Am. J. Hum. Genet. 64:1754–1764, 1999

Sibling-Based Tests of Linkage and Association for Quantitative Traits

David B. Allison,[1] Moonseong Heo,[1] Norman Kaplan,[2] and Eden R. Martin[3]

[1]Obesity Research Center, St. Luke's–Roosevelt Hospital Center, Columbia University College of Physicians and Surgeons, New York; [2]Biostatistics Branch, Department of Health and Human Resources, National Institute of Environmental Health Sciences, National Institutes of Health, Bethesda; and [3]Section Medical Genetics, Department of Medicine Duke University Medical Center, Durham, NC

Summary

The transmission/disequilibrium test (TDT) developed by Spielman et al. can be a powerful family-based test of linkage and, in some cases, a test of association as well as linkage. It has recently been extended in several ways; these include allowance for implementation with quantitative traits, allowance for multiple alleles, and, in the case of dichotomous traits, allowance for testing in the absence of parental data. In this article, these three extensions are combined, and two procedures are developed that offer valid joint tests of linkage and (in the case of certain sibling configurations) association with quantitative traits, with use of data from siblings only, and that can accommodate biallelic or multiallelic loci. The first procedure uses a mixed-effects (i.e., random and fixed effects) analysis of variance in which sibship is the random factor, marker genotype is the fixed factor, and the continuous phenotype is the dependent variable. Covariates can easily be accommodated, and the procedure can be implemented in commonly available statistical software. The second procedure is a permutation-based procedure. Selected power studies are conducted to illustrate the relative power of each test under a variety of circumstances.

Introduction

Locating genes influencing quantitative traits in humans remains a challenging task. Linkage studies have substantial power problems under many plausible scenarios (Blackwelder and Elston 1982; Risch and Merikangas 1996; Allison and Schork 1997; Collins et al. 1997). Association studies, when the marker locus is the trait locus or is both linked and associated with the trait locus, can be substantially more powerful but are subject to confounding due to population "admixture" or "stratification" (Ewens and Spielman 1995). In this report, we will use the term "linkage disequilibrium" to refer to cases in which there is an association between alleles at two different loci that are linked.

The transmission/disequilibrium test (TDT) developed by Spielman et al. (1993) can be a powerful family-based test of linkage and, in some cases, a test of association as well as linkage. As a test of linkage, the TDT has high power in the presence of association and will not give false positives due to the presence of inappropriate controls. The TDT has been extended in a number of ways (e.g., see Curtis and Sham 1995; Sham and Curtis 1995; Morris et al. 1997). In the context of the present report, two developments are noteworthy. First, Allison (1997) has extended the TDT for use with quantitative (i.e., continuously distributed) traits, and Rabinowitz (1997) has further extended this effort, to allow for the inclusion of multiallelic loci. Second, Spielman and Ewens (1998), Curtis (1997), and Boehnke and Langefeld (1998) recently have developed family-based tests of linkage and association that do not require information about parents; all that they require is that sibships contain at least one affected sibling and one unaffected sibling and that not all siblings have the same genotype.

The purpose of the present study is to develop family-based tests of linkage, in some cases of both association and linkage, that can be applied to quantitative traits among siblings, in the absence of parental information. This is important because parental information will often be unavailable, because of financial limitations or other practical constraints. This is especially so when the trait of interest is studied among older individuals; for example, it would be very difficult to recruit an adequate sample with parents when the phenotype is rate of lean-body-mass loss (sarcopenia [Rosenberg 1997]) among elderly individuals.

The remainder of this report consists of three sections. First, the logical justification for the proposed procedures is presented, and the statistical approaches are described. Second, a discussion of power and sample-size requirements for the procedures is presented with some

Received June 11, 1998; accepted for publication March 25, 1999; electronically published May 4, 1999.

Address for correspondence and reprints: Dr. David B. Allison, Obesity Research Center, St. Luke's/Roosevelt Hospital, 1090 Amsterdam Avenue, 14th Floor, New York, NY 10025. E-mail: dba8@columbia.edu

0002-9297/99/6406-0029$02.00

illustrative simulations. Finally, a discussion of related issues is offered.

Proposed Approach: Justification and Statistical Models

If an association is observed to exist between phenotypes and marker-locus genotypes, this association can logically be due to several things:

1. the phenotype causes variations in the marker genotype;
2. the marker genotype (either directly or indirectly through intermediary phenotypes) causes variations in the phenotype;
3. the marker locus is both linked to and *associated with another locus that causes variations in the phenotype;*
4. the alleles at a marker locus are associated with (but not linked to) some other inherited factor that causes variation in the phenotype.

Only associations due to points 2 and 3 are useful in the localization of genes affecting the trait, and so the goal is to conduct a test of association that rules out points 1 and 4. Point (1) can be ruled out a priori, both as being logically impossible because genotype precedes phenotype in time and because it is a fundamental axiom of causality that cause must precede effect (Hume 17XX [1988]). Allison (1997) has pointed out that conditioning on parents' genotype at the marker locus is sufficient to eliminate point 4 as a possible explanation for an observed association between offspring's phenotype and genotype at the marker locus. This is because, in the absence of linkage and conditional on parental genotypes at the marker locus, there is no population association between offspring's genotypes at the marker locus and other inherited factors, by the law of independent assortment (of course, observed significant sample associations can also occur by chance).

Furthermore, it is obvious that, when sibships consist of full sibs, the probabilities of genotypes of siblings within a sibship depend totally on the genotypes of their parents. With unknown parental genotypes, conditioning on sibship—that is, controlling for the effects of membership in a particular sibship—is equivalent to controlling for parental genotypes, since all siblings within a sibship have the same parents and, therefore, the same parental genotypes. This eliminates the possibility of confounding by population stratification. Previous authors (Curtis 1997; Boehnke and Langefeld 1998; Spielman and Ewens 1998) have used this idea to develop sibship tests of linkage and, for certain data structures, tests of linkage and association.

