Am. J. Hum. Genet. 65:236–245, 1999

A Test of Transmission/Disequilibrium for Quantitative Traits in Pedigree Data, by Multiple Regression

Varghese George,[1] Hemant K. Tiwari,[2] Xiaofeng Zhu,[3] and Robert C. Elston[2]

[1]Division of Biostatistics, Medical College of Wisconsin, Milwaukee; [2]Department of Epidemiology and Biostatistics, Case Western Reserve University, Cleveland; and [3]Department of Preventive Medicine and Epidemiology, Loyola University Medical Center, Chicago

Summary

The transmission/disequilibrium (TD) test (TDT), proposed, by Spielman et al., for binary traits is a powerful method for detection of linkage between a marker locus and a disease locus, in the presence of allelic association. As a test for linkage disequilibrium, the TDT makes the assumption that any allelic association present is due to linkage. Allison proposed a series of TD-type tests for quantitative traits and calculated their power, assuming that the marker locus is the disease locus. All these tests assume that the observations are independent, and therefore they are applicable, as a test for linkage, only for nuclear-family data. In this report, we propose a regression-based TD-type test for linkage between a marker locus and a quantitative trait locus, using information on the parent-to-offspring transmission status of the associated allele at the marker locus. This method does not require independence of observations, thus allowing for analysis of pedigree data as well, and allows adjustment for covariates. We investigate the statistical power and validity of the test by simulating markers at various recombination fractions from the disease locus.

Introduction

The transmission/disequilibrium (TD) test (TDT) of Spielman et al. (1993) is a viable alternative to other existing sampling designs for testing for linkage between marker loci and associated dichotomous disease traits when marker genotypes are known for both the parents and affected offspring from independent nuclear families. As a test for linkage, transmissions from heterozygous parents to their affected children are used in the analysis. The test compares the frequency of marker alleles transmitted to affected children versus the frequency of marker alleles not transmitted, using a χ^2 statistic. Thus, for the purpose of testing for linkage, the TDT has the advantage of not requiring data on either multiple affected family members or unaffected sibs (although transmission to unaffected sibs should also studied if there is any doubt about segregation distortion, or meiotic drive, at the marker locus). The TDT can also be used directly as a test for association in the presence of linkage if the sample consists of nuclear families with a single affected offspring, as has been suggested by Spielman and Ewens (1996). However, if families in the sample have several affected sibs, then it is not a valid test of association, because of the sibs' lack of independence. Martin et al. (1997) have proposed a method for testing the null hypothesis of no association and no linkage against the alternative hypothesis of association in the presence of linkage, using all affected siblings from independent nuclear families. They considered the set of transmissions to affected sibs in the whole family, rather than the transmissions to each child separately, thus retaining the necessary independence property. This method is more powerful than the TDT applied to a single affected child in each nuclear family. Cleves et al. (1997) have proposed the use of an exact TDT for multiallelic markers, using both an exact algorithm and Markov-chain Monte Carlo simulation. Their simulation studies showed not only that exact tests improve the validity and power of the TDT but also that the power further increases if the nuclear families each consist of two affected sibs and if only those parents who transmit the same allele to both sibs are included in the sample for analysis.

Allison (1997) has proposed various TD-type tests for use with quantitative traits measured in members of independent nuclear families. These tests accommodate either selected sampling or sampling based on selection of extreme phenotypes among the offspring. All these tests assume that the observations are independent (for the purpose of testing for linkage). Allison has shown that, when a candidate gene in disequilibrium with the marker

Received November 10, 1998; accepted for publication May 6, 1999; electronically published June 3, 1999.

Address for correspondence and reprints: Dr. Varghese George, Division of Biostatistics, Medical College of Wisconsin, 8701 Watertown Plank Road, Milwaukee, WI 53226. E-mail: george@mcw.edu

© 1999 by The American Society of Human Genetics. All rights reserved.
0002-9297/99/6501-0030$02.00

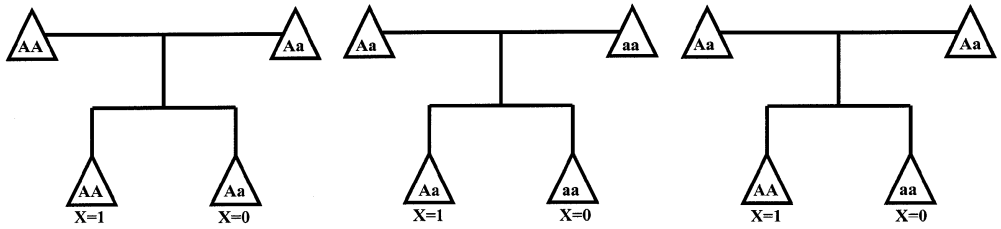

Figure 1 Offspring who are informative for linkage, from relevant parental matings. "A" is the associated allele of interest, "X=1" corresponds to A transmitted from a heterozygous parent, and "X=0" corresponds to A not transmitted from a heterozygous parent.

is available, these tests are more powerful than two common sib-pair linkage tests proposed by Haseman and Elston (1972) and Risch and Zhang (1996). He further has noted that, in the presence of disequilibrium, the sample size needed to map a gene when these tests are used could be orders of magnitude smaller than the sample size needed when the sib-pair methods that assume equilibrium are used.

