Annals of human genetics

doi: 10.1111/j.1469-1809.2008.00502.x

Robust Quantitative Trait Association Tests in the Parent-Offspring Triad Design: Conditional Likelihood-Based Approaches

J.-Y. Wang[1] and J. J. Tai[2]*

[1]*Department of Healthcare Administration, Asia University, Taichung 41354, Taiwan*
[2]*Division of Biostatistics, College of Public Health, National Taiwan University, Taipei 10055, Taiwan*

Summary

Association studies, based on either population data or familial data, have been widely applied to mapping of genes underlying complex diseases. In family-based association studies, using case-parent triad families, the popularly used transmission/disequilibrium test (TDT) was proposed for avoidance of spurious association results caused by other confounders such as population stratification. Originally, the TDT was developed for analysis of binary disease data. Extending it to allow for quantitative trait analysis of complex diseases and for robust analysis of binary diseases against the uncertainty of mode of inheritance has been thoroughly discussed. Nevertheless, studies on robust analysis of quantitative traits for complex diseases received relatively less attention. In this paper, we use parent-offspring triad families to demonstrate the feasibility of establishment of the robust candidate-gene association tests for quantitative traits. We first introduce the score statistics from the conditional likelihoods based on parent-offspring triad data under various genetic models. By applying two existing robust procedures we then construct the robust association tests for analysis of quantitative traits. Simulations are conducted to evaluate empirical type I error rates and powers of the proposed robust tests. The results show that these robust association tests do exhibit robustness against the effect of misspecification of the underlying genetic model on testing powers.

Keywords: candidate-gene, exponential family, family-based association test, mode of inheritance, score test

Introduction

With the large amount of genetic variants that are available for use in association analyses nowadays, association studies, designed using either population data or familial data, have been proved useful in mapping of genes underlying complex diseases (Kruglyak & Lander, 1995; Risch & Merikangas, 1996; Schork et al., 2001; Cordell & Clayton, 2005). Traditional population-based studies are easier to collect data in practice, but often challenged for the population stratification problem that causes spurious results. To adjust for population stratification, methods using information of unlinked markers, for example, were proposed by several researchers (Devlin & Roeder, 1999; Pritchard & Rosenberg, 1999; Reich & Goldstein, 2001). Instead of use of statistical methods to adjust for population stratification in population-based association studies, on the other hand, family-based association studies avoid the stratification problem by use of the matched data structure generated from the familial data. For example, for case-parent triad families a matched case-pseudocontrol data structure can be created (Falk & Rubinstein, 1987). Based on this data structure a family-based association test, the transmission/disequilibrium test (TDT), was proposed to avoid the stratification problem (Spielman et al., 1993). Methods extending the TDT to allow for various situations, such as multi-allele locus (Sham & Curtis, 1995; Bickeboller & Clerget-Darpoux, 1995; Kaplan et al., 1997), missing parental data (Spielman & Ewens, 1998; Horvath & Laird, 1998; Weinberg, 1999), or genotyping errors (Gordon et al., 2001; Douglas et al., 2002), have been widely proposed. Pros and cons of the population-based and family-based association approaches have been extensively discussed in the literature (Gauderman et al., 1999; Teng & Risch, 1999; McGinnis et al., 2002; Tabor et al., 2002; Cardon & Palmer, 2003).

*Corresponding author: Dr. John Jen Tai, Division of Biostatistics, College of Public Health, National Taiwan University, Room 538, No. 17, Xu-Zhou Road, Taipei, 10055 Taiwan. Fax: +886-2-33668042. E-mail: jjtai@ntu.edu.tw

By specifying the genetic model (mode of inheritance) of a disease, Schaid & Sommer (1993) showed that the TDT is identical to the score test derived from a conditional likelihood under the additive model. Because score tests are powerful in local alternatives (Cox & Hinkley, 1974), the TDT can be regarded as an optimal test for the additive model. Analogously, for dominant and recessive models, the optimal score tests can also be derived based on the corresponding conditional likelihoods. In practice, however, since the genetic model for a complex disease under investigation is usually unknown, developing a robust test that has relatively stable power over all plausible genetic models is thus required.

For binary disease traits, Zheng et al. (2002) developed TDT-type robust association tests using case-parent triad data. When dealing with complex diseases, the phenotype of an individual is likely to be measured as a quantitative trait, such as bone mineral density used in diagnosis of bone disorders (Deng et al., 2002) and bronchial responsiveness or the numbers of eosinophils in airway tissues used in allergic asthma studies (Zhang et al., 1999). For these quantitative traits, several association tests have been proposed by researchers (Allison, 1997; Rabinowitz, 1997; Xiong et al., 1998; Abecasis et al., 2000; Monks & Kaplan, 2000; Sun et al., 2000; Liu et al., 2002; Alcais & Abel, 2003; Diao & Lin, 2006). In particular, the likelihood-based QTDT/orthogonal model proposed by Abecasis et al. (2000) and the TDT-based method proposed by Monks & Kaplan (2000) are feasible in wider family structures than other methods and are widely used (Li et al., 2008). Along with the idea of Li et al. (2008), the two tests are allele-based association tests which means that they set genotype scores according to the count of alleles and so provide evidence of association for an allele at a marker locus. Thus they are expected to have powerful performances for the additive model rather than for other models. In practical studies, when a genetic model cannot be certainly specified, we may consider the additive model as an intermediate solution. However, such a measure still cannot avoid suffering the potential loss of power in some situations. A method that is robust against the influence on testing power due to misspecification of genetic models is therefore required. In this paper, we will apply two robust procedures to establish robust candidate-gene association tests for quantitative traits using parent-offspring triad data.

We will first demonstrate the feasibility of using conditional likelihood of parent-offspring triad data to extract association information between a candidate-gene and the quantitative trait, and then derive the optimal score tests under three typical genetic models. Based on the robust procedures suggested by Gastwirth (1966, 1985); Davies (1977) and Freidlin et al. (1999), we will construct two robust statistics for assessing putative association between a candidate gene and a quantitative trait. The quantitative trait under investigation is assumed to have a distribution which belongs to the exponential family rather than a normal distribution. Statistical powers of the proposed robust tests are compared with the optimal score test under the correct model and with the score tests under incorrect models. According to the simulation results, the proposed robust tests do exhibit robustness with acceptable powers.