Herein, we construct and evaluate two such tests, with samples of unrelated sibships. The first is a mixed-effects model; the second is a permutation test. The null hypothesis tested is that there is no linkage between the marker locus and the trait locus. However, the null hypothesis to be tested becomes lack of both linkage and association with the trait locus in sibships of minimal configurations—a minimally configured sibship consists of two and only two siblings with different genotypes and phenotypes. Both tests provide inferences conditional on sibships and, therefore, control for any possible population stratification. The mixed-effects model controls for such stratification by evaluating variation (of phenotypes) due to genotype within a sibship. The permutation test also controls for such stratification, because the permutation will be performed within each sibship, which therefore leads to inferences conditional on sibship.

This is important because such conditional inferences make both tests valid for testing of linkage; that is, conditional on sibship, siblings are randomly assigned to genotypes, and the probability of genotypes at two or more loci can only be dependent if they are linked. This implies that tests of the association between a genetic marker and a phenotype that are conducted conditional on sibship are valid tests of linkage regardless of whether the sibships are all of minimal configuration. Regarding the assertion that only sibships of minimal configuration yield valid tests of association between alleles at two loci, the logic has been previously elaborated by Curtis (1997), Boehnke and Langefeld (1998), and Spielman and Ewens (1998). This association occurs when there exists a probabilistic dependency between alleles at two loci in the total population (Elston 1998).

This approach is fundamentally different than tests of association that allow family data to be used but that do not condition on sibships. For example, Trégouët et al. (1997) have offered a generalized estimating equations (GEE) method of analysis of data for association in which some of the data are correlated. In this context, their method is not controlling for stratification (as they acknowledge) and is therefore a valid test of association but not of linkage. It is closely akin to the procedure developed by George and Elston (1987) and implemented in the SAGE (1997) software.

Mixed-Effects Model

Consider a collection of J sibships such that each sibship consists of two or more full (i.e., non-MZ) siblings. The number of siblings in the jth sibship will be denoted by K_j. Let n_{ij} denote the number of siblings in the jth sibship with the ith genotype (with $\Sigma_i n_{ij} = K_j$). The phenotype for the kth sibling of the ith genotype in the jth sibship will be denoted by Y_{ijk}. The marker locus M is assumed to have alleles $M_1, M_2, \ldots, M_m$, so that there are as many as $\frac{m(m+1)}{2}$ genotypes in total. In this situation,

the genotype can be considered to be a fixed factor A with $\frac{m(m+1)}{2}$ levels, and the sibship can be considered to be a random factor B with J levels. Therefore, a two-factor mixed-effects model (i.e., one with fixed and random effects) for the phenotype can be formulated as follows (e.g., see Burdick and Graybill 1992; Neter et al. 1996):

$$Y_{ijk} = \mu + \alpha_i + \beta_j + (\alpha\beta)_{ij} + e_{ijk} , \qquad (1)$$

where $i = 1, \ldots, I[= \frac{m(m+1)}{2}]$; $j = 1, \ldots, J$; $k = 1, \ldots, n_{ij}$ ($\geqslant 0$); $\Sigma_i n_{ij} = K_j$; μ is a constant; the effect sizes α_i are for the fixed (genotype) factor A; the effect sizes β_j are for the random (sibship) factor B; and the interaction effects $(\alpha\beta)_{ij}$ are also random.

With this formulation, testing of genotypic effects on the phenotype is conditional on sibship, since the sibship effect is introduced into the model as being random. However, under almost any sampling scheme, the number of individuals within each sibship with each genotype will not be constant across all sibships. This implies that the model in equation (1) is unbalanced; that is, the n_{ij} are not all equal. Statistical inferences under this mixed unbalanced model can be complex (Burdick and Graybill 1992). Although there are many approaches to estimation of variance components and to testing of their significance, we restricted ourselves to testing the mixed-effects model by an analysis of variance (ANOVA)–based F-test (Burdick and Graybill 1992, chap. 6).

Permutation Test

We next consider an alternative approach, by using a test statistic, S, whose distribution can be approximated by an appropriate χ^2. Let Y_{jk} be the trait value of the kth child in the jth sibship, where $j = 1, \ldots, J$ and $k = 1, \ldots, K_j$. And let N_{ijk} be the number of copies of the ith allele (not genotype) in the kth child in jth sibship, where $i = 1, \ldots, m$ (the total number of alleles). Then the permutation mean, μ_{ij}, of the trait value for the ith allele observed in the jth sibship can be computed as

$$\mu_{ij} = \frac{1}{K_j!}\sum_{p=1}^{K_j!}\sum_{k=1}^{K_j}\tilde{Y}_{p:jk}N_{ijk} = \frac{1}{K_j}\sum_{k=1}^{K_j}Y_{jk}\sum_{k=1}^{K_j}N_{ijk} ,$$

where $\tilde{Y}_{p:jk}$ is the pth permuted trait value for the kth child in the jth sibship. Specifically, the permutation is performed within sibships so that $\tilde{Y}_{p:j}$ becomes a vector such that $\tilde{Y}_{p:j} = (\tilde{Y}_{p:j1}, \ldots \tilde{Y}_{p:jK_j})'$. Similarly, the permutation variance V_{ij} of the trait value for the ith allele observed in the jth sibship can be computed as

