is the recombination fraction, and $\Delta = P(ND) - pq$ is the linkage disequilibrium between the trait locus and the marker locus. Assume that the mean trait values are μ_0 , μ_1 , and μ_2 , for individuals with genotypes dd , Dd , and DD at the trait locus, respectively. Let G_t be the offspring's genotype at the trait locus; Q , the offspring's trait value; and Y , the corresponding index function defined above. We assume randomly sampled offspring-parent pairs. For simplicity, we do not use subscripts here. With the above notation, we have

$$E((Q-c)Y) = (\mu_2 - c)(P(DD, Y = 1) - P(DD, Y = -1)) + (\mu_1 - c)(P(Dd, Y = 1) - P(Dd, Y = -1)) + (\mu_0 - c)(P(dd, Y = 1) - P(dd, Y = -1)).$$

For any genotype G_t at the trait locus, we can calculate $P(G_t, Y = \pm 1)$ using Tables 2, 3, and 4 in George *et al.* (1999) (see Appendix). Then, we can show that

$$E((Q-c)Y) = -(2\theta - 1)\Delta(q(\mu_2 - \mu_1) + (1-q)(\mu_1 - \mu_0)).$$

Thus, $E((Q-c)Y) = 0$ when $(2\theta - 1)\Delta = 0$. $T_1(c)$ can be used as a test for linkage between the marker locus and the quantitative trait locus.

Combining data from different families

In practical studies, we might have families with both parents available, one parent available, or no parents available. Similar to Sun *et al.* (1999), we classify the families as follows:

1. Genotypes available for both parents.
2. Genotypes available for only one parent and one offspring.
3. Genotypes available for multiple offspring and at most one parent.

The first group of families can be analyzed by $T_2(c)$ and the second group of families can be analyzed by $T_1(c)$ or $T_1^*(c)$ depending on whether assumption A1 or assumption A2 holds or neither assumption holds.

The families with multiple offspring can be analyzed with the TDT for sibs. Allison *et al.* (1999) proposed two test statistics for detecting linkage between the marker locus and the trait locus: the mixed-effects model based test and the permutation test. Table 1 of Allison *et al.* (1999) showed that the permutation test is generally more powerful than the mixed-effects model based test although exceptions exist. For the permutation test, the weighted trait values (weighted by the number of M alleles), w_0 , in a sibship is compared with the expected weighted trait value, A_0 , by permutation conditional on the genotypes and the trait values of the offspring. Let $s_0 = \Sigma(w_0 - A_0)$ with summation across all the sibships. The variance, σ_0^2 , of s_0 was given in Allison *et al.* (1999). The same statistic can also be derived by conditioning on the trait values of the sibs and by permuting their genotypes. The latter conceptualization is consistent with our derivation of $T_2(c)$ and $T_1(c)$.

As in Sun *et al.* (1999), we propose the following statistic to combine the three types of families when either assumption A1 or assumption A2 above holds:

$$z(c) = \frac{s_2(c) + s_1(c) + s_0}{\sqrt{\sigma_2^2(c) + \sigma_1^2(c) + \sigma_0^2}}$$

z has an approximate standard normal distribution under no linkage when the number of families is large. When the number of families is small, simulations can be used to obtain an empirical p-value.

If neither assumption A1 nor assumption A2 holds, we should replace $s_1(c)$ and $\sigma_1^2(c)$ by $s_1^*(c)$ and $\sigma_1^{*2}(c)$ respectively. Note that, for qualitative traits, $z(1/2)$ is the same as the combined test of Sun *et al.* (1999).

SIMULATIONS

In this section, we study the validity and the power of the various TDTs using simulations. As in Sun *et al.* (1999), we considered three demographic models: population stratification, assortative mating, and differential gender mating preferences. We assumed that the population under study consisted of two sub-populations. A family belongs to the first population with probability 0.8 and belongs to the second population with probability 0.2. The haplotype frequencies of Nd, ND, Md, and MD were 0.2, 0.0, 0.0, and 0.8 in the first population, and 0.8, 0.0, 0.0, and 0.2 in the second population. For the assortative mating model, we assumed that 80% of the families were formed through random mating and 20% of the families were formed through assortative mating where the father's trait value and the mother's trait value differed by at most 5. For the differential gender mating preference model, we assumed that males with a trait value less than 0 mated randomly with females, while males with a trait value greater than 0 could only mate with females with a trait value greater than 0. To see the effects of family structure in the two populations on the statistics, we assumed that there are four offspring in each family in the first population and two offspring in each family in the second population. The recombination fraction between the marker locus and the trait locus was $\theta = 0, 0.1-0.5$ in steps of 0.1.