Methods

Notations

Suppose that in order to extract information of association between a bi-allelic candidate locus and a quantitative trait, n independent parent-offspring triad families are sampled. In each family, genotypes of the parents and their offspring at the candidate locus and the measurement of quantitative trait of the offspring are collected. Let M label the mutant allele and m the normal allele at the candidate locus. This type of data can be classified into 10 categories by the parental mating type and the offspring genotype as shown in Table 1. Let $g_p = i$ represent the parents with the ith parental mating type, $g_c = j$ the offspring with the jth genotype, y_{ijk} the measurement of the quantitative trait of the kth offspring member in the category (i, j), where $i = 1, \ldots, 6$, $j = 0, 1, 2$, $k = 1, \ldots, n_{ij}$,

Parental mating types (g_p)	Offspring genotypes (g_c)	Offspring quantitative traits (y_{ijk})	Number of triad families
$MM \times MM$ $(i = 1)$	MM $(j = 2)$	$y_{121}, \ldots, y_{12n_{12}}$	n_{12}
$MM \times Mm$ $(i = 2)$	MM $(j = 2)$	$y_{221}, \ldots, y_{22n_{22}}$	n_{22}
	Mm $(j = 1)$	$y_{211}, \ldots, y_{21n_{21}}$	n_{21}
$MM \times mm$ $(i = 3)$	Mm $(j = 1)$	$y_{311}, \ldots, y_{31n_{31}}$	n_{31}
$Mm \times Mm$ $(i = 4)$	MM $(j = 2)$	$y_{421}, \ldots, y_{42n_{42}}$	n_{42}
	Mm $(j = 1)$	$y_{411}, \ldots, y_{41n_{41}}$	n_{41}
	mm $(j = 0)$	$y_{401}, \ldots, y_{40n_{40}}$	n_{40}
$Mm \times mm$ $(i = 5)$	Mm $(j = 1)$	$y_{511}, \ldots, y_{51n_{51}}$	n_{51}
	mm $(j = 0)$	$y_{501}, \ldots, y_{50n_{50}}$	n_{50}
$mm \times mm$ $(i = 6)$	mm $(j = 0)$	$y_{601}, \ldots, y_{60n_{60}}$	n_{60}

Table 1 Classification of parent-offspring triad families according to the parental mating types and the offspring genotypes.

and n_{ij} the number of offspring members in the category (i, j). The number of triad families with parental mating type i in the sample is denoted by n_i and $n_i = \sum_{j=0}^{2} n_{ij}$.

Construction of Conditional Likelihood for Triad Families

Schaid & Sommer (1993) proposed a conditional likelihood method for test of association between a candidate gene and a disease. This method was developed to eliminate spurious association results due to population stratification and to accommodate for departure from Hardy-Weinberg equilibrium in the candidate-gene locus. In the light of their method, here we also take the conditional approach to extract association information to develop the method for analysis of quantitative traits. For an offspring member k in the category (i, j) in Table 1, the conditional probability of observing the offspring genotype $g_c = j$ given parental mating type $g_p = i$ and quantitative trait γ_{ijk} is calculated by

$$\Pr(g_c = j \mid g_p = i, \gamma_{ijk}) = \frac{\Pr(\gamma_{ijk} \mid g_p = i, g_c = j)\Pr(g_c = j \mid g_p = i)}{\sum_{j'} \Pr(\gamma_{ijk} \mid g_p = i, g_c = j')\Pr(g_c = j' \mid g_p = i)}. \quad (1)$$

Assume that if the studied candidate-gene locus is rightly the quantitative trait locus (QTL) responsible for the susceptibility of the disease, then the genotypic values of *MM*, *Mm* and *mm*, which are denoted by $u_j, j = 2, 1, 0$, respectively, should associate with the variation of the quantitative trait. Let d be the displacement effect between the genotypes *MM* and *mm*, and t the degree of dominance, then u_2 and u_1 can be expressed as

$$u_2 = u_0 + d, \quad u_1 = u_0 + td. \quad (2)$$

When $t = 0$, it represents the underlying genetic model as recessive; when $t = 0.5$, the model is additive; when $t = 1$, the model is dominant. These three typical genetic models will be taken into account in the following derivation of association tests for analysis of quantitative traits. Without loss of generality, we assume that d will be greater than 0 if the candidate-gene association is present. Therefore, to test the association between the candidate-gene and the quantitative trait is equivalent to testing whether d is equal to 0 (i.e., $u_2 = u_1 = u_0$) or not. Because the parameter t vanishes when $d = 0$, but co-exists with d when $d > 0$, t plays a role of nuisance parameter in testing the effect of d. This type of hypothesis was discussed by Davies (1977).

Assume that the distribution of the quantitative trait belongs to the exponential family, as that assumed in the study of Liu et al. (2002), then given the offspring genotype $g_c = j$ the probability density function $\Pr(\gamma_{ijk} \mid g_p = i, g_c = j)$ in (1) has the form (McCullagh & Nelder, 1989)

$$\Pr(\gamma_{ijk} | g_p = i, g_c = j) = \exp\left\{\frac{\gamma_{ijk}\theta_j - b(\theta_j)}{a(\phi)} + c(\gamma_{ijk}, \phi)\right\} \quad (3)$$

for some specific functions $a(\cdot)$, $b(\cdot)$, and $c(\cdot)$, where ϕ is the dispersion parameter and θ_j is the canonical parameter that is a function of the genotypic value u_j. The distribution defined in (3) implies that the quantitative traits of all subjects in the population have the same distribution form but with different means which depend on the genotypes of subjects. If the quantitative trait is normally distributed, then $\theta_j = u_j$, $a(\phi) = \sigma^2$, $b(\theta_j) = u_j^2/2$ and $c(\gamma_{ijk}, \phi) = -\frac{1}{2}\{\gamma_{ijk}^2/\sigma^2 + \log(2\pi\sigma^2)\}$, where u_j and σ^2 are the mean and variance of γ_{ijk} given $g_c = j$. If the quantitative trait follows a gamma distribution, then $\theta_j = -1/u_j$, $a(\phi) = 1/v$, $b(\theta_j) = \log(u_j)$ and $c(\gamma_{ijk}, \phi) = v\log(v) + (v-1)\log(\gamma_{ijk}) - \log(\Gamma(v))$, where v is the shape parameter of the gamma distribution, and the mean and variance of γ_{ijk} given $g_c = j$ are u_j and u_j^2/v, respectively. These two distributions will be assumed for evaluation of the power performances of the following proposed methods in *Simulations and Results*. Let $p_{j|i}$ denote the transmission probability $\Pr(g_c = j \mid g_p = i)$ and $g(\cdot)$ a specific link function between the canonical parameter and the genotypic value, i.e., $\theta_j = g(u_j)$ for $j = 0, 1, 2$. Because u_2 and u_1 are determined by d, t and u_0, according to (1) the conditional likelihood of the offspring members is set up as

$$L(d, t, u_0; \gamma) = \prod_{i,j}\prod_{k}^{n_{ij}} \frac{\exp\left\{\frac{\gamma_{ijk}\theta_j - b(\theta_j)}{a(\phi)} + c(\gamma_{ijk}, \phi)\right\} p_{j|i}}{\sum_{j'} \exp\left\{\frac{\gamma_{ijk}\theta_{j'} - b(\theta_{j'})}{a(\phi)} + c(\gamma_{ijk}, \phi)\right\} p_{j'|i}}. \quad (4)$$