$$\begin{aligned}
V_{ij} &= \frac{1}{K_j!}\sum_{p=1}^{K_j!}\left(\sum_{k=1}^{K_j}\tilde{Y}_{p:jk}N_{ijk} - \mu_{ij}\right)^2 \\
&= \frac{1}{K_j!}\sum_{p=1}^{K_j!}\left(\sum_{k=1}^{K_j}\sum_{k'=1}^{K_j}\tilde{Y}_{p:jk}N_{ijk}\tilde{Y}_{p:jk'}N_{ijk'}\right) - \mu_{ij}^2 \\
&= \frac{1}{K_j!}\sum_{p=1}^{K_j!}\left(\sum_{k=1}^{K_j}\tilde{Y}_{p:jk}N_{ijk}^2 + \sum_{k=1}^{K_j}\sum_{k'\neq k}^{K_j}\tilde{Y}_{p:jk}N_{ijk}\tilde{Y}_{p:jk'}N_{ijk'}\right) - \mu_{ij}^2 \\
&= \frac{1}{K_j}\sum_{k=1}^{K_j}Y_{jk}^2\sum_{k=1}^{K_j}N_{ijk}^2 + \frac{1}{K_j(K_j-1)}\left[\left(\sum_{k=1}^{K_j}Y_{jk}\right)^2 - \sum_{k=1}^{K_j}Y_{jk}^2\right] \\
&\quad \times\left[\left(\sum_{k=1}^{K_j}N_{ijk}\right)^2 - \sum_{k=1}^{K_j}N_{ijk}^2\right] - \mu_{ij}^2 .
\end{aligned}$$

This shows that V_{ij} can be expressed in terms of the quantities

$$\sum_{k=1}^{K_j}Y_{jk}, \sum_{k=1}^{K_j}Y_{jk}^2, \sum_{k=1}^{K_j}N_{ijk}, \text{ and } \sum_{k=1}^{K_j}N_{ijk}^2 .$$

The permutation statistic S can now be written as

$$S = \frac{m-1}{m}\sum_{i=1}^{m}\frac{\left[\sum_{j=1}^{J}\left(\sum_{k=1}^{K_j}Y_{jk}N_{ijk} - \mu_{ij}\right)\right]^2}{\sum_{j=1}^{J}V_{ij}} ,$$

which, under the null hypothesis of no linkage, can be approximated by a χ^2 distribution with $m - 1$ df.

Simulations and Results

In this section we attempt to provide some sense of the relative power of the two proposed procedures—the ANOVA-based F-test for the mixed-effects model and the test statistic S for the permutation test. We also attempt to evaluate the extent to which they appropriately hold the type I–error rate to the nominal levels. We assume that the marker locus being studied is the trait locus—that is, there is complete association between marker and trait alleles, and the recombination fraction is 0. For each phenotype, let h^2 (i.e., the locus-specific heritability) denote the variance proportion that is attributable to the locus in question.

The power of the proposed procedure will depend on either the noncentrality parameter of the F distribution, for the mixed-effects model, or the distribution for the permutation-test statistic S, given a particular design.

Computation of the noncentrality parameter for a mixed effects–model ANOVA can be extremely challenging when the design is unbalanced (Searle 1987). Therefore, we chose to produce power estimates via simulation. The simulation studies that we present are not intended to be an exhaustive sampling of the parameter space of interest but, rather, are meant to provide a few illustrative examples. The specific parameter sets for which the simulations were conducted were specifically chosen to illustrate key features of the test. The effect of genotype was tested via ANOVA-based *F*-ratios calculated from the type III sum of squares in SPSS's (1997) general linear-model routine. The type III sum of squares calculates the sum of squares of an effect in a design adjusted for all other effects being modeled (Searle 1987, chap. 12). The calculation of the test statistic *S* was programmed in S-plus (Becker et al. 1988).

Base Model

We began selecting a reasonable representation of a situation that an investigator might actually encounter, which we labeled as "base model." It included the following specifications: (1) The biallelic ($m = 2$) locus under study explained 10% of the total phenotypic variance (i.e., $h^2 = .1$. (2) The alleles at the locus under study acted in an additive fashion. (3) An additive polygenic component (A^2) explained an additional 30% of the phenotypic variance. (4) A nonadditive (dominance [D^2]) polygenic component explained an additional 10% of the phenotypic variance. (5) A shared common environmental component (C^2) explained an additional 20% of the phenotypic variance. (6) The remaining phenotypic variance, 30%, was explained by a residual environmental component (E^2) uncorrelated among siblings. (7) The frequency of allele M_1, the allele associated with greater values of the trait, denoted as p, was set equal to .2.

For this base model, data were simulated for four situations: a constant sibship size of two, for 75 sibships; a constant sibship size of three, for 50 sibships; a constant sibship size of four, for 38 sibships; and a constant sibship size of 5, for 30 sibships. For convenience, the sibship size was kept constant within each simulation. The sibship sizes selected are those that are relatively common in the population (Eggebeen 1992). The number of sibships was selected to maintain the total number of individuals under study as close to 150 as possible allowing a comparison of the relative power of using different sibship sizes given the same number of subjects studied overall. For each situation, power values were based on results from 1,000 simulated data sets.

Finally, it should be noted that the number of sampled sibships listed in table 1 is the total number of sibships sampled, not the number that were informative; that is, sibships in which all siblings have the same genotype add no information to a test of the genotypic effects (similarly, a sibship in which all siblings have the same phenotype also add no information, although, theoretically, this situation will not occur with truly continuous traits). If only informative sibships were selected, power would obviously be higher. Unfortunately, one does not know in advance of doing a study whether sibships will be informative. The probability that any randomly selected sibship with K siblings will be uninformative is a function of K and p. Specifically, if random mating and a biallelic locus are assumed, the probability that a randomly selected family will be informative is

$$P(I) = 1 - \sum_{t=1}^{6} \sum_{i=1}^{3} P(M_t)P(G_i|M_t)^K ,$$

where $P(M_t)$ is the probability of parental mating type t, $P(G_i|M_t)$ is the probability of offspring having genotype G_i, given parents of mating type t, and K is the number of siblings in the sibship (see the Appendix).