We assumed an additive model at the trait locus in both populations. Let

$$\mu_{ad}(i) = -a(i), \quad \mu_{aD}(i) = 0, \quad \mu_{DD}(i) = a(i), \quad i = 1, 2,$$

in the i -th population. The additive variance is $\sigma_a^2(i) = 2p_D(i)p_a(i)a^2(i)$ in the i -th population. Let $\sigma_e^2(i)$ be the residual trait variance. In our simulations, we assumed no residual sibling correlation ($\sigma_s^2(i) = 0$). For convenience, the total trait variances in both populations ($\sigma^2(i) = \sigma_a^2(i) + \sigma_e^2(i)$) were fixed at 100. The additive variance explains 30% of the total variance ($\sigma_a^2(1) = 30$) in the first population and 20% in the second population ($\sigma_a^2(2) = 20$).

Two hundred families were sampled for analysis. We considered two sampling schemes: random sampling and extremal sampling. For random sampling, the 200 families were assumed to be a random sample from the general population. For extremal sampling, we assumed that 200 families with at least one offspring having trait value greater than 0.0 were sampled for analysis. This sampling scheme is related to the threshold model in quantitative genetics in which individuals with a trait value above a certain threshold are considered to be affected. The type I error rate was set at 0.05. In each situation, 10,000 repetitions were used to approximate the type I error rate under the null hypothesis of no linkage and the power under the alternative hypothesis of linkage.

Let AQ be the population mean of the trait values and \bar{Q} be the sample mean of the trait values of all the offspring under study. We analyzed the data using $T_{21} = T_2(AQ)$, $T_{22} = T_2(\bar{Q})$ proposed by Rabinowitz (1997), and T_{23} : the test proposed by Xiong *et al.* (1998). We also analyzed the data using the permutation based TDT (T_o) for sibs of Allison *et al.* (1999) and $T_{11} = T_1(AQ)$, $T'_{11} = T'_1(AQ)$, $T_{12} = T_1(\bar{Q})$, $T'_{12} = T'_1(\bar{Q})$ given above. When we analyzed the data with TDTs with one parent, we assumed that the father was missing with probability 0.8 and the mother was missing with probability 0.2 given one of them was missing. The power of each test statistic was the fraction of times the absolute value of the test statistic was at least 1.96. Tables 1 and 2 show the power and the type I error of the different tests under random sampling and extremal sampling, respectively.

Table 1. Simulated power and type I error estimates (with 10 000 simulations) of the tests under random sampling for three different demographic models: A. Population stratification, B. Assortative mating, and C. Differential gender mating preferences. The number of families under study is 200

θ	Different tests*							
	T_{21}	T_{22}	T_{23}	T_0	T_{11}	T'_{11}	T_{12}	T'_{12}
A. Population stratification								
0.5	4.92	5.06	5.00	4.78	5.05	4.94	5.09	4.78
0.4	51.4	49.9	52.2	43.5	27.4	19.0	25.9	18.1
0.3	97.9	97.5	98.2	95.3	77.5	59.5	74.8	57.1
0.2	100	100	100	100	98.0	89.3	97.1	87.1
0.1	100	100	100	100	100	98.6	99.9	98.0
0.0	100	100	100	100	100	99.9	100	99.7
B. Assortative mating								
0.5	4.71	4.61	4.72	4.75	4.96	4.65	4.75	4.61
0.4	51.1	49.8	51.9	43.4	28.2	19.6	27.0	18.6
0.3	97.7	97.2	98.1	94.7	77.9	59.6	75.4	56.7
0.2	100	100	100	100	98.3	90.3	97.7	88.2
0.1	100	100	100	100	100	98.9	99.9	98.4
0.0	100	100	100	100	100	99.9	100	99.7
C. Differential gender mating preference								
0.5	4.98	4.92	4.95	4.42	5.34	4.71	6.99	4.51
0.4	48.2	45.9	48.8	40.8		20.0		19.2
0.3	96.6	95.7	96.8	92.9		56.0		52.8
0.2	100	100	100	100		87.3		84.0
0.1	100	100	100	100		98.1		96.7
0.0	100	100	100	100		99.8		99.5