The Score Tests for Recessive, Additive and Dominant Models

By means of regular calculation steps, in Appendix I, we derive the general form of the score statistic as

$$U_{H_0} = \frac{g'(\mu)}{a(\phi)} \sum_{i,j}\sum_{k=1}^{n_{ij}} (\gamma_{ijk} - \mu) \cdot \left(\frac{\partial u_j}{\partial d} - \sum_{j'=0}^{2} \frac{\partial u_{j'}}{\partial d} p_{j'|i}\right), \quad (5)$$

where $\mu = u_2 \Pr(MM) + u_1 \Pr(Mm) + u_0 \Pr(mm)$ is the overall population trait mean, $g'(\mu)$ is the first derivative of $g(\mu)$ with respect to μ and $\partial u_j/\partial d$ is the derivative of u_j with respect to d for $j = 0, 1, 2$ in which $\partial u_2/\partial d = 1$, $\partial u_1/\partial d = t$ and $\partial u_0/\partial d = 0$. Furthermore, the variance of U_{H_0} is

$$Var\left(U_{H_0}\right) = \left[\frac{g'(\mu)}{a(\phi)}\right]^2 \sum_{i,j}\sum_{k=1}^{n_{ij}} (\gamma_{ijk} - \mu)^2 \times Var\left(\frac{\partial u_j}{\partial d} - \sum_{j'=0}^{2} \frac{\partial u_{j'}}{\partial d} p_{j'|i}\right). \quad (6)$$

Based on (5) and (6), we obtain the score test statistic $T = \hat{U}_{H_0}[Var\left(\hat{U}_{H_0}\right)]^{-1/2}$ by substituting the sample mean $\bar{\gamma} = \frac{1}{n}\sum_{ij}\sum_{k}^{n_{ij}} \gamma_{ijk}$ for the population trait mean μ. In addition, by setting $t = 0$ (recessive), 0.5 (additive) and 1 (dominant), the corresponding score statistics for test of association under each of

the three genetic models are approximated by

$$T_{REC} = \frac{\left\{2\sum_{k=1}^{n_{22}} \frac{y_{22k}-\bar{y}}{S} - 2\sum_{k=1}^{n_{21}} \frac{y_{21k}-\bar{y}}{S} + 3\sum_{k=1}^{n_{42}} \frac{y_{42k}-\bar{y}}{S} - \sum_{k=1}^{n_{41}} \frac{y_{41k}-\bar{y}}{S} - \sum_{k=1}^{n_{40}} \frac{y_{40k}-\bar{y}}{S}\right\}}{\sqrt{4n_2+3n_4}}, \quad (7)$$

$$T_{ADD} = \frac{\left\{\sum_{k=1}^{n_{22}} \frac{y_{22k}-\bar{y}}{S} - \sum_{k=1}^{n_{21}} \frac{y_{21k}-\bar{y}}{S} + 2\sum_{k=1}^{n_{42}} \frac{y_{42k}-\bar{y}}{S} - 2\sum_{k=1}^{n_{40}} \frac{y_{40k}-\bar{y}}{S} + \sum_{k=1}^{n_{51}} \frac{y_{51k}-\bar{y}}{S} - \sum_{k=1}^{n_{50}} \frac{y_{50k}-\bar{y}}{S}\right\}}{\sqrt{n_2+2n_4+n_5}}, \quad (8)$$

and

$$T_{DOM} = \frac{\left\{\sum_{k=1}^{n_{42}} \frac{y_{42k}-\bar{y}}{S} + \sum_{k=1}^{n_{41}} \frac{y_{41k}-\bar{y}}{S} - 3\sum_{k=1}^{n_{40}} \frac{y_{40k}-\bar{y}}{S} + 2\sum_{k=1}^{n_{51}} \frac{y_{51k}-\bar{y}}{S} - 2\sum_{k=1}^{n_{50}} \frac{y_{50k}-\bar{y}}{S}\right\}}{\sqrt{3n_4+4n_5}}, \quad (9)$$

where $S^2 = \frac{1}{n}\sum_{ij}\sum_{k}^{n_{ij}} (y_{ijk}-\bar{y})^2$ is the sample variance of the quantitative trait, and it is verified that the three test statistics are all asymptotically distributed as the standard normal distribution by use of the law of large number and Slutsky's theorem (Appendix I).

Note that the T_{ADD} can also be expressed simply by (Appendix II)

$$T_{ADD} = \frac{2\left\{\sum_{i,j}\sum_{k}^{n_{ij}} \frac{y_{ijk}-\bar{y}}{S}\left[j - E\left(g_c|g_p=i\right)\right]\right\}}{\sqrt{n_2+2n_4+n_5}}.$$

This result indicates that in analysis of the candidate-gene association test with a quantitative trait the T_{ADD} detects significance of the distortion between the distributions of observed and expected offspring genotypes with respective weights $(y_{ijk}-\bar{y})/S$. In fact, all of the proposed score statistics T_{REC}, T_{ADD} and T_{DOM} are identical to the score statistics derived by Schaid and Sommer (1994) for binary traits except that here additional weights are added.

Robust Association Tests

When the underlying genetic model for the quantitative trait locus is known, a correct score test can be distinctly selected for association analysis. However, when the model is unknown, we would wonder which test is applicable. Developing a robust test that can be used in this uncertain situation is thus necessary. Here, we adopt the maximin efficiency robust procedure proposed by Gastwirth (1966) to establish such a test. Among all plausible models, this procedure calculates the asymptotic relative efficiency (ARE) of a specific test to an optimal test at first. Then, the minimum ARE of the specific test relative to all optimal tests can then be figured out. The Maximin Efficiency Robust Test (*MERT*) is defined as the test which has highest minimum ARE across all plausible models.

To apply the maximin efficiency robust procedure to our study we need first to calculate the pairwise correlations of the three score statistics. Under the null hypothesis of no association, when n is sufficiently large, the pairwise correlation coefficients of the three statistics are shown to converge to ρ_{12}, ρ_{13} and ρ_{23}, respectively, as follows (Appendix III)

$$\begin{cases} \mathrm{corr}(T_{REC}, T_{ADD}) \to \dfrac{2n_4+2n_2}{\sqrt{3n_4+4n_2}\sqrt{n_2+2n_4+n_5}} \equiv \rho_{12} \\ \mathrm{corr}(T_{REC}, T_{DOM}) \to \dfrac{n_4}{\sqrt{3n_4+4n_2}\sqrt{3n_4+4n_5}} \equiv \rho_{13} \\ \mathrm{corr}(T_{ADD}, T_{DOM}) \to \dfrac{2n_4+2n_5}{\sqrt{3n_4+4n_5}\sqrt{n_2+2n_4+n_5}} \equiv \rho_{23} \end{cases} \quad (10)$$

The asymptotic results of these correlations are identical to those obtained by Zheng et al. (2002) for binary traits. This is the case because, as aforementioned, basically the three score statistics obtained in (7), (8) and (9) have the same formulations as those for binary traits except that an extra weight is added to each offspring. Among the three correlation coefficients in (10), it can be immediately verified that ρ_{13} is the smallest one. That is, the recessive and dominant models are the extreme pair of these plausible genetic models. Therefore, based on the result of Gastwirth (1970, 1985) the *MERT* can be developed through a linear combination of two extreme test statistics, T_{REC} and T_{DOM}, provided that the necessary and sufficient condition $\rho_{12}+\rho_{23} \geq 1+\rho_{13}$ holds. The condition can be verified by a short calculation. We thus obtain the *MERT* for test of association for a quantitative trait as follows

$$MERT = \frac{T_{REC}+T_{DOM}}{\sqrt{2\,(1+\rho_{13})}},$$

which is asymptotically distributed as the standard normal distribution.