The results of the power studies are shown in table 1. As can be seen, when the gene explains only 10% of the variance and the conditions described above hold, studying 150 subjects yields generally adequate power at the two-tailed $\alpha = .05$ but generally inadequate power for more-stringent values of α. It is noteworthy that, in this situation, switching from sibling pairs to larger sibships increases power even for the same number of subjects. This is consistent with many other gene-mapping situations (Todorov et al. 1997) and expectations from theory (Keppel 1982, p. 535). Although $\alpha = .05$ is not necessarily the most appropriate value to use for power studies in genetics, results for a reduced significance level, such as $\alpha = .01$ or $\alpha = .001$, show the same qualitative results (table 1); in other words, the relative performances are the same, and it is the *pattern* of results that is important in the present context. Researchers planning new research are encouraged to estimate power by using the α value appropriate to their situation. However, it is interesting that, in this situation, the added power derived from the use of large sibships is confined primarily to the increase from two to more than two siblings per sibship; in the current situation, there is only a modest gain in going beyond trios.

Extreme Sampling

It has been shown, in many contexts, that sampling of phenotypically extreme individuals and relatives can markedly increase the power of gene-mapping studies (Eaves and Meyer 1994; Risch and Zhang 1995, 1996; Allison and Schork 1997; Allison et al. 1998). Most demonstrations of this phenomenon have been limited to the case of sibling pairs for which the definition of

Table 1

Power Results from Mixed Effects–Model ANOVA and from Permutation-Test Statistic *S*, for Simulated Phenotypes When Total Number of Alleles (M and m) Is 2

No. of Simulations	ρ[a]	K	J	QTL Variance[b]		Effect of Genotype			Frequency of m	Comment	Power					
											Mixed Effects–Model ANOVA, for α=			Statistic *S*, for α =		
				Additive	Dominance	MM	Mm	mm			.05	.01[c]	.001[c]	.05[c]	.01[c]	.001[c]
1,000	.42	2	75	.10	.00	−.224	.335	.894	.2	Base	.56	.32	…	.69	.35	…
1,000	.42	3	50	.10	.00	−.224	.335	.894	.2	Base	.71	.43	.17	.82	.59	.24
1,000	.42	4	38	.10	.00	−.224	.335	.894	.2	Base	.75	.48	.19	.88	.69	.33
1,000	.42	5	30	.10	.00	−.224	.335	.894	.2	Base	.78	.52	.20	.88	.72	.39
1,000	.00	4	38	.10	.00	−.224	.335	.894	.2	No correlation	.51	.27	…	.66	.40	.14
1,000	.42	4	38	.03	.07	−.065	−.065	1.549	.2	Recessive m	.77	.58	.33	.37	.17	…
1,000	.42	4	38	.09	.01	−.237	.422	.422	.2	Dominant m	.80	.54	.27	.86	.64	.29
1,000	.42	4	38	.00	.10	−.316	.316	−.316	.5	Overdominance	.92	.77	.47	…	…	…
1,000	.42	4	38	.10	.00	−.447	.000	.447	.5	Additive	.77	.53	.24	.89	.71	.39
1,000[d]	.42	4	38	.10	.00	−.224	.335	.894	.2	Extreme	.98	.93	.78	1.00	.99	.90
10,000	.38	4	38	.00	.00	.000	.000	.000	.2	Null	.0455	.0087	.0009	.0468	.0075	.0003
10,000[d]	.38	4	38	.00	.00	.000	.000	.000	.2	Null/extreme	.0461	.0101	.0019	.0496	.0078	.0009
10,000	.38	4	38	.00	.00	.000	.000	.000	.9/.05[e]	Null/population stratification	.0485	.0114	.0016	.0484	.0059	.0001
…[f]		1	50	.10	.00	−.224	.335	.894	.2	Parent based	.26	…	…	.26	…	…

[a] A value of .42 occurs if, for example, the QTL explains 10% of the phenotypic variance and $A^2 = .3$, $D^2 = .1$, $C^2 = .2$, and $E^2 = .3$.

[b] $h^2 = \sigma_A^2 + \sigma_D^2$.

[c] An ellipsis (…) denotes that the estimated power is <.10.

[d] Extreme sampling; for this situation, the "best" 38 were selected from 200 initial sibships, on the basis of the Mahalanobis distance score of phenotypes among siblings in each sibship

[e] .90 is the frequency in population 1; .05 is the frequency in population 2.

[f] Powers for the parent-based TDT have been analytically calculated.

extreme concordance and discordance is somewhat simpler. To our knowledge, there has been no development of a criterion for extreme sampling in the context of larger sibships.

For this purpose, we chose to use the Mahalanobis distance, calculated on the basis of the phenotypic scores of all siblings within each sibship. The Mahalanobis distance can be defined as a standardized (by the inverse of the covariance matrix) distance between any two points in multidimensional space; that is, $(\mathbf{X}_i - \mathbf{X}_j)^T\Sigma^{-1}(\mathbf{X}_i - \mathbf{X}_j)$, for points $\mathbf{X}_i$ and $\mathbf{X}_j$ in a multidimensional space, where Σ is the covariance matrix of the points (or the multivariates). In the present study, however, that is the standardized distance between a point and the center of a multivariate distribution. This method assigns to each sibship a score indicating its degree of "multivariate unusualness" (Krzanowski 1988, p. 234). Selecting the sibships with the highest values of the Mahalanobis distance seems to be a reasonable way to select phenotypically extreme sibships with more than two siblings per sibship. Using this approach, we simulated 200 randomly sampled sibships of size four, given the base model described above, and then selected the 38 sibships with the highest value of the Mahalanobis distance. This dramatically increased the power for both the mixed-effects model and the permutation test (table 1). Especially, the permutation test now had an excellent power even for $\alpha = .001$ and had perfect power for $\alpha = .05$ (table 1).