A disadvantage of all TDT methods is that they can detect linkage between the marker locus and the disease trait only if there is association between the disease locus and alleles at the linked locus. As has been pointed out by Allison (1997), these methods typically require that one has a marker that is extremely close to the trait locus, in order to have adequate power. As a test for linkage disequilibrium, the TDT makes the assumption that any allelic association present is due to linkage disequilibrium. The tests cannot distinguish between linkage disequilibrium and, in the presence of linkage, allelic association due to other causes (such as population admixture or selection). Thus, the interpretation of these tests, other than as a test for linkage in the presence of allelic association, is questionable.

In this paper we propose a TDT for assessment of the linkage between a marker locus and a quantitative trait, by regressing the trait on the parental transmission of an allele of interest, assuming that the alleles at the marker locus are associated with those at the disease locus. As in the original TDT, it is not necessary to assume that this association is due to linkage disequilibrium; the allelic association could be due to other factors, such as selection, population admixture, or stratification. The model is a simple linear-regression model that can simultaneously assess the effects that other relevant covariates have on the trait. Linkage is assessed by evaluation of the effect of the transmission or nontransmission of the associated allele from at least one of the parents. An advantage of this method is that there is no restriction on either the family structure sampled (except for the limitation of the method that is used to compute the likelihood functions for the analysis) or the affection status of individuals in the pedigree. We use simulation to investigate the power of the proposed method for markers at various distances from the trait locus.

Methods

The method that we propose here is a linear-regression approach with the disease trait, which is assumed to be continuous, as the dependent variable (Y). The primary independent variable in the model is, for each individual, the transmission status (X) of the associated allele at a given locus. As in other TD-based tests, only offspring who are informative for linkage will be used in the analysis. This includes all offspring of a homozygous × heterozygous mating and all homozygous offspring of a heterozygous × heterozygous mating. The heterozygous offspring of a heterozygous × heterozygous mating have each allele transmitted from one parent but not transmitted from the other parent and, therefore, are considered to be noninformative. If we let A be the associated allele and assume that the marker locus is diallelic (if there are more than two alleles at the locus, they can be treated as A and not-A), the transmission-status variable X takes on a value of 1 if the offspring possesses allele A transmitted from a heterozygous parent and takes on a value of 0 otherwise. Informative offspring are shown in figure 1, where not-A is denoted as "a."

If we let p denote the population frequency of the associated marker allele, under the assumption of random mating of the parents, then the probability that an offspring is informative for linkage is given by

$$4p^3(1-p) + 4p(1-p)^3 + 2p^2(1-p)^2 = 2p(1-p)(2-3p+3p^2) \ . \qquad (1)$$

In addition to X, we can incorporate the additional covariates $C_1, \ldots, C_k$ into the model. The coefficient of X measures linkage by assessing the association between the transmission status of allele A from heterozygous

Table 1

Empirical Power and Type I Error of Proposed Method, Based on Simulation of 2,000 Replicates

No. of Families	No. of Offspring	Δ	θ	5% Significance	1% Significance	0.5% Significance
200	4	.10				
				Empirical Power		
			.00	89.7	73.1	65.9
			.01	88.4	72.2	64.1
			.05	82.7	63.3	54.7
				Type I Error (%)		
			.50	6.4	1.2	.7
200	4	.01				
			θ	Empirical Power		
			.00	7.6	2.1	1.2
			.01	7.9	1.6	.9
			.05	8.0	2.5	1.5
				Type I Error (%)		
			.50	5.2	.7	.4
200	2	.10				
			θ	Empirical Power		
			.00	52.3	29.5	21.9
			.01	50.9	27.8	20.3
			.05	46.1	22.8	17.4
				Type I Error (%)		
			.50	5.8	1.6	.9
200	2	.01				
			θ	Empirical Power		
			.00	6.2	1.6	1.0
			.01	5.8	1.0	.6
			.05	6.5	1.3	.8
				Type I Error (%)		
			.50	6.3	1.3	.6
100	4	.10				
			θ	Empirical Power		
			.00	64.3	41.4	32.5
			.01	62.2	40.1	31.9
			.05	55.7	33.2	26.1
				Type I Error (%)		
			.50	5.1	1.3	.7
100	4	.01				
			θ	Empirical Power		
			.00	7.5	1.5	1.0
			.01	7.4	1.8	1.2
			.05	6.8	1.5	.8