On the other hand, Davies (1977) proposed to choose the maximum of the score statistics over all plausible models as a robust statistic. Freidlin et al. (1999) and Zheng and Chen (2005) suggested that a preferable way to select a robust statistic in

practice is to simply take the maximum of the score statistics at two extreme models and an additional intermediate model. In the light of their approaches, here we consider

$$MAX = \max\{T_{REC}, T_{ADD}, T_{DOM}\}$$

for testing association between a candidate gene and the quantitative trait. In the real analysis, we have to simulate the asymptotic distribution of MAX under the null hypothesis to obtain the critical value for testing. To simulate this distribution we can assume the joint distribution of the three statistics T_{REC}, T_{ADD} and T_{DOM} have a multivariate normal distribution with mean vector 0 and a covariance matrix $\sum$ which has value 1 for the diagonal elements (i.e., the variances of the three statistics are all 1) and values for the off-diagonal elements (i.e., the covariances of the three statistics) are ρ_{12}, ρ_{13} and ρ_{23}, respectively. This assumption allows that each statistic, either T_{REC}, T_{ADD} or T_{DOM}, has an asymptotic normal distribution with mean 0 and variance 1 as discussed before. In the simulation, for a given value of allele frequency p the numbers of families with parental mating types 2, 4 and 5 (i.e., n_2, n_4 and n_5) will be generated and thus ρ_{12}, ρ_{13} and ρ_{23} are determined using (10). Based on the multivariate normal distribution with mean vector 0 and the covariance $\sum$ as mentioned above, the distribution of MAX under the null hypothesis can be simulated and the critical value corresponding to a prescribed significance level α can then be determined.

In practical situations when the potential genetic model is unknown, the parameter t (degree of dominance) could range from 0 to 1. The two robust statistics $MERT$ and MAX proposed above can be applied. However, if t can be specified in a more restricted region, these robust tests should be modified. For example, if there is prior information provided by other studies to ensure that the dominant model is fairly excluded, then a modified robust statistic $MERT_{\mathrm{RA}}$ corresponding to the $MERT$ should be used. Denote $MERT_{\mathrm{RA}} = (T_{REC} + T_{ADD})/\sqrt{2(1+\rho_{12})}$ and $MAX_{\mathrm{RA}} = \max\{T_{REC}, T_{ADD}\}$ as the robust test statistics developed by the two robust procedures in case of the dominant model being excluded, and $MERT_{\mathrm{AD}} = (T_{ADD} + T_{DOM})/\sqrt{2(1+\rho_{23})}$ and $MAX_{\mathrm{AD}} = \max\{T_{ADD}, T_{DOM}\}$ as the robust test statistics in case of the recessive model being excluded. Comparisons of robustness and power performances among the tests mentioned in this section are discussed in the following section.

Simulations and Results

Simulation Settings

To investigate the robustness and power performances of our proposed tests and other existing tests, we conducted a series of simulations under two scenarios. Two existing tests were selected, the likelihood-based QTDT/orthogonal model (OM) proposed by Abecasis et al. (2000), and the TDT-based association test (MK) proposed by Monks & Kaplan (2000). In both scenarios, three allele frequencies $p = 0.1$, 0.2, and 0.5 of the candidate gene M were evaluated and the population was assumed to be in Hardy-Weinberg equilibrium. For each allele frequency parental genotypes of a family were first generated based on the assumption of random mating, and the offspring genotype was then assigned according to the Mendelian transmission. For the first scenario, we assumed that conditional on the offspring genotype $MM(j = 2)$, $Mm(j = 1)$, or $mm(j = 0)$, the quantitative trait of the offspring was produced based on the normal distribution with a mean value of u_2, u_1 or u_0 and variance 1. By assigning $u_0 = 0$, u_2 and u_1 were determined respectively by $u_2 = d$ and $u_1 = td$ under a specified genetic model ($t = 0$ for recessive models, $t = 0.5$ for additive models and $t = 1$ for dominant models). For the given values of t and d, associated with the genotypic value u_0 and the allele frequency p, the variance of the genotypic value can be calculated. Dividing this genotype variance by the variance of the quantitative trait (here, it is assigned value 1), the corresponding heritability H^2 can then be estimated. Because power performances are diverse in different genetic models, to let the robust tests have a fair basis to be compared across the three genetic models, we selected the values of d to allow the powers of the three optimal tests T_{REC}, T_{ADD} and T_{DOM} to reach a level around 80%. For example, for allele frequency $p = 0.1$, we set $d = 2.8$ ($H^2 = 7.2\%$) for the recessive model, $d = 1.12$ ($H^2 = 5.3\%$) for the additive model, and $d = 0.6$ ($H^2 = 5.2\%$) for the dominant model as shown in Table 2. Then based on these settings, the powers of the three optimal tests are obtained with the values 0.801, 0.808 and 0.810, respectively (see Table 3).

Table 2 Simulation settings for the displacement effect d and corresponding heritability H^2 under the three frequencies of allele M and various genetic models for quantitative traits with normal or gamma distributions.

Normal distributions	Allele frequency					
	$p = 0.1$		$p = 0.2$		$p = 0.5$	
Genetic models	d	H^2 (%)	d	H^2 (%)	d	H^2 (%)
Null	0	0	0	0	0	0
Recessive	2.80	7.2	1.10	4.4	0.50	4.5
Additive	1.12	5.3	0.84	5.3	0.66	5.2
Dominant	0.60	5.2	0.48	5.0	0.50	4.5
Gamma distributions	allele frequency					
	$p = 0.1$		$p = 0.2$		$p = 0.5$	
Genetic models	d	H^2 (%)	d	H^2 (%)	d	H^2 (%)
Null	0	0	0	0	0	0
Recessive	8.5	8.2	3.2	4.7	1.5	5.0
Additive	3.2	5.4	2.4	5.4	2.0	5.9
Dominant	1.7	5.3	1.4	5.3	1.5	5.0

In the second scenario, we assumed that conditional on the offspring genotype *MM*, *Mm*, or *mm*, the quantitative trait was generated from a gamma distribution with the scale parameter $v = 1$ and the mean value u_2, u_1 or u_0. Here, u_0 was set as 8.0, and u_2 and u_1 were determined respectively by $u_2 = d$ and $u_1 = td$ given values of d and t. For the three allele frequencies $p = 0.1$, 0.2, and 0.5 of the candidate gene *M*, again, values of d and the corresponding heritability H^2 were selected such that the powers of the three optimal score tests can reach a level around 80% under the true models (Table 2).

Empirical type I error rates and powers of all tests were evaluated by 10,000 replicates at significance level 0.05. In each replicate, 300 parent-offspring families were generated. The asymptotic critical values of the *MAX* were obtained by simulating the multivariate normal distributions of (T_{REC}, T_{ADD}, T_{DOM}). Asymptotic critical values of the other tests were 1.96.