The Effect of Residual Correlation

The quantities A^2, D^2, and C^2 come together to influence the residual correlation among siblings (i.e., the correlation among siblings' phenotypes that is conditional on genotype). Specifically, the residual correlation ρ is $\rho = (\frac{A^2}{2} + \frac{D^2}{4} + C^2)/(1 - h^2)$ (Neale and Cardon 1992). As has been shown elsewhere, the power of variance-components linkage approaches for detection of genes in sib-pair designs is generally increased when there is greater residual correlation (Schork 1993). In the context of the mixed-effects model, we would also expect increased residual correlation to increase power, because this increases the magnitude of the random effect of the sibship and, thereby, decreases both the residual variance and the denominator of the *F*-statistic testing the effect of genotype. To observe the influence that sibling residual correlation has on the sibling-based tests (proposed herein), we simulated the base model, again using four siblings per sibship and 38 sibships, but we set the values of A^2, C^2, and D^2 at 0 and set E^2 at .9. This yielded a residual correlation of 0. When the residual correlation was reduced from its initial value to 0, power decreased substantially, for both the mixed-effects model and the permutation test (table 1). This illustrates that, all other things being equal, phenotypes with greater "familiality" will be able to be mapped with greater power by use of sibling-based tests.

Comparison with a Parent-Based TDT

It is interesting to ask what the relative power of these sibling-based tests would be if compared with a parental TDT (Allison 1997) that used the same amount of resources. In the TDT, at least three individuals must be genotyped within each nuclear family: the father, the mother, and one or more offspring. We therefore calculated the power to detect a quantitative-trait locus (QTL) under the "base" model described earlier, with 50 nuclear families each consisting of two parents and one offspring; that is, we used the same number of individuals (150) that was used in the preceding simulations. Power was calculated by methods described by Allison (1997) and was checked via simulation. The analytic and simulation results agreed very closely, and the former are reported herein. Table 1 shows that, when only one sibling per nuclear family is used, the parent-based TDT is dramatically less powerful than both sibling-based tests, on a per person basis (although the sib-based tests require more phenotyping). This may be due to the difference between the expected number of informative observations that is useful for the parent-based test versus the sibling-based test; for example, in the current situation, where $p = .2$, with 50 nuclear families (a total of 150 individuals) for the parent-based test for a bialleic locus, the expected number of informative observations is 27 data units, whereas, in contrast, with 75 sibling pairs (a total of 150 individuals) for the sibling-based test for a bialleic locus, the expected number of informative observations is 42 data units. Alternatively, one might conjecture that the reduced power of the parent-based TDT might be due to the fact that, conceptually, a trio yields the equivalent of one known phenotype and one unknown phenotype (relating to the untransmitted alleles), rather than the three phenotypes that one gets from three siblings.

Effect of Dominance at the Marker Locus

Because the correlation, among siblings, for the phenotype (the total correlation, not the residual correlation) is a function of A^2, C^2, D^2 *and* the effects of the marker locus (Neale and Cardon 1992), the random effect of sibship and the fixed QTL effect will be correlated to some extent. This colinearity affects the power of the mixed-model test for the fixed effect. All other things being equal, the lower this degree of colinearity, the more powerful the test. It is noteworthy, then, that the additive component of the QTL effect contributes more strongly to sibling correlations than the dominance component does; specifically, one half of the additive

variance of the QTL is contributed to the sibling covariance, whereas only one fourth of the dominance variance of the QTL is contributed to the sibling covariance.

Therefore, we simulated the power of the test by using the base model but varying, in three ways, the mode of inheritance. First, we set the mode of inheritance to be recessive such that the mean of the "wild-type" homozygotes and the mean of the heterozygotes was equal and only the mean of individuals homozygous for the increasing allele was higher. Second, we set the mode of inheritance to be dominant such that the mean of the wild-type homozygotes was less than those of individuals with the other two genotypes, with the means of the latter two genotypes being the same. Third, we selected an extreme case of dominance—namely, overdominance. In this model, only the heterozygote had any elevation in the mean of the phenotype, and we changed the value of p from .2 to .5. This particular configuration of means and gene frequency was selected to yield the highest possible dominance variance at the QTL, given the fixed QTL effect of 10%. We intentionally selected such a pattern in order to observe the power under an extreme situation, in which all of the QTL variance was nonadditive. Such overdominance (in which 100% of the QTL variance is nonadditive), although presumably rare, is not unheard of; for example, the callipyge sheep has been shown to exhibit such a pattern of inheritance for a quantitative trait relating to body composition (Georges and Cockett 1996). In addition, to ensure that any power difference observed in this third model was not due solely to increasing the allele frequency from .2 to .5, we also evaluated power for the strictly additive case with an allele frequency of .5.

In these cases where the ratio of dominance variance to additive variance at the QTL was increased dramatically while the total variance due to the QTL was kept constant, power was increased for the mixed-effects model but was decreased for the permutation test (table 1). Thus, when all other factors are held constant, QTL with more dominance effects will be detectable with greater power when the mixed-effects model is used. This effect appears to be due to dominance and not to changing the allele frequency to .5, since the model in which the allele frequency was changed but additivity remained unchanged resulted in no substantial change. However, it should be pointed out that the extreme situation of 100% of variance at the QTL being nonadditive need not occur, given overdominance, and was intentionally selected to yield a situation in which the relative power of each proposed test in response to variance in the degree of dominance would be portrayed in stark relief.