* AQ and Q are the population and the sample mean trait values, respectively. $T_{21} = T_2(AQ)$, $T_{22} = T_2(\bar{Q})$, T_{23} : the test proposed by Xiong *et al.* (1998), T_0 : the permutation based TDT for sibs of Allison *et al.* (1999), $T_{11} = T_1(AQ)$, $T'_{11} = T'_1(AQ)$, $T_{12} = T_1(\bar{Q})$, $T'_{12} = T'_1(\bar{Q})$.

Table 2. Simulated power and type I error estimates (with 10 000 simulations) of the tests under extremal sampling for three different demographic models: A. Population stratification, B. Assortative mating, and C. Differential gender mating preferences. The number of families under study is 200

θ	Different tests*							
	T_{21}	T_{22}	T_{23}	T_0	T_{11}	T'_{11}	T_{12}	T'_{12}
A. Population stratification								
0.5	4.96	4.80	4.85	4.98	5.16	4.47	4.63	4.66
0.4	33.4	15.0	19.5	13.7	14.9	11.0	9.50	7.77
0.3	87.6	43.1	57.8	41.3	46.4	32.5	22.6	16.3
0.2	99.7	77.4	91.6	77.1	81.0	62.6	44.9	31.7
0.1	100	95.3	99.7	96.4	97.5	87.7	69.0	50.8
0.0	100	99.6	100	99.9	99.9	98.0	86.1	67.8
B. Assortative mating								
0.5	4.82	4.97	5.15	4.95	4.92	4.83	4.88	4.93
0.4	33.9	13.6	18.4	13.0	15.7	11.5	9.04	7.17
0.3	86.9	40.6	56.1	39.9	46.1	33.0	22.0	16.2
0.2	99.7	73.8	90.9	74.7	81.7	62.6	43.0	30.2
0.1	100	93.5	99.5	95.7	97.1	87.1	66.8	48.3
0.0	100	99.4	100	99.8	99.9	98.3	85.4	67.7
C. Differential gender mating preference								
0.5	5.08	4.90	5.00	4.97	11.9	4.60	6.14	4.63
0.4	30.6	12.9	17.8	12.9		9.76		7.51
0.3	83.1	36.3	53.4	37.5		25.7		15.2
0.2	99.5	68.7	88.3	71.7		56.6		27.1
0.1	100	91.1	99.3	93.9		82.8		43.5
0.0	100	98.6	100	99.8		96.3		60.8

* AQ and Q are the population and the sample mean trait values, respectively. $T_{21} = T_2(AQ)$, $T_{22} = T_2(\bar{Q})$, T_{23} : the test proposed by Xiong *et al.* (1998), T_0 : the permutation based TDT for sibs of Allison *et al.* (1999), $T_{11} = T_1(AQ)$, $T'_{11} = T'_1(AQ)$, $T_{12} = T_1(\bar{Q})$, $T'_{12} = T'_1(\bar{Q})$.

The type I error rates: In all the simulations, the simulated type I error rates of T_{21} , T_{22} , T_{23} , T_0 , T'_{11} and T'_{12} were close to the pre-specified type I error rate of 5%. The simulated type I error rates of T_{11} and T_{12} were close to 5% under the population stratification model and the assortative mating model. In both the population stratification model and the assortative mating model, males and females with the same genotype at the marker locus had the same mating preference. Under the differential gender mating preference model, the simulated type I error rates of T_{11} and T_{12} were 5.34% and 6.99%, respectively, for random sampling. For the extremal sampling considered here, the type I error rate of T_{11} and T_{12} were 11.9% and 6.14%, respectively for these two statistics. Thus, T_{11} and T_{12} are not valid when neither assumption A1 nor assumption A2 holds.