Empirical Type I Error Rates and Powers of the Proposed Tests

Table 3 summarizes the simulation results on empirical type I error rates and powers of our proposed tests and the OM and MK tests when the quantitative traits were generated from normal distributions. The empirical type I error rates of these tests are all close to the prescribed nominal significance level 0.05. Each optimal score test does outperform others in power under the model which is assumed true. For example, in Table 3 in the situation of $p = 0.2$, T_{REC} has the best power performance with a level of 0.799 when the true model is recessive. But its power drastically drops down to the level of 0.307 when the true model is additive, and down even more to the level of 0.067 when the true model is dominant. Similar results are also observed when the quantitative trait was generated from gamma distributions (Table 4). In sum, it is as expected that T_{REC} is the test having the lowest power when the true model is dominant, and vice versa; when the true model is recessive T_{DOM} is the test having the lowest power. The two robust tests *MERT* and *MAX*, in contrast, have relatively stable power performances over the three genetic models. For example, when $p = 0.1$ in Table 3, the powers of the *MERT* are all above 0.5, and the powers of the *MAX* maintain at least 0.7 in various genetic models. The *MAX* is generally more powerful than the *MERT*. The T_{ADD} also shows robustness in most situations, but its power performance is relatively low, compared with *MERT* and *MAX*, under the recessive model. Both OM and MK have comparable power performances with T_{ADD} due to the fact that they were constructed based on the additive model assumption. In general, the power of OM was slightly higher than powers of MK and T_{ADD}.

If the dominant or recessive model is excluded from the plausible genetic models, then we can use a more specific robust test $MERT_{\mathrm{RA}}$, MAX_{RA}, $MERT_{\mathrm{AD}}$ or MAX_{AD} to replace the *MERT* or *MAX* to increase the testing power. This idea is justified by the simulation results. For example, for $p = 0.5$ and under the recessive model in Table 3, the power of $MERT_{\mathrm{RA}}$ is 0.750 which is superior to the power of *MERT*

Table 3 Comparisons of empirical type I error rates and powers among three score tests, six robust tests and two popularly used tests (OM and MK) in the scenario that the quantitative trait is normally distributed and the number of triad families is 300 for three allele frequencies 0.1, 0.2, and 0.5 of the candidate gene.

True model	T_{REC}	T_{ADD}	T_{DOM}	*MERT*	*MAX*	$MERT_{\mathrm{RA}}$	$MERT_{\mathrm{AD}}$	MAX_{RA}	MAX_{AD}	*OM*†	*MK*‡
Allele frequency 0.1											
NULL	0.048	0.051	0.052	0.048	0.051	0.048	0.051	0.050	0.052	0.051	0.051
REC	0.801	0.323	0.058	0.685	0.765	0.732	–	0.771	–	0.315	0.241
ADD	0.223	0.808	0.781	0.683	0.754	0.614	0.804	0.739	0.808	0.821	0.791
DOM	0.076	0.777	0.810	0.534	0.743	–	0.803	–	0.804	0.789	0.765
Allele frequency 0.2											
NULL	0.048	0.051	0.050	0.051	0.051	0.050	0.050	0.050	0.049	0.050	0.051
REC	0.799	0.327	0.061	0.592	0.725	0.701	–	0.746	–	0.322	0.304
ADD	0.307	0.807	0.745	0.762	0.757	0.680	0.795	0.750	0.801	0.822	0.805
DOM	0.067	0.733	0.804	0.543	0.732	–	0.786	–	0.787	0.740	0.733
Allele frequency 0.5											
NULL	0.049	0.052	0.046	0.051	0.050	0.049	0.050	0.052	0.049	0.051	0.051
REC	0.805	0.565	0.067	0.564	0.722	0.750	–	0.766	–	0.572	0.566
ADD	0.563	0.802	0.565	0.802	0.745	0.750	0.753	0.767	0.767	0.814	0.804
DOM	0.063	0.562	0.802	0.562	0.721	–	0.751	–	0.765	0.571	0.563

†the likelihood-based QTDT/orthogonal model proposed by Abecasis et al. (2000).
‡the TDT-based association test proposed by Monks & Kaplan (2000).

Table 4 Comparisons of empirical type I error rates and powers among three score tests, six robust tests and two popularly used tests (OM and MK) in the scenario that the quantitative trait follows a gamma distribution and the number of triad families is 300 for three allele frequencies 0.1, 0.2, and 0.5 of the candidate gene.

True model	T_{REC}	T_{ADD}	T_{DOM}	*MERT*	*MAX*	$MERT_{\mathrm{RA}}$	$MERT_{\mathrm{AD}}$	MAX_{RA}	MAX_{AD}	*OM*†	*MK*‡
Allele frequency 0.1											
NULL	0.054	0.050	0.050	0.051	0.053	0.051	0.049	0.053	0.050	0.050	0.048
REC	0.801	0.349	0.060	0.694	0.769	0.738	–	0.774	–	0.343	0.246
ADD	0.225	0.794	0.771	0.660	0.744	0.585	0.793	0.727	0.796	0.810	0.758
DOM	0.090	0.772	0.804	0.533	0.738	–	0.796	–	0.797	0.782	0.737
Allele frequency 0.2											
NULL	0.054	0.049	0.047	0.052	0.050	0.051	0.049	0.051	0.049	0.049	0.050
REC	0.791	0.339	0.060	0.596	0.724	0.697	–	0.740	–	0.338	0.310
ADD	0.300	0.792	0.725	0.734	0.739	0.650	0.779	0.726	0.784	0.804	0.777
DOM	0.077	0.747	0.808	0.556	0.739	–	0.795	–	0.797	0.754	0.731
Allele frequency 0.5											
NULL	0.050	0.051	0.048	0.051	0.052	0.050	0.051	0.053	0.053	0.052	0.051
REC	0.817	0.591	0.064	0.593	0.743	0.771	–	0.785	–	0.597	0.593
ADD	0.556	0.800	0.561	0.800	0.740	0.739	0.749	0.761	0.765	0.811	0.802
DOM	0.068	0.554	0.801	0.552	0.718	–	0.745	–	0.763	0.560	0.557

†the likelihood-based QTDT/orthogonal model proposed by Abecasis et al. (2000)
‡the TDT-based association test proposed by Monks & Kaplan (2000)

0.564. This shows that if the dominant model can be excluded from the plausible models, use of the $MERT_{\mathrm{RA}}$ for test of association would acquire higher power than use of the *MERT*. Similarly, the MAX_{AD} is overall more powerful than the *MAX* when the recessive model can be excluded.

Power Performances of the Proposed Tests Over the Whole Range of Degree of Dominance

To investigate the characteristic of power performance of each proposed test over the range of degree of dominance t, we considered two situations of $p = 0.2$ and 0.5. For each situation, values of t in the whole range [0, 1] with a step size 0.1 were investigated. The displacement effect d was set subject to the assumption that the value of heritability was around 5% for each fixed value of t.