Type I–Error Rate

To confirm that the procedures held the type I–error rate at or below the nominal α values, we simulated models under the null hypothesis of no linkage (i.e., when the means of the genotypes are equal). For the simulations under the null hypothesis, given that the expected proportion of trials with significant results was much smaller, we increased the number of trials to 10,000, to increase the precision of our estimates. As can be seen in table 1, both the mixed-effects model and the permutation test performed quite adequately in this regard.

In addition, we simulated the null scenario, incorporating the extreme sampling procedure described above. This is especially important in order to evaluate the extent to which the observed increase in power with extreme sampling is either a "true" increase in power or an artifact of an inflated type I–error rate due to the extreme sampling's violation of assumptions of normality and independence (Wilcox 1997). It is noteworthy and reassuring that, as indicated in table 1, the extreme sampling employed herein did not affect the type I–error rate at all for α values as low as .01; for α values lower than that, the number of simulations conducted is too small to allow firm conclusions. It should be noted, however, that we are not claiming that any form of extreme sampling with any sample size will not affect the type I error rate; it is possible that other types of extreme sampling, with smaller sample sizes, might have untoward effects on the type I–error rate.

Finally, to see whether the type I–error rates of the two tests herein are robust to population stratification, we assumed two populations, 1 and 2, with the following conditions.

1. The frequencies of allele m are .90 and .05 for populations 1 and 2, respectively.
2. The population means are 1 and −1 for populations 1 and 2, respectively, with the same within-population variance (i.e., 1).
3. There is no genotypic effect for either population.

We simulated 19 sibships, with four siblings each, for each population and combined them into a single sample of 38 sibships with four siblings each, for 10,000 times. The results are shown in table 1. For both tests, the actual significance levels are seen to be near the nominal levels, although the permutation test tended to be a little conservative for lower significance levels.

Simulations with Multiallelic Loci

Finally, we evaluated the two proposed procedures' power with multiallelic loci. We chose to test our base model with $K = 4$ as the number of siblings in each

sibship, the only modification being that allele M_2 was relabeled as $M_2, M_3, \ldots, M_a$, where a was three to six. This, combined with allele M_1, gives us a multiallelic locus with the number of alleles being three to six and with the number of possible genotypes being 6–21. The frequencies of the alleles from M_2 to M_a are all the same; that is, $\frac{1-p}{a-1}$. The effects of the alleles were set to be identical; that is, if one were to collapse alleles M_2–M_a into a single category, then the multiallelic models tested would be identical to the base model. However, from the data analyst's point of view, there is no way of knowing a priori that the alleles M_2–M_a should be collapsed, and we therefore tested them as separate alleles, as has been described above.

The results of these simulations are presented in table 2. As can be seen, the power of the mixed effects–model test decreases as the number of alleles increases. This is expected, given both the greater number of degrees of freedom in the test and the unchanging sizes of effects. Although the power of the permutation test also decreases as the number of alleles increases, the drop is less than that for the mixed-effects model; therefore, with a larger number of alleles, the permutation test performs better. Whether this exact pattern of results would hold with a different distribution of allele frequencies is unknown.

Table 2

Comparison of Powers between Mixed Effects–Model ANOVA and Permutation-Test Statistic *S*, under Base Model for Multiple Alleles (with 1,000 Simulations), for *K* = 4 and *J* = 38

	Power, for α=		
Test and Total No. of Alleles	.05	.01	.001[a]
Mixed effects–model ANOVA:			
3	.64	.40	.16
4	.53	.28	…
5	.42	.21	…
6	.36	.16	…
Permutation-test statistic *S:*			
3	.71	.42	.11
4	.66	.37	.13
5	.65	.37	…
6	.58	.31	…

[a] An ellipsis (…) denotes that the estimated power is <.10.

Discussion

The procedures proposed in this paper allow the advantages of the TDT to be applied to sibling data when the outcome variable is a quantitative trait, even in the absence of parental information. There are several strengths to the procedures proposed. For the mixed-effects model, four advantages come to mind. First, the mixed-effects model can be readily implemented in commonly available software packages that have mixed effects–model procedures such as SAS and SPSS. Second, covariates such as age, sex, and measured life-style factors could be easily incorporated. Third and similarly, additional loci can be incorporated either as fixed factors or, when interaction terms are included, to test for epistasis (Frankel and Schork 1996). Fourth, the procedure also accommodates sibships of any and varying size (as long as there are at least two sibs). For the permutation test, the two major advantages are greater power in some cases and the fact that it is distribution free—that is, no assumptions about the marginal phenotype distribution is required.

For readers interested in real-world examples, Benjamin et al. (1996) and Mitchell et al. (1998) conducted tests, using sib pairs, that are valid sibling-based tests for quantitative traits (although they do not develop the general theory and procedures, as we have done herein). More recently, Chung et al. (in press) have applied both the mixed-effects model and the permutation test to data, examining the association between uncoupling protein-3 polymorphisms and obesity-related phenotypes.

Although the mixed-effects model has relatively little power for multiple alleles, because it is essentially assessing genotypic effects rather than allelic effects, genotypic effects (even those for the multiallelic markers) can be of great interest in terms of inferring, for example, mode of inheritance. Moreover, as we have shown, when there is nonadditivity at the QTL, power can be markedly enhanced by consideration of the genotypic, rather than the allelic, effects. Furthermore, it may not be wise to try to detect every single allelic effect when some alleles have very low frequency. Rather, by treating such alleles as a group (Kaplan et al. 1998) and applying the mixed-effects model, one may detect the genotypic effects, without loss of great power. The permutation test can, of course, be used for the detection of allele effects for multiallelic markers, and the mixed-effects model could also be adapted to this purpose.