(*continued*)

Table 1 (continued)

No. of Families	No. of Offspring	Δ	θ	5% Significance	1% Significance	0.5% Significance
100	4	.01				
				Type I Error (%)		
			.50	5.8	1.2	.6
100	2	.10				
			θ	Empirical Power		
			.00	30.3	14.2	10.3
			.01	29.6	11.3	7.6
			.05	24.3	10.3	7.2
				Type I Error (%)		
			.50	6.2	1.1	.6
100	2	.01				
			θ	Empirical Power		
			.00	6.4	1.5	.9
			.01	5.6	1.4	.6
			.05	5.8	1.5	.8
				Type I Error (%)		
			.50	5.1	1.4	.7

parents and the disease trait, after adjusting for the covariates. Thus, the model for analysis is, for the ith member in a pedigree of n individuals,

$$Y_i = \alpha + \beta_1 C_{1i} + \ldots + \beta_k C_{ki} + \gamma X_i + E_i \ , \qquad (2)$$

where $(E_1,\ldots,E_n)^T$ is distributed as a multivariate normal with mean vector **0**. The familial correlations among pedigree members are incorporated by means of the association model proposed by George and Elston (1987), which assumes that the residual E_i is composed of two additive and uncorrelated random components: a familial effect, G_i, and an individual-specific residual effect, R_i. These effects are assumed to be normally distributed (after transformation, if necessary), with means 0 and variances σ^2_g and σ^2_r, respectively. As a first approximation, the familial effect G_i is assumed to lead to a correlation structure such as would be expected, under random mating, from polygenic inheritance. Thus the residual correlation between a pair of jth-degree relatives is taken to be of the form $f^2/2^j$, where

$$f^2 = \sigma^2_g/(\sigma^2_g + \sigma^2_r) \ . \qquad (3)$$

For a pedigree of n individuals, the joint likelihood function is given by

$$\mathbf{L} = \int \ldots \int \prod_i f_g(G_i) f_y(Y_i \mid G_i) dG_1 \ldots dG_n$$
$$= \int \ldots \int \prod_i f_g(G_i) \varphi(y_i - G_i - \mu_y, \sigma^2_r) dG_1 \ldots dG_n \ ,$$

where $\mu_y = \alpha + \beta_1 C_{1i} + \ldots + \beta_k C_{ki} + \gamma X_i$, and $\varphi(z,\sigma^2)$ is the density, at z, of a normal distribution with mean 0 and variance σ^2. All integrations are from $-\infty$ to ∞, and the product is over all n individuals in the pedigree. The function $f_g(G_i)$ takes on one of two forms, depending on whether the person has no parents in the pedigree or both parents in the pedigree. In the former case, it is the population distribution of G_i, which is $\varphi(G_i,\sigma^2_g)$; in the second case, it is the distribution of the person's polygenic effect, conditional on those of his or her parents (G_f and G_m), and is given by $\varphi[G_i - (G_f + G_m)/2, \sigma^2_g/2]$. The joint likelihood of the sample of independent pedigrees will be the product of the likelihoods of the individual pedigrees.

The Elston-Stewart algorithm (Elston and Stewart 1971) can be used to compute the likelihood presented above. The maximum-likelihood estimates of the parameters, as well as the standard errors (SEs) of the estimates, are computed by numerical methods. These procedures are implemented in the program ASSOC of the

Table 2

Parental Two-Locus Mating–Type Probabilities and Joint Probabilities of Offspring's Trait and Marker Genotypes, Conditional on Parental Two-Locus Mating Type, When Marker Mating Type Is AA × Aa