In Figure 1, it is obviously shown that the test statistic T_{REC} or T_{DOM} loses its power rapidly if the presumed value of t is far from the true value (i.e., the true genetic model). The T_{ADD} loses substantial power to detect a putative association if the value of t is close to 0 (i.e., the recessive model). In contrast, the *MERT* and *MAX* have more stable power performances over the whole range of t, especially the *MAX*. The result for $p = 0.5$ is similar to that for $p = 0.2$ and it shows a symmetric pattern as shown in Figure 2. In particular, the power curve of the *MERT* coincides with the curve of the T_{ADD}. Further comments on this result are given in the *Discussion* section.

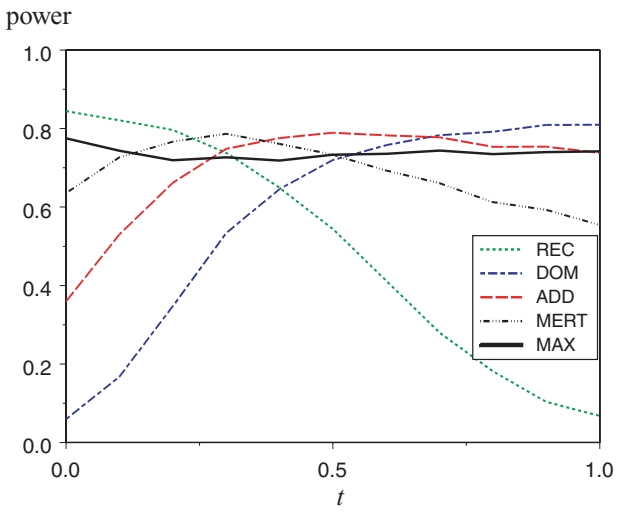

Figure 1 Power performances of the T_{REC}, T_{ADD}, T_{DOM}, *MERT* and *MAX* over the range of degree of dominance t from 0 to 1 by controlling heritability around 5% for each genetic model when $p = 0.2$.

Discussion

Most of the family-based association tests may encounter the problem of loss of power when the underlying genetic model of the studied disease is not commensurate with that by which a test is developed. Association tests that are developed based

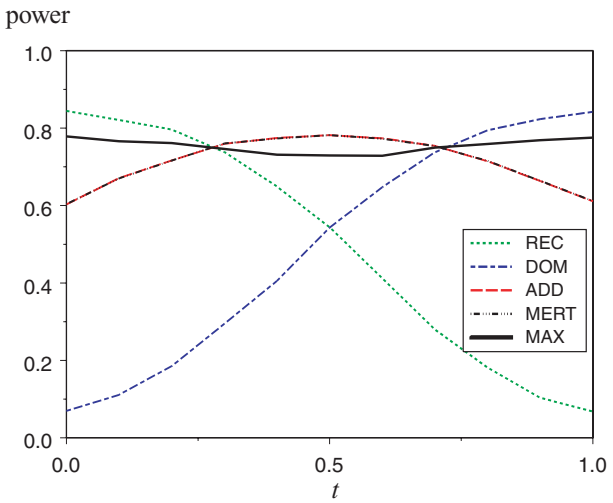

Figure 2 Power performances of the T_{REC}, T_{ADD}, T_{DOM}, *MERT* and *MAX* over the range of degree of dominance t from 0 to 1 by controlling heritability around 5% for each genetic model when $p = 0.5$.

on allele-counting methods, such as the TDT, can essentially be regarded as derived under the additive model, and would thus have better power performance under the additive model than under others. However, these methods could lose substantial powers if the additive model is not true. For example, the TDT is powerless under recessive models if the frequency of the mutant allele is low either for binary disease traits (Camp, 1997; Lange & Laird, 2002) or quantitative traits (Lange et al., 2002; Purcell et al., 2003). Ideally, if it were possible, we would like to specify an appropriate genetic model for the quantitative trait so that an optimal test can be used for test of association. However, in realistic studies it is usually problematic to do so, and such a method may suffer from the risk of model misspecification. Our simulation results show that once the genetic model is misspecified, a test could critically lose its power in detection of any association candidate gene. Hence, development of robust association tests that are applicable in practice deserves more attention. Issues on robustness have been popularly addressed in statistical inference when there exist nuisance parameters that interfere with the powers of the studied tests. In genetic studies, this issue has also been investigated widely by researchers (Wang et al., 1999; Gastwirth & Freidlin, 2000; Kraft, 2001; Zheng et al., 2002; Xu et al., 2003; Diao & Lin, 2005; Tai & Hou, 2006; Yan et al., 2008; Wang et al., 2008). To obtain robust statistics over all plausible genetic models, we have referred to the studies of Davies (1977) and Zheng & Chen (2005). According to the study of Davies (1977) we should take the maximum score statistic over the whole range of a nuisance parameter as the robust statistic. Nevertheless, by the study of Zheng & Chen (2005) a maximum score statistic at two most extreme models and an intermediate model almost performs as well as the maximum statistic proposed by Davies (1977). Based on their studies, we thus conduct the robust statistic *MAX* in our study by taking the maximum of the three score statistics under the recessive, additive and dominant models only. In addition, we also use the maximin efficiency robust procedure (Gastwirth, 1966, 1985) to establish the other robust statistic *MERT* for an alternative choice beyond the *MAX*. This statistic can be regarded as a combined statistic of the two score statistics under the recessive and dominant models. The robustness of the test comes from the compromise between the two extreme models on powers.

Simulation results show that the two proposed robust tests have relatively stable and acceptable powers under various genetic models for test of association between a candidate gene and a quantitative trait. While the *MERT* is somewhat easier in computation, its power performance is overall slightly lower than that of *MAX*. Note that the association tests T_{ADD}, OM and MK, which are constructed based on additive model assumption, also show robustness in most situations. In Figure 2, it can be seen that the power curve of T_{ADD} coincides with the curve of *MERT*. This may be due to the fact that T_{ADD} is the optimal test derived at the middle point $t = 0.5$ of the degree of dominance. Note that when $p < 0.5$, the correlation of (T_{REC}, T_{ADD}) is lower than that of (T_{ADD}, T_{DOM}). On the contrary, when $p > 0.5$ the correlation of (T_{REC}, T_{ADD}) is higher than that of (T_{ADD}, T_{DOM}). It is only when $p = 0.5$ that the correlation coefficients of (T_{REC}, T_{ADD}) and (T_{ADD}, T_{DOM}) are equal. Therefore, it is in the condition when the true model is additive and in particular when the gene frequency $p = 0.5$, the *MERT* and T_{ADD} would almost have identical power performances. However, this issue still needs more investigation.

Looking into the estimated powers in Tables 3 and 4, it shows that T_{ADD}, OM and MK perform comparable to, or even better than, *MAX* or *MERT* when the true model is additive or dominant. The merit of *MAX* and *MERT* is their stability in power performances. However, their advantage is almost under the recessive model only. In the light of these results, it indicates that association tests based on additive model assumption are still a valid method even though the underlying genetic model may not be additive. In other words, in practical studies researchers are encouraged to calculate all three model-specific score tests. If the test statistic based on additive model assumption is distinctly low, the robust tests proposed here may be taken into account as a useful tool for validation of the results.