One question that arises when one is considering the application of these procedures is which of the two techniques described herein should be used. Because, depending on the circumstances, each technique has advantages over the other, we are unable to make a blanket recommendation; that is, on the basis of the simulation results, the mixed-effects model performs relatively better when the degree of dominance is greater, and, in contrast, the permutation test performs better when the degree of dominance is minimal. Because investigators are not likely to know exactly the architecture of the

traits that they are studying, they are generally unable to choose which technique is the best for them a priori.

The obvious alternative—and the one that we suspect most investigators will adopt—is simply to run both analyses and to declare the results significant if either test produces a P value below one's chosen significance value for α. Although we do not explicitly advocate against this, investigators should be aware that such an approach will increase the overall type I–error rate under the null hypothesis, and, consequently, investigators adopting this approach should certainly make readers aware of this fact when publishing the results. The degree to which inflation of the type I–error rate is induced by conducting both the permutation test and the mixed effects–model test is unknown. If the two tests statistics were perfectly correlated, then there would be no inflation at all; if the correlation were 0 (a very unlikely possibility), then the type 1–error rate would be $1 - (1 - \alpha)^2$. Obviously, the reality lies somewhere between the two—and (we suspect) closer to the former.

Several extensions of this procedure might be envisioned in the future. First, it should be possible to incorporate multiple phenotypes into the analyses, via either procedures that correctly adjust the type I–error rate for multiple univariate analyses (Allison and Beasley 1998) or true multivariate procedures (Markel and Corley 1994). Second, in the event that the assumption of normality is not met and cannot be satisfied either with a transformation or by appeal to large-sample theory, nonparametric tests (e.g., see Good 1995) may be used, including the permutation test presented herein.

A third approach might be to develop a truly multipoint (as opposed to a simply multilocus) test that simultaneously uses the information from several loci to estimate both the point on a chromosome at which the QTL is most likely to lie and its confidence interval. This procedure might also be extended to analyze data from related sibships (e.g., sets of cousins). Currently, the model provided assumes that sibships are independent. However, dependent sibships might be accommodated by inclusion of additional nesting factors.

Finally, it should be noted that the procedure that has been developed herein assumes that the siblings are truly full siblings. The use of sibships including half-siblings allows for potential confounding. Fortunately, given a reasonable number of polymorphic markers genotyped on the siblings, inclusion of half-siblings due to nonpaternity can be ruled out with high probability (Jamieson 1994; Boehnke and Cox 1997; Goring and Ott 1997); however, this might be a more serious issue for studies using only a few single-nucleotide polymorphisms each for one or two candidate genes.

In conclusion, we note that previous work (Spielman et al. 1993; Allison 1997; Curtis 1997; Spielman and Ewens 1998) was extended to the development of tests that have the desirable properties of the TDT, can be applied to quantitative traits, and do not require parental information. These models should be flexible, allowing adaptation to many circumstances and additional developments.

Acknowledgments

We are grateful to Drs. Michael C. Neale and Hongyu Zhao for their helpful comments. This work was supported, in part, by National Institutes of Health grants R01DK51716, P30DK26687, and R29DK47256.

Appendix

The informative event I is the situation in which, for a sibship consisting of K (i.e., more than one) siblings, not all siblings have the same genotype. For a biallelic marker with both unknown parental mating type M_t ($t = 1,2,\ldots,6$) and sibling genotypes G_i ($i = 1,\ldots,3$), the probability $P(I)$ can be drawn as follows, where $G_i^{(k)}$ is the genotype of the kth ($k = 1,2,\ldots,K$) sibling in the sibship:

$$\begin{aligned} P(I) &= 1 - P(\bar{I}) \\ &= 1 - \sum_{t=1}^{6}\sum_{i=1}^{3} P\left[M_t \cap \left\{G_i^{(1)} = G_i^{(2)} = \cdots = G_i^{(K)}\right\}\right] \\ &= 1 - \sum_{t=1}^{6}\sum_{i=1}^{3} P\left[\left\{G_i^{(1)} = G_i^{(2)} = \cdots = G_i^{(K)}\right\} | M_t\right] P(M_t) \\ &= 1 - \sum_{t=1}^{6}\sum_{i=1}^{3} P[G_i | M_t]^K P(M_t) \ , \end{aligned}$$

where $\bar{I}$ is the complement of I—that is, is a noninformative event—and the last of the preceding equations is due to independence of sibling genotypes that is conditional on parental mating-type. The probabilities $P(\mathrm{M}_t)$ and $P(G_i | M_t)$ can be found in table A1, where p is the frequency of allele A and $q = 1 - p$.

Table A1

			$P(G_i \mid M_t)$		
t	M_t	$P(M_t)$	AA ($i = 1$)	Aa ($i = 2$)	aa ($i = 3$)
1	AA, AA	p^4	1	0	0
2	AA, Aa	$4p^3q$	$\frac{1}{2}$	$\frac{1}{2}$	0
3	AA, aa	$2p^2q^2$	0	1	0
4	Aa, Aa	$4p^2q^2$	$\frac{1}{4}$	$\frac{1}{2}$	$\frac{1}{4}$
5	Aa, aa	$4pq^3$	0	$\frac{1}{2}$	$\frac{1}{2}$
6	aa, aa	q^4	0	0	1

References

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

Allison DB, Beasley M (1998) A method and computer program for controlling the family-wise alpha rate in gene association studies involving multiple phenotypes. Genet Epidemiol 15:87–101