Parental Two-Locus Mating Type	Probability of Mating Type	Joint Probability of Offspring's Trait and Marker Genotypes								
		DD			Dd			dd		
		AA	Aa	Aa	AA	Aa	aa	AA	Aa	aa
AD/AD × AD/aD	$4\,h_1^3\,h_3$	1/2	1/2	0	0	0	0	0	0	0
AD/Ad × AD/aD	$8\,h_1^2\,h_2\,h_3$	1/4	1/4	0	1/4	1/4	0	0	0	0
AD/AD × AD/ad	$4\,h_1^3\,h_4$	$(1-\theta)/2$	$\theta/2$	0	$\theta/2$	$(1-\theta)/2$	0	0	0	0
AD/Ad × Ad/aD	$4\,h_1^2\,h_2\,h_3$	$\theta/2$	$(1-\theta)/2$	0	$(1-\theta)/2$	$\theta/2$	0	0	0	0
Ad/Ad × AD/aD	$4\,h_1\,h_2^2\,h_3$	0	0	0	1/2	1/2	0	0	0	0
AD/Ad × Ad/ad	$4\,h_1^2\,h_2\,h_4$	0	0	0	1/2	1/2	0	0	0	0
AD/Ad × AD/ad	$8\,h_1^2\,h_2\,h_4$	$(1-\theta)/4$	$\theta/4$	0	1/4	1/4	0	$\theta/4$	$(1-\theta)/4$	0
AD/Ad × Ad/aD	$8\,h_1\,h_2^2\,h_3$	$\theta/4$	$(1-\theta)/4$	0	1/4	1/4	0	$(1-\theta)/4$	$\theta/4$	0
Ad/Ad × AD/ad	$4\,h_1\,h_2^2\,h_4$	0	0	0	$(1-\theta)/2$	$\theta/2$	0	$\theta/2$	$(1-\theta)/2$	0
Ad/Ad × Ad/aD	$4\,h_2^3\,h_3$	0	0	0	$\theta/2$	$(1-\theta)/2$	0	$(1-\theta)/2$	$\theta/2$	0
AD/Ad × Ad/ad	$8\,h_1\,h_2^2\,h_4$	0	0	0	l/4	1/4	0	1/4	1/4	0
Ad/Ad × Ad/ad	$4\,h_2^3\,h_4$	0	0	0	0	0	0	1/2	1/2	0

SAGE (1998) software package. The correlation structure given by equation (3) can easily be extended to allow for more-complex models incorporating various types of common environmental effects (Elston et al. 1992). If one wishes to use only nuclear families instead of extended pedigrees, then SUDAAN (Shah et al. 1997) or the Mixed Model procedure in SAS (Littell et al. 1996) can be used to perform the analysis, if it is assumed that there is some form of reasonable correlation structure within families.

The regression coefficient γ in equation (2) can be expressed as a function of the recombination fraction (θ) and the disequilibrium parameter (Δ). Let A and a be the alleles at the marker locus, with $P(\mathrm{A}) = p$, and let D and d be the alleles at the trait locus, with $P(\mathrm{D}) = q$. Let the trait genotypic means be μ, ν, and $-\mu$, corresponding to genotypes DD, Dd, and dd, respectively. Then it can be shown (for derivation, see the Appendix) that

$$\gamma = 2(1-2\theta)\Delta\{(1-p+p^2)[\mu+(1-2q)\nu] + (1-2p)\Delta\nu\}[p(1-p)(2-3p+3p^2)]^{-1} \,. \quad (4)$$

Thus, under the assumption of population association (i.e., $\Delta \neq 0$), testing the null hypothesis that $\gamma = 0$ is equivalent to the test for no linkage (i.e., $\theta = .5$).

Denote the maximum-likelihood estimate of γ as $\hat{\gamma}$, and denote its SE as $\mathrm{SE}(\hat{\gamma})$. The maximum-likelihood estimates have an asymptotic normal distribution, provided that the log-likelihood function is not appreciably far from being quadratic (Lindsey 1996). Therefore, it is reasonable to conclude that $\hat{\gamma}$ is asymptotically normal, with mean γ. Thus, the test statistic for testing for linkage is $\mathrm{T} = \hat{\gamma}/\mathrm{SE}(\hat{\gamma})$, where T is asymptotically distributed as standard normal.

We can also perform a likelihood-ratio test for linkage, by comparing the maximized likelihood functions under the null and alternative hypotheses. Denote the log likelihoods under the null hypothesis of no linkage and under the alternate hypothesis of linkage as l_0 and l_1, respectively. Under the null hypothesis of no linkage, the likelihood-ratio statistic $2(l_1 - l_0)$ is asymptotically distributed as χ^2 with 1 df (Lindsey 1996).

Simulation Study of Power and Validity

To evaluate the power of the proposed method, we simulated sets of 2,000 replicate samples, each sample consisting of either 100 or 200 nuclear families, each of contained two parents and either two or four offspring, under the assumption of random mating of the parents. A quantitative trait was simulated with a polygenic effect, a random environmental effect, and a major-gene (the trait locus) effect for each individual. We assumed that these three effects are independent of each other. The parents' polygenic component was randomly assigned on the basis of a normal distribution with mean 0 and variance .25. The offspring's polygenic components were derived by adding to the average of the parents' polygenic effects a random deviation with mean 0 and variance .125. The random environmental effect was simulated on the basis of a normal distribution with mean 0 and variance .5. The major-gene effect at the trait locus (D) was additive, with $\mu = .7071$ and $\nu = 0$, with allele frequencies $P(\mathrm{D}) = P(\mathrm{d}) = .5$, resulting in a major-locus variance of .25. Thus the polygenic effect

Table 3

Parental Two-Locus Mating–Type Probabilities and Joint Probabilities of Offspring's Trait and Marker Genotypes, Conditional on Parental Two-Locus Mating Type, When Marker Mating Type Is AA × aa