For analyzing a quantitative trait, parent-offspring triad data may be collected through an ascertainment procedure other than random sampling according to prescribed thresholds, which are determined by the upper and/or lower tails of the offspring trait distribution. As mentioned by Liu et al. (2002),

under these circumstances, an estimate of the population trait mean can be obtained by adjusting the sample mean with an offset term, which is available by a *priori* knowledge about the central tendency of the quantitative trait distribution. Note that the proposed tests are still valid if the estimation of population trait mean is biased. However, a biased estimation will cause negative effects on the powers of these tests. The robust methods proposed in this paper have focused on the triad families; future study on the extension of these to wider family data is under investigation. In addition, the construction of a robust association test based on linear or nonlinear models, so that some other covariates of interest could be involved as well, may also deserve further efforts.

Acknowledgments

The authors would like to thank the editor and two anonymous reviewers for their constructive comments that greatly improved this paper. This work was supported in part by grants AU-97-H-03 (JYW) from Asia University and NSC-96-2628-M-002–025-MY3 (JJT) from National Science Council, Taiwan.

References

Abecasis, G. R., Cardon, L. R. & Cookson, W. O. C. (2000) A general test of association for quantitative traits in nuclear families. *Am J Hum Genet* **66**, 279–292.

Alcais, A. & Abel, L. (2003) Removing phenotypic distribution assumptions from tests of linkage disequilibrium for quantitative traits. *Genet Epidemiol* **24**, 191–199.

Allison, D. B. (1997) Transmission-disequilibrium tests for quantitative traits. *Am J Hum Genet* **60**, 676–690.

Bickeboller, H. & Clerget-Darpoux, F. (1995) Statistical properties of the allelic and genotypic transmission/disequilibrium test for multiallelic markers. *Genet Epidemiol* **12**, 865–870.

Camp, N. J. (1997) Genomewide transmission/disequilibrium testing – Consideration of the genotypic relative risks at disease loci. *Am J Hum Genet* **61**, 1424–1430.

Cardon, L. R. & Palmer, L. J. (2003) Population stratification and spurious allelic association. *Lancet* **361**, 598–604.

Cordell, H. J. & Clayton, D. G. (2005) Genetic epidemiology 3 - Genetic association studies. *Lancet* **366**, 1121–1131.

Cox, D. R. & Hinkley, D. V. (1974) *Theoretical Statistics*. London: Chapman and Hall.

Davies, R. B. (1977) Hypothesis testing when a nuisance parameter is present only under alternative. *Biometrika* **64**, 247–254.

Deng, H. W., Livshits, G., Yakovenko, K., Xu, F. H., Conway, T., Davies, K. M., Deng, H. & Recker, R. R. (2002) Evidence for a major gene for bone mineral density/content in human pedigrees identified via probands with extreme bone mineral density. *Ann Hum Genet* **66**, 61–74.

Devlin, B. & Roeder, K. (1999) Genomic control for association studies. *Biometrics* **55**, 997–1004.

Diao, G. & Lin, D. Y. (2005) A powerful and robust method for mapping quantitative trait loci in general pedigrees. *Am J Hum Genet* **77**, 97–111.

Diao, G. & Lin, D. Y. (2006) Improving the power of association tests for quantitative traits in family studies. *Genet Epidemiol* **30**, 301–313.

Douglas, J. A., Skol, A. D. & Boehnke, M. (2002) Probability of detection of genotyping errors and mutations as inheritance inconsistencies in nuclear-family data. *Am J Hum Genet* **70**, 487–495.

Falk, C. T. & Rubinstein, P. (1987) Haplotype relative risks - an easy reliable way to construct a proper control sample for risk calculations. *Ann Hum Genet* **51**, 227–233.

Freidlin, B., Podgor, M. J. & Gastwirth, J. L. (1999) Efficiency robust tests for survival or ordered categorical data. *Biometrics* **55**, 883–886.

Gastwirth, J. L. (1966) On the robust procedures. *J Am Statist Assoc* **61**, 929–948.

Gastwirth, J. L. (1970) On robust rank tests. In: *Nonparametric techniques in statistical inference*. M. L. Puri (ed.) London: Cambridge University Press.

Gastwirth, J. L. (1985) The use of maximin efficiency robust-tests in combining contingency tables and survival analysis. *J Am Statist Assoc* **80**, 380–384.

Gastwirth, J. L. & Freidlin, B. (2000) On power and efficiency robust linkage tests for affected sibs. *Ann Hum Genet* **64**, 443–453.

Gauderman, W. J., Witte, J. S. & Thomas, D. C. (1999) Family-based association studies. *J Natl Cancer Inst Monogr* **26**, 31–37.

Gordon, D., Heath, S. C., Liu, X. & Ott, J. (2001) A transmission disequilibrium test that allows for genotyping errors in the analysis of single nucleotide polymorphism data. *Am J Hum Genet* **69**, 371–380.

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.

Kaplan, N. L., Martin, E. R. & Weir, B. S. (1997) Power studies for the transmission/disequilibrium tests with multiple alleles. *Am J Hum Genet* **60**, 691–702.

Kraft, P. (2001) A robust score test for linkage disequilibrium in general pedigrees. *Genet Epidemiol* **21**, S447–S452.

Kruglyak, L. & Lander, E. S. (1995) High-resolution genetic-mapping or complex traits. *Am J Hum Genet* **56**, 1212–1223.

Lange, C., Demeo, D. L. & Laird, N. M. (2002) Power and design considerations for a general class of family-based association tests: Quantitative traits. *Am J Hum Genet* **71**, 1330–1341.

Lange, C. & Laird, N. M. (2002) Power calculations for a general class of family-based association tests: Dichotomous traits. *Am J Hum Genet* **71**, 575–584.

Li, Y. W., Martin, E. R. & Li, Y. J. (2008) EMK: A novel program for family-based allelic and genotypic association tests on quantitative traits. *Ann Hum Genet* **72**, 388–396.

Liu, Y., Tritchler, D. & Bull, S. B. (2002) A unified framework for transmission-disequilibrium test analysis of discrete and continuous traits. *Genet Epidemiol* **22**, 26–40.

McCullagh, P. & Nelder, J. A. (1989) *Generalized Linear Models.* London: Chapman and Hall.

McGinnis, R., Shifman, S. & Darvasi, A. (2002) Power and efficiency of the TDT and case-control design for association scans. *Behav Genet* **32**, 135–144.

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.