Allison DB, Heo M, Schork NJ, Wong SL, Elston RC (1998) Extreme selection strategies in gene mapping studies of oligogenic quantitative traits do not always increase power. Hum Hered 48:97–107

Allison DB, Schork NJ (1997) Selected methodological issues in meiotic mapping of obesity genes in humans: issues of power and efficiency. Behav Genet 27:401–421

Becker RA, Chambers JM, Wilks AR (1988) The new S language: a programming environment for data analysis and graphs. Wadsworth & Brooks/Cole, Pacific Grove, CA

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

Boehnke M, Cox NJ (1997) Accurate inference of relationships in sib-pair linkage studies. Am J Hum Genet 61:423–429

Boehnke M, Langefeld CD (1998) Genetic association mapping based on discordant sib pairs: the discordant-alleles test. Am J Hum Genet 62:950–961

Burdick RK, Graybill FA (1992) Confidence intervals on variance components. Marcel Dekker, New York

Chung WK, Luke A, Cooper RS, Rotini C, Vidal-Puig A, Rosenbaum M, Chua M, et al. Genetic and physiologic analysis of the role of UCP3 in energy homeostasis. Diabetes (in press)

Collins FS, Guyer MS, Chakravarti A (1997) Variations on a theme: cataloging human DNA sequence variation. Science 278:1580–1581

Curtis D (1997) Use of siblings as controls in case-control association studies. Ann Hum Genet 61:319–333

Curtis D, Sham PC (1995) A note on the application of the transmission disequilibrium test when a parent is missing. Am J Hum Genet 56:811–812

Eaves LJ, Meyer J (1994) Locating human quantitative trait loci: guidelines for the selection of sibling pairs for genotyping. Behav Genet 24:443–455

Eggebeen DJ (1992) Changes in sibling configurations for American preschool children. Soc Biol 39:27–44

Elston RC (1998) Linkage and association. Genet Epidemiol 15:565–576

Ewens WJ, Spielman RS (1995) The transmission/disequilibrium test: history, subdivision, and admixture. Am J Hum Genet 57:455–464

Frankel WN, Schork NJ (1996) Who's afraid of epistasis? Nat Genet 14:371–373

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

Georges M, Cockett N (1996) The ovine callipyge locus: a paradigm illustrating the importance of non-Mendelian genetics in livestock. Reprod Nutr Dev 36:651–657

Good PI (1995) Permutation tests: a practical guide to resampling methods for testing hypotheses. Springer-Verlag, New York

Goring HH, Ott J (1997) Relationship estimation in affected sib pair analysis of late-onset disease. Eur J Hum Genet 5: 69–77

Hume D (1748 [1988]) An enquiry concerning human understanding. Prometheus Books, Amherst, MA

Jamieson A (1994) The effectiveness of using co-dominant polymorphic allelic series for (1) checking pedigrees and (2) distinguishing full-sib pair members. Anim Genet 25 Suppl 1:37–44

Kaplan NL, Martin ER, Morris RW, Weir BS (1998) Marker selection for the transmission/disequilibrium test, in recently admixed populations. Am J Hum Genet 62:703–712

Keppel G (1982) Design and analysis, 2d ed. Prentice-Hall, Englewood Cliffs, NJ

Krzanowski WJ (1988) Principles of multivariate analysis. Oxford Science, Oxford

Markel PD, Corley RP (1994) A multivariate analysis of repeated measures: linkage of the albinism gene (Tyr) to a QTL influencing ethanol-induced anesthesia in laboratory mice. Psychiatr Genet 4:205–210

Mitchell BD, Blangero J, Comuzzie AG, Almasy LA, Shuldiner AR, Silver K, Stern MP, et al (1998) A paired sibling analysis of the beta-3 adrenergic receptor and obesity in Mexican Americans. J Clin Invest 101:584–587

Morris AP, Whittaker JC, Curnow RN (1997) A likelihood ratio test for detecting patterns of disease-marker association. Ann Hum Genet 61:335–350

Neale MC, Cardon LR (1992) Methodology for genetic studies of twins and families. Kluwer Academic, Dordrecht, the Netherlands

Neter J, Kutner MH, Nachtshem CJ, Wasserman W (1996) Applied linear statistical models, 4th ed. Irwin, Chicago

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

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

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

Rosenberg IH (1997) Sarcopenia: origins and clinical relevance. J Nutr 127 Suppl 5:990S-991S

SAGE (1997) Statistical analysis for genetic epidemiology, release 3.1. Computer program package software. Department of Epidemiology and Biostatistics, Rammelkamp Center for Education and Research, MetroHealth Campus, Case Western Reserve University, Cleveland

Schork NJ (1993) Extended multipoint identity-by-descent analysis of human quantitative traits: efficiency, power, and modeling considerations. Am J Hum Genet 53: 1306–1319

Searle SR (1987) Linear models for unbalanced data. John Wiley & Sons, New York

Sham PC, Curtis D (1995) An extended transmission/disequilibrium test (TDT) for multi-allele marker loci. Ann Hum Genet 59:323–336

Spielman RS, Ewens WJ (1998) A sibship test for linkage in the presence of association: the sib transmission/disequilibrium test. Am J Hum Genet 62:450–458

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

SPSS Inc (1997) SPSS advanced statistics, version 7.5. SPSS, Chicago

Todorov AA, Province MA, Borecki IB, Rao DC (1997) Trade-off between sibship size and sampling scheme for detecting quantitative trait loci. Hum Hered 47:1–5

Trégouët D-A, Ducimetière P, Tiret L (1997) Testing association between candidate-gene markers and phenotype in related individuals, by use of estimating equations. Am J Hum Genet 61:189–199

Wilcox RR (1997) Introduction to robust estimation and hypothesis testing. Academic Press, San Diego