Parental Two-Locus Mating Type	Probability of Mating Type	Joint Probability of Offspring's Trait and Marker Genotypes								
		DD			Dd			dd		
		AA	Aa	aa	AA	Aa	aa	AA	Aa	aa
AD/aD × aD/aD	$4\ h_1\ h_3^3$	0	1/2	1/2	0	0	0	0	0	0
AD/aD × aD/aD	$8\ h_1\ h_3^2\ h_4$	0	1/4	1/4	0	1/4	1/4	0	0	0
AD/ad × aD/aD	$4\ h_1\ h^2{}_3\ h_4$	0	$(1-\theta)/2$	$\theta/2$	0	$\theta/2$	$(1-\theta)/2$	0	0	0
Ad/aD × aD/aD	$4\ h_2\ h_3^3$	0	$\theta/2$	$(1-\theta)/2$	0	$(1-\theta)/2$	$\theta/2$	0	0	0
AD/aD × ad/ad	$4\ h_1\ h_3\ h_4^2$	0	0	0	0	1/2	1/2	0	0	0
Ad/ad × aD/aD	$4\ h_2\ h_3^2\ h_4$	0	0	0	0	1/2	1/2	0	0	0
AD/ad × aD/ad	$8\ h_1\ h_3\ h_4^2$	0	$(1-\theta)/4$	$\theta/4$	0	1/4	1/4	0	$\theta/4$	$(1-\theta)/4$
Ad/aD × aD/ad	$8\ h_2\ h_3^2\ h_4$	0	$\theta/4$	$(1-\theta)/4$	0	1/4	1/4	0	$(1-\theta)/4$	$\theta/4$
AD/ad × ad/ad	$4\ h_1\ h_4^3$	0	0	0	0	$(1-\theta)/2$	$\theta/2$	0	$0\theta/2$	$(1-\theta)/2$
Ad/aD × ad/ad	$4\ h_2\ h_3\ h_4^2$	0	0	0	0	$\theta/2$	$(1-\theta)/2$	0	$(1-\theta)/2$	$\theta/2$
Ad/ad × aD/ad	$8\ h_2\ h_3\ h_4^2$	0	0	0	0	1/4	1/4	0	1/4	1/4
Ad/ad × ad/ad	$4\ h_2\ h_4^3$	0	0	0	0	0	0	0	1/2	1/2

and the major gene each accounted for 25% of the total phenotypic variance, whereas the random environmental effect accounted for 50%. The marker allele A was simulated with $P(\mathrm{A}) = .4$ and $P(\mathrm{a}) = .6$. Linkage-disequilibrium coefficients, Δ's, of .1 and .01 were achieved by use of $P(\mathrm{Ad}) = .3$ and .21, respectively; $P(\mathrm{ad}) = .2$ and .29, respectively; $P(\mathrm{AD}) = .1$ and .19, respectively; and $P(\mathrm{aD}) = .4$ and .31, respectively. The simulation of the genotypes of the trait loci and the marker loci was repeated with $\theta = .0$, .01, .05, and .5, under the assumption that there was no crossover interference. No covariates were included in the simulation. We used the program ASSOC in SAGE (1998) to perform the analyses. Empirical power for linkage testing by test statistic T was computed at significance levels .005, .01, and .05. The results of the simulation are given intable 1. Under moderate disequilibrium (i.e., $\Delta = .1$), the power is very high when θ is close to 0, and it decreases as θ approaches .5. When the marker and the trait locus are weakly associated ($\Delta = .01$), the power is, as expected, extremely low (very close to the nominal significance level), regardless of the value of θ. It should be pointed out that the power depends on both the family structure and the number of informative offspring in the sample size. Because the families were generated under the assumption of random mating, the number of informative offspring varied among replicates in the simulation. In our simulation, the frequency of the associated marker allele, p, was taken to be .4 and, hence, by equation (1), the probability of an offspring being informative was 61%. For a fixed total number of offspring, families with larger sibship size should increase this probability, because of familial correlation and linkage disequilibrium. Our simulation study confirms these facts; the power corresponding to 100 families each containing four offspring is higher than that of 200 families each containing two offspring. Therefore, because the cost of genotyping is cheaper, per offspring, in larger families (there being, on average, fewer parents to type), it is cost-effective to take larger families into the study.

$\theta = .5$ corresponds to the null hypothesis of no linkage, and the corresponding power is the empirical type I error. The empirical type I errors for various significance levels are also given in table 1. These values are close to the specified levels, regardless of the value of Δ, thus supporting the validity of the method. Asymptotically, the T statistic and the likelihood ratio–test statistic are normally distributed, but one may question the properties that these tests have for finite samples. The simulation results show that our proposed method is valid and quite powerful when the sample size is reasonably large.