Comparing the power of the different tests: Under random sampling, T_{21} , T_{22} , and T_{23} have roughly the same power with T_{22} slightly less powerful than the other two tests. $T_{11}(T'_{11})$ is slightly more powerful than $T_{12}(T'_{12})$, respectively. Under extremal sampling, T_{21} can be much more powerful than T_{22} and T_{23} . Similarly, $T_{11}(T'_{11})$ can be much more powerful than $T_{12}(T'_{12})$, respectively. But this conclusion is not general. There exist situations where T_{23} and T_{22} are more powerful than T_{21} . Similarly, there exist situations where $T_{11}(T'_{11})$ is less powerful than $T_{12}(T'_{12})$, respectively.

DISCUSSION

In this paper, we first proposed a new class of TDTs to test for linkage for quantitative traits applicable to families with both parents. The standard TDT for qualitative traits of Spielman *et al.* (1993) is a special case of this class of tests. If the population mean, AQ , of the trait values in the general population is known, $T_2(AQ)$ has roughly the same power as the test proposed by Rabinowitz (1997) and the test proposed by Xiong *et al.* (1998) under random sampling. For non-random sampling, $T_2(AQ)$ can be much more powerful than the other available tests. We also proposed TDT type tests for quantitative traits applicable to families with one parent. We showed theoretically and by simulations that the proposed tests are valid for linkage between the marker locus and the quantitative trait locus. Finally, we proposed a method to combine data from families with both parents available, one parent available and no parent available but with multiple offspring. The tests proposed in this paper unify the TDT for quantitative and qualitative traits.

George *et al.* (1999) proposed maximum likelihood methods for testing linkage for quantitative traits based on multiple regression. In order for their method to be valid, the individuals under study need to be a random sample from the general population. It is also assumed that the trait values have a normal distribution. When these conditions are satisfied, their methods presumably will be more powerful than the methods proposed in this paper. However, their methods are not applicable to selectively sampled samples. The statistics proposed in this paper are applicable to any kind of samples. Neither do they need any assumptions about the distribution of the quantitative traits.

In the above sections, we assumed that there are two alleles at the marker locus of interest. As in TDT for qualitative traits, this assumption is natural if there is a particular allele of interest and we can group the other alleles into one group. Sometimes we do not have a particular allele of interest. Several investigators (Allison, 1997; Allison *et al.* 1999; Rabinowitz, 1997; Xiong, 1998) proposed chi-square tests by jointly considering multiple alleles at the marker locus. It is not obvious how to combine data from different types of families using this approach. Instead we propose the following method as in Spielman & Ewens (1998) and Sun *et al.* (1999) for qualitative traits. Suppose that there are k alleles at the marker locus. For each of the k alleles, we can group the other alleles into one group and calculate the z-score for that particular allele. The final score is the maximum, z_{\max} , of the absolute z-scores over all the alleles. Since z_{\max} is no longer normally distributed, Bonferroni

correction can be used to approximate the p -value (Spielman & Ewens 1998). When multiple loci are considered, we can study each locus separately and obtain the p -value using Bonferroni correction or treat each haplotype as an allele the same as for the combined TDT for qualitative traits.

When one suspects that environmental risk factors are related to the quantitative trait of interest, one could replace c in $T_1(c)$ or $T_2(c)$ by the fitted values, $\hat{X}_{i,j}$, of $X_{i,j}$ obtained by regressing $X_{i,j}$ on the available covariates as proposed in Rabinowitz (1997). Similarly, for the TDT for sibs, one could replace $X_{i,j}$ by its residual $X_{i,j} - \hat{X}_{i,j}$. Then the data from different types of families could be similarly combined. The validity and the increase in power of this approach need to be further studied and is an area of future research.

This research is partly supported by NIH DK53392 (FZ Sun), GM59507, and HD36834 (HY Zhao).