Pritchard, J. K. & Rosenberg, N. A. (1999) Use of unlinked genetic markers to detect population stratification in association studies. *Am J Hum Genet* **65**, 220–228.
Purcell, S., Cherny, S. S. & Sham, P. C. (2003) Genetic Power Calculator: design of linkage and association genetic mapping studies of complex traits. *Bioinformatics* **19**, 149–150.
Rabinowitz, D. (1997) A transmission disequilibrium test for quantitative trait loci. *Hum Hered* **47**, 342–350.
Reich, D. E. & Goldstein, D. B. (2001) Detecting association in a case-control study while correcting for population stratification. *Genet Epidemiol* **20**, 4–16.
Risch, N. & Merikangas, K. (1996) The future of genetic studies of complex human diseases. *Science* **273**, 1516–1517.
Schaid, D. J. & Sommer, S. S. (1993) Genotype relative risks – methods for design and analysis of candidate-gene association studies. *Am J Hum Genet* **53**, 1114–1126.
Schaid, D. J. & Sommer, S. S. (1994) Comparison of statistics for candidate-gene association studies using cases and parents. *Am J Hum Genet* **55**, 402–409.
Schork, N. J., Fallin, D., Thiel, B., Xu, X., Broeckel, U., Jacob, H. J. & Cohen, D. (2001) The future of genetic case-control studies. *Adv Genet* **42**, 191–212.
Sham, P. C. & Curtis, D. (1995) An extended transmission/disequilibrium test (TDT) for multi-allele marker loci. *Ann Hum Genet* **59**, 323–336.
Spielman & Ewens (1998) A sibship test for linkage in the presence of association: The sib transmission/disequilibrium test. *Am J Hum Genet* **62**, 450–458.
Spielman, R. S., McGinnis, R. E. & Ewens, W. J. (1993) Transmission test for linkage disequilibrium – the insulin gene region and insulin-dependent diabetes-mellitus (IDDM). *Am J Hum Genet* **52**, 506–516.
Sun, F. Z., Flanders, W. D., Yang, Q. H. & Zhao, H. Y. (2000) Transmission/disequilibrium tests for quantitative traits. *Ann Hum Genet* **64**, 555–565.
Tabor, H. K., Risch, N. J. & Myers, R. M. (2002) Candidate-gene approaches for studying complex genetic traits: practical considerations. *Nature Reviews Genetics* **3**, 391–A396.
Tai, J. J. & Hou, C. D. (2006) On the combination of transmission/disequilibrium test and mean test for linkage detection using affected sib pairs. *Computational Statistics & Data Analysis* **50**, 1072–1089.
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 Research* **9**, 234–241.
Wang, J. Y., Hou, C. D. & Tai, J. J. (2008) A robust linkage analysis method using combined allele sharing and transmission disequilibrium information from case-parent tetrad families. *Ann Hum Genet* **72**, 575–587.
Wang, J., Guerra, R. & Cohen, J. (1999) A statistically robust variance-components approach for quantitative trait linkage analysis. *Ann Hum Genet* **63**, 249–262.
Weinberg, C. R. (1999) Allowing for missing parents in genetic studies of case-parent 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* **62**, 431–452.
Xu, X., Tian, L. & Wei, L. J. (2003) Combining dependent tests for linkage or association across multiple phenotypic traits. *Biostatistics* **4**, 223–229.
Yan, L. K., Zheng, G. & Li, Z. (2008) Two-stage group sequential robust tests in family-based association studies: controlling type I error. *Ann Hum Genet* 557–565.
Zhang, Y. M., Lefort, J., Kearsey, V., Silva, J. R. L. E., Cookson, W. O. C. M. & Vargaftig, B. B. (1999) A genome-wide screen for asthma-associated quantitative trait loci in a mouse model of allergic asthma. *Hum Mol Genet* **8**, 601–605.
Zheng, G. & Chen, Z. H. (2005) Comparison of maximum statistics for hypothesis testing when a nuisance parameter is present only under the alternative. *Biometrics* **61**, 254–258.
Zheng, G., Freidlin, B. & Gastwirth, J. L. (2002) Robust TDT-type candidate-gene association tests. *Ann Hum Genet* **66**, 145–155.

Appendix I

Derivation of the Score Statistics Under Recessive, Additive and Dominant Models

For the exponential family setting in equation (3), The canonical parameter θ_j is a function of the genotypic value u_j. Given an offspring genotype j, the quantitative trait y_{ijk} of an offspring follows a distribution with mean u_j as that defined in equation (2). Based on the data structure in Table 1, taking the natural logarithm of the conditional likelihood in equation (4), we obtain

$$\ell = \sum_{i,j} \ell_{ij}(\theta_j) \propto \sum_{i,j} \sum_{k=1}^{n_{ij}} \left\{ \frac{y_{ijk}\theta_j - b(\theta_j)}{a(\phi)} + c(y_{ijk}, \phi) \right\} - \sum_{i,j} \sum_{k=1}^{n_{ij}} \log \left[\sum_{j'=0}^{2} f(y_{ijk}; \theta_{j'}) p_{j'|i} \right], \quad \text{(A1)}$$

where $f(y_{ijk}; \theta_{j'}) = \Pr(y_{ijk} | g_p = i, g_c = j') = \exp\{(y_{ijk}\theta_{j'} - b(\theta_{j'}))/a(\phi) + c(y_{ijk}, \phi)\}$. The score statistic becomes

$$U = \frac{\partial \ell}{\partial d} = \frac{1}{a(\phi)} \left\{ \sum_{i,j} \sum_{k=1}^{n_{ij}} \left(y_{ijk} - b'(\theta_j) \right) \left(\frac{\partial \theta_j}{\partial d} \right) \right\} - \sum_{i,j} \sum_{k=1}^{n_{ij}} \frac{\sum_{j'=0}^{2} \left[\left(\frac{y_{ijk} - b'(\theta_{j'})}{a(\phi)} \right) \left(\frac{\partial \theta_{j'}}{\partial d} \right) f(y_{ijk}; \theta_{j'}) p_{j'|i} \right]}{\sum_{j'=0}^{2} f(y_{ijk}; \theta_{j'}) p_{j'|i}}, \quad \text{(A2)}$$

where $b'(\theta_j)$ is the first derivative of $b(\theta_j)$ with respect to θ_j. Suppose that θ_j and u_j are linked through a function $g(\cdot)$, i.e., $\theta_j = g(u_j)$. For $j = 0$, 1 and 2, $\partial\theta_j/\partial d = g'(u_j)(\partial u_j/\partial d)$, where $g'(u_j)$ is the first derivative of $g(u_j)$ with respect to u_j, $\partial u_2/\partial d = 1$, $\partial u_1/\partial d = t$ and $\partial u_0/\partial d = 0$. Under the null hypothesis $H_0 : d = 0$, $u_2 = u_1 = u_0 = u_\bullet$ and $\theta_2 = \theta_1 = \theta_0 =$

$\theta_\bullet$ such that the overall population trait mean μ becomes $\mu = E(y) = u_2 \Pr(MM) + u_1 \Pr(Mm) + u_0 \Pr(mm) = u_\bullet$ and thus $g'(u_j) = g'(u_\bullet) = g'(\mu)$ for $j = 0, 1, 2$. In addition, since it is known that $b'(\theta_j) = u_j$ (McCullagh & Nelder, 1989), here we have $b'(\theta_j) = u_j = u_\bullet = \mu$ for all j under H_0. Therefore, the score statistic in (A2) under H_0 turns out to be

$$
\begin{aligned}
U_{H_0} &= \frac{1}{a(\phi)} \sum_{