Discussion

The method that we propose here is more general than those that have been described by Allison (1997). Our method allows for arbitrary pedigree structure and nonindependence of observations. Even though the transmissions of marker alleles from parents to offspring are independent, the disease trait may still be dependent among family members, because of polygenic and common environmental effects. Therefore it is necessary to account for some reasonable form of correlation structure in the analysis. We assumed that the correlation structure among the pedigree members was as would be expected, under random mating, from polygenic inheritance and random residual error, but this is not a nec-

Table 4

Parental Two-Locus Mating–Type Probabilities and Joint Probabilities of Offspring's Trait and Marker Genotypes, Conditional on Parental Two-Locus Mating Type, When Marker Mating Type Is Aa × Aa

		Joint Probability of Offspring's Trait and Marker Genotypes								
		DD			Dd			dd		
Parental Two-Locus Mating Type	Probability of Mating Type	AA	Aa	aa	AA	Aa	aa	AA	Aa	aa
AD/aD × AD/aD	$4\ h_1^2\ h_3^2$	1/4	1/2	1/4	0	0	0	0	0	0
AD/aD × AD/ad	$8\ h_1^2\ h_3\ h_4$	$(1-\theta)/4$	1/4	$\theta/4$	$\theta/4$	1/4	$(1-\theta)/4$	0	0	0
AD/aD × Ad/aD	$8\ h_1\ h_2\ h_3^2$	$\theta/4$	1/4	$(1-\theta)/4$	$(1-\theta)/4$	1/4	$\theta/4$	0	0	0
AD/aD × Ad/ad	$8\ h_1\ h_2\ h_3\ h_4$	0	0	0	1/4	1/2	1/4	0	0	0
AD/ad × AD/ad	$4\ h_1^2\ h_4^2$	$(1-\theta)^2/4$	$[\theta(1-\theta)]/2$	$\theta^2/4$	$[\theta(1-\theta)]/2$	$[\theta^2+(1-\theta)^2]/2$	$[\theta(1-\theta)]/2$	$\theta^2/4$	$[\theta(1-\theta)]/2$	$(1-\theta)^2/4$
Ad/aD × AD/ad	$8\ h_1\ h_2\ h_3\ h_4$	$[\theta(1-\theta)]/4$	$[\theta^2+(1-\theta)^2]/4$	$[\theta(1-\theta)]/4$	$[\theta^2+(1-\theta)^2]/4$	$\theta(1-\theta)$	$[\theta^2+(1-\theta)^2]4$	$[\theta(1-\theta)]/4$	$[\theta^2+(1-\theta)^2]/4$	$[\theta(1-\theta)]/4$
Ad/aD × Ad/aD	$4\ h_2^2\ h_3^2$	$\theta^2/4$	$[\theta(1-\theta)]/2$	$(1-\theta)^2/4$	$[\theta(1-\theta)]/2$	$[\theta^2+(1-\theta)^2]/2$	$[\theta(1-\theta)]/2$	$(1-\theta)^2/4$	$[\theta(1-\theta)]/2$	$\theta^2/4$
Ad/ad × AD/ad	$8\ h_1\ h_2\ h_4^2$	0	0	0	$(1-\theta)/4$	1/4	$\theta/4$	$\theta/4$	1/4	$(1-\theta)/4$
Ad/ad × Ad/aD	$8\ h_2^2\ h_3\ h_4$	0	0	0	$\theta/4$	1/4	$(1-\theta)/4$	$(1-\theta)/4$	1/4	$\theta/4$
Ad/ad × Ad/ad	$4\ h_2^2\ h_4^2$	0	0	0	0	0	0	1/4	1/2	1/4

essary assumption (Elston et al. 1992). If the sample consists of only nuclear families, the correlation structure can be modeled by the mixed-model approach in software packages such as SAS or SUDAAN. For the purpose of power calculations, Allison (1997) assumed that the marker locus is the disease locus itself. When we ran the simulation under that assumption, using our method, the power was extremely high (>99%) for all combinations of family and offspring sizes. Another advantage of our method is that it allows us to estimate the effects of other relevant covariates simultaneously with the detection of linkage.

As with other TD-type tests, this new test can be very powerful when markers are extremely close to the disease locus, so that they are in disequilibrium with it. In principle, this test should be more powerful than the conventional TDT using McNemar's test after dichotomizing the disease trait. We can include relevant covariates, their interactions, and interactions with the marker allele in the model and then test their significance. Furthermore, a regression model such as this will enable us to use regression diagnostics to check the model's validity with respect to the various distributional assumptions. Our simulation study substantiates the fact that the proposed test is valid and very powerful, especially when the marker locus is close to the disease locus.