APPENDIX

In this appendix, we calculate $P(G_t, Y = \pm 1)$ for any given genotype, G_t , at the quantitative trait locus. Four offspring-parents-trio genotypes: (NM, NN \times NM), (NM, NN \times MM), (MM, NM \times NM), and (MM, NM \times MM), result in $Y = 1$. Similarly, four other offspring-parents-trio genotypes: (NN, NM \times NN), (NN, NM \times NM), (NM, MM \times NN), and (NM, MM \times NM), result in $Y = -1$. In the above notation, the first genotype is the offspring's genotype at the marker locus; the second genotype is the available parent's genotype and the third genotype is the unavailable parent's genotype. Thus, for any offspring's genotype, G_t , at the trait locus, we have

$$\begin{aligned} P(G_t, Y = 1) &= P(G_t, NM, NN \times NM) + P(G_t, NM, NN \times MM) \\ &\quad + P(G_t, MM, NM \times NM) + P(G_t, MM, NM \times MM). \end{aligned} \quad (1)$$

Similarly,

$$\begin{aligned} P(G_t, Y = -1) &= P(G_t, NN, NM \times NN) + P(G_t, NN, NM \times NM) \\ &\quad + P(G_t, NM, MM \times NN) + P(G_t, NM, MM \times NM). \end{aligned} \quad (2)$$

Note

$$P(G_t, NM, NN \times MM) = P(NN) \times P(MM) \times P(G_t) = p^2(1-p)^2 P(G_t).$$

The terms in which at least one parent is heterozygous can be calculated using Tables 2, 3 and 4 in George *et al.* (1999). Here we just summarize the results.

$$P(DD, NN, NN \times NM) = h_1 p[-\theta\Delta + h_1(1-p)],$$

$$P(Dd, NN, NN \times NM) = p[(h_1 - h_2)\theta\Delta + 2h_1 h_2(1-p)],$$

$$P(dd, NN, NN \times NM) = h_2 p[\theta\Delta + h_2(1-p)].$$

$$P(DD, NM, NN \times NM) = h_1 p[\theta\Delta + h_3 p],$$

$$P(Dd, NM, NN \times NM) = p[-(h_1 - h_2)\theta\Delta + p(h_1 h_4 + h_2 h_3)],$$

$$P(dd, NM, NN \times NM) = h_2 p[-\theta\Delta + h_4 p].$$

$$P(DD, NN, NM \times NM) = (h_2 h_3 + \theta\Delta)^2 + h_1 h_3[2\theta\Delta + h_3(h_1 + 2h_2)],$$

$$P(Dd, NN, NM \times NM) = 2\theta(1-\theta)\Delta^2 - 2\theta\Delta(h_1 h_3 - h_2 h_4) + 2h_3 h_4 p^2,$$

$$P(dd, NN, NM \times NM) = (h_1 h_4 - \theta\Delta)^2 + h_2 h_4[-2\theta\Delta + h_4(2h_1 + h_2)].$$

$$P(DD, MM, NM \times NM) = (h_1 h_4 - \theta\Delta)^2 + h_1 h_3[-2\theta\Delta + h_1(h_3 + 2h_4)],$$

$$P(Dd, MM, NM \times NM) = 2\theta(1-\theta)\Delta^2 + 2\theta\Delta(h_1 h_3 - h_2 h_4) + 2h_1 h_2(1-p)^2,$$

$$P(dd, MM, NM \times NM) = (h_2 h_3 + \theta\Delta)^2 + h_2 h_4[2\theta\Delta + h_2(2h_3 + h_4)].$$

$$P(DD, MM, MM \times NM) = h_3(1-p)[\theta\Delta + h_3 p],$$

$$P(Dd, MM, MM \times NM) = (1-p)[(h_4 - h_3)\theta\Delta + 2h_3 h_4 p],$$

$$P(dd, MM, MM \times NM) = h_4(1-p)[- \theta\Delta + h_4 p].$$

$$P(DD, NM, MM \times NM) = h_3(1-p)[- \theta\Delta + h_1(1-p)],$$

$$P(Dd, NM, MM \times NM) = (1-p)[(h_3 - h_4)\theta\Delta + (1-p)(h_1 h_4 + h_2 h_3)],$$

$$P(dd, NM, MM \times NM) = h_4(1-p)[\theta\Delta + h_2(1-p)].$$

Substituting the above equations into Equations (1) and (2), we can obtain $P(G_t, Y = \pm 1)$ for any given genotype, G_t , at the quantitative trait locus.