The method proposed here assumes that the trait values—or, more correctly, the residuals—are normally distributed. If there is significant departure from normality, then the Box-Cox power transformation or the general class of power transformations proposed by George and Elston (1988) can be incorporated to induce approximate normality. This is an option available in the ASSOC program.

All TDT methods test for linkage in the presence of population association. As is evident from the expression for γ, in equation (4), the test can have power only if $\Delta \neq 0$. Therefore, in the absence of the knowledge of the presence of any population association, no TDT method can detect linkage. It is, however, possible to perform a two-stage procedure in which, at the first stage, a test of association is performed, and in which, if significant association is found, we can proceed with the second stage, of testing for linkage. The procedure is outlined as follows:

1. Regress the disease trait Y_i on the covariates and the association variable Z_i, using the model

$$Y_i = \alpha + \beta_1 C_{1i} + \ldots + \beta_k C_{ki} + \delta Z_i + E_i \,, \qquad (5)$$

where Z_i is defined as 1 if the associated allele A is present in the ith individual and is defined as 0 otherwise. Perform the likelihood-ratio test or the t-test to assess the significance of association.

2. If significant association is found in stage 1, then regress Y_i on the covariates and X_{li}, using the model given in equation (2), and perform the test for linkage, as indicated above.

An advantage of using this two-stage procedure is that we need perform the second-stage analysis only if significant allelic association is found in the first stage. We can include in the sample the parents and offspring who are noninformative for linkage, when performing the test for association in the first stage.

The method proposed here can easily be extended for analysis of linkage between marker loci and binary traits, by modeling the trait by means of a logistic-regression approach. We replace Y_i in either equation (2) or equation (5) by the logit-function $\log_e, p/(1-p)$, where p is the probability of the disease, under the model assumption that the effect of linkage is additive with respect to the logit function. Again, for nuclear families, we can use the SAS or SUDAAN software packages to perform the analysis, using some form of correlation structure among family members. However, incorporating the correlation structure and computing the likelihood for extended pedigrees may be more difficult to implement. Finally, it should be pointed out that, when a genome scan is performed with tests on a large number of markers within a region of interest, appropriate adjustments of the significance level must be made, to account for multiple testing and to control the genomewide significance level.

Acknowledgments

This study was supported in part by U.S. Public Health Service research grants GM 28353, from the National Institute of General Medical Sciences, and HL 55055, from the National Heart, Lung and Blood Institute, and by resource grant RR03655 from the National Center for Research Resources. The authors thank the reviewers for their valuable critiques, which helped us to make substantial changes and improvements to an earlier version of this report.

Appendix

Let A and a, the alleles at the marker locus, have probabilities p and $1-p$, respectively, and let D and d, the alleles at the trait locus, have probabilities q and $1-q$, respectively. Δ is defined as $\Delta = P(\mathrm{AD}) - pq$. The haplotype frequencies with respect to the two loci, in terms of the allelic frequencies and Δ, are given by

$$h_1 = P(\mathrm{AD}) = pq + \Delta ,$$
$$h_2 = P(\mathrm{Ad}) = p(1-q) - \Delta ,$$
$$h_3 = P(\mathrm{aD}) = (1-p)q - \Delta ,$$
$$h_4 = P(\mathrm{ad}) = (1-p)(1-q) + \Delta .$$

Thus, $\Delta = h_1 h_4 - h_2 h_3$. Let the trait genotypic means μ, ν, and $-\mu$ correspond to genotypes DD, Dd, and dd, respectively. Without loss of generality, we can assume that the model given by equation (2) contains no covariates and that X_i and E_i in equation (2) are independent. Then parameter γ in equation (4) can be determined as follows:

$$\begin{aligned}\gamma &= E(Y_i|X_i = 1) - E(Y_i|X_i = 0)\\ &= \mu P(\mathrm{DD}|X_i = 1) + \nu P(\mathrm{Dd}|X_i = 1)\\ &\quad -\mu P(\mathrm{dd}|X_i = 1) - \mu P(\mathrm{DD}|X_i = 0)\\ &\quad -\nu P(\mathrm{Dd}|X_i = 0) + \mu P(\mathrm{dd}|X_i = 0)\\ &= \mu\left[\frac{P(\mathrm{DD},X_i = 1)}{P(X_i = 1)} - \frac{P(\mathrm{dd},X_i = 1)}{P(X_i = 1)}\right.\\ &\quad \left. - \frac{P(\mathrm{DD},X_i = 0)}{P(X_i = 0)} + \frac{P(\mathrm{dd},X_i = 0)}{P(X_i = 0)}\right]\\ &\quad +\nu\left[\frac{P(\mathrm{Dd},X_i = 1)}{P(X_i = 1)} - \frac{P(\mathrm{Dd},X_i = 0)}{P(X_i = 0)}\right] .\end{aligned}$$

Because $P(X_i = 1) = P(X_i = 0) = p(1-p)(2 - 3p + 3p^2)$ for the informative matings given in figure 1, we have

$$\begin{aligned}\gamma = [1/P(X_i = 1)]\{&\mu[P(\mathrm{DD},X_i = 1) - P(\mathrm{dd},X_i = 1)\\ &-P(\mathrm{DD},X_i = 0) + P(\mathrm{dd},X_i = 0)]\\ &+\nu[P(\mathrm{Dd},X_i = 1) - P(\mathrm{Dd},X_i = 0)]\} .\end{aligned} \tag{A1}$$

Let the offspring genotype at the trait locus be denoted by G_t, the offspring genotype at the marker locus be G_m, the parental mating type for the marker locus be M_m, and the parental two-locus mating type be M_{mt}. Then

$$\begin{aligned}P(G_t,X_i = 1) &= P(G_t,\mathrm{AA},\mathrm{AA} \times \mathrm{Aa})\\ &\quad +P(G_t,\mathrm{Aa},\mathrm{Aa} \times \mathrm{aa})\\ &\quad +P(G_t,\mathrm{AA},\mathrm{Aa} \times \mathrm{Aa}) ,\\ P(G_t,X_i = 0) &= P(G_t,\mathrm{Aa},\mathrm{AA} \times \mathrm{Aa})\\ &\quad +P(G_t,\mathrm{aa},\mathrm{Aa} \times \mathrm{aa})\\ &\quad +P(G_t,\mathrm{aa},\mathrm{Aa} \times \mathrm{Aa}) ,\end{aligned} \tag{A2}$$

where $P(G_t,G_m,M_m)$ is the joint probability of offspring trait and marker genotypes and of the parental mating type at the marker locus. Now, $P(G_t,G_m,M_m)$ can be calculated by noting that

$$P(G_t,G_m,M_m) = \sum P(M_{mt})P(G_t,G_m|M_{mt}) , \tag{A3}$$

where the summation is over all values from the set of two-locus parental mating types that are consistent with both the mating type M_m and the genotypes of the offspring that correspond to the trait and marker loci. The probabilities $P(M_{mt})$ and $P(G_t,G_m|M_{mt})$ are given in tables 2, 3, and 4, for the informative matings AA × Aa, Aa × aa, and Aa × Aa, respectively, which are adapted from the work of Zhu (1999). When equations (A2) and (A3) and the probabilities given in the tables 2–4 are used, equation (A1) simplifies to

$$\begin{aligned}\gamma = &\frac{2(1-2\theta)\Delta}{p(1-p)(2-3p+3p^2)}\\ &\times\{(1-p+p^2)[\mu + (1-2q)\nu] + (1-2p)\nu\Delta\} .\end{aligned}$$

References

Allison DB (1997) Transmission-disequilibrium tests for quantitative traits. Am J Hum Genet 60:676–690

Cleves MA, Olson JM, Jacobs KB (1997) Exact transmission-disequilibrium tests with multiallelic markers. Genet Epidemiol 14:337–347

Elston RC, George V, Severtson F (1992) The Elston-Stewart algorithm for continuous genotypes and environmental factors. Hum Hered 42:16–27

Elston RC, Stewart J (1971) A general model for the genetic analysis of pedigree data. Hum Hered 21:523–542

George V, Elston RC (1987) Testing the association between polymorphic markers and quantitative traits in pedigrees. Genet Epidemiol 4:193–201

George V, Elston RC (1988) Generalized modulus power transformations. Commun Stat Theory Methods 17:2933–2952

Haseman JK, Elston RC (1972) The investigation of linkage between a quantitative trait and a marker locus. Behav Genet 2:3–19

Lindsey JK (1996) Parametric statistical inference. Oxford University Press, New York

Littell RC, Milliken GA, Stroup WW, Wolfinger RD (1996) SAS system for linear models. SAS Institute, Cary, NC

Martin ER, Kaplan NL, Weir BS (1997) Tests for linkage and association in nuclear families. Am J Hum Genet 61: 439–448

Risch NJ, Zhang H (1996) Mapping quantitative trait loci with extreme discordant sib pairs: sampling considerations. Am J Hum Genet 58:836–843

SAGE (1998) Statistical analysis for genetic epidemiology, release 3.1. Department of Epidemiology and Biostatistics, Case Western Reserve University, Cleveland

Shah BV, Barnwell BG, Bieler GS (1997) SUDAAN user's manual, release 7.5. Research Triangle Institute, Research Triangle Park, NC

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

Zhu X (1999) Association between traits and genetic markers. PhD diss, Department of Epidemiology and Biostatistics, Case Western Reserve University, Cleveland