REFERENCES

- Abecasis, G. R., Cardon, L. R., Cookson, O. C. (2000). A general test of association for quantitative traits in nuclear families. *Am. J. Hum. Genet.* **66**, 279–292.
- Allison, D. B. (1997). Transmission-disequilibrium tests for quantitative traits. *Am. J. Hum. Genet.* **60**, 676–690.
- Allison, D. B., Heo, M., Kaplan, N., Martin, E. R. (1999). Sibling-based tests of linkage and association for quantitative traits. *Am. J. Hum. Genet.* **64**, 1754–1764.
- Boehnke, M., Langefeld, C. D. (1998). Genetic association mapping based on discordant sib pairs: the discordant-allele test. *Am. J. Hum. Genet.* **62**, 950–961.
- Curtis, D. R. (1997). Use of siblings as controls in case-control association studies. *Ann. Hum. Genet.* **61**, 319–333.
- Duffy, D. (1995). Screening a 2 cM genetic map for allelic association: a simulated oligogenic trait. *Genet. Epidemiol.* **12**, 595–600.
- Ewens, W. J., Spielman, R. S. (1995). The transmission/disequilibrium test: history, subdivision and admixture. *Am. J. Hum. Genet.* **57**, 455–464.
- George, V., Tiwari, H. K., Zhu, X., Elston, R. C. (1999). A test of transmission/disequilibrium for quantitative traits in pedigree data by multiple regression. *Am. J. Hum. Genet.* **65**, 236–245.
- Horvath, S., 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. (1999). The transmission/disequilibrium test and parental-genotype reconstruction: the reconstruction-combined transmission/disequilibrium test. *Am. J. Hum. Genet.* **64**, 861–870.
- Lazzeroni, L. C., Lange, K. (1998). A conditional inference framework for extending the transmission/disequilibrium test. *Hum. Hered.* **48**, 67–81.
- Lunetta, K. L., Faraone, S. V., Biederman, J., Laird, N. M. (2000). Family-based tests of association and linkage that use unaffected sibs, covariates, and interactions. *Am. J. Hum. Genet.* **66**, 605–614.
- Monks, S. A., Kaplan, N. L. (2000). Removing the sampling restrictions from family-based tests of association for a quantitative trait locus. *Am. J. Hum. Genet.* **66**, 576–592.
- Rabinowitz, D. (1997). A transmission disequilibrium test for quantitative trait loci. *Hum. Hered.* **47**, 342–350.
- Risch, N., Merikangas, K. (1996). The future of genetic studies of complex human diseases. *Science* **273**, 1516–1517.
- Schaid, D. J. (1996). General score tests for associations of genetic markers with disease using cases and their parents. *Genet. Epidemiol.* **13**, 423–450.
- Sham, P. C., Curtis, D. R. (1995). An extended transmission/disequilibrium test (TDT) for multi-allelic marker loci. *Ann. Hum. Genet.* **59**, 323–336.
- Spielman, R. S., McGinnis, R. E., Ewens, W. J. (1993). Transmission test for linkage disequilibrium: the insulin gene region and insulin-dependent diabetes mellitus. *Am. J. Hum. Genet.* **52**, 506–516.
- Spielman, R. S., Ewens, W. J. (1996). The TDT and other family-based tests for linkage disequilibrium and association. *Am. J. Hum. Genet.* **59**, 983–989.
- Spielman, R. S., 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., 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., Khoury, M. J. (1999). Transmission disequilibrium test (TDT) when only one parent is available: the 1-TDT. *Am. J. Epidemiol.* **150**, 97–104.
- Teng, J., Risch, N. (1999). The relative power of family-based and case-control designs for linkage disequilibrium studies of complex human diseases II. Individual genotyping. *Genome Res.* **8**, 1273–1288.
- Weinberg, C. R. (1999). Allowing for missing parents in genetic studies of case-parental triads. *Am. J. Hum. Genet.* **64**, 1186–1193.
- Xiong, M. M., Krushkal, J., Boerwinkle, E. (1998). TDT statistics for mapping quantitative trait loci. *Ann. Hum. Genet.* **65**, 431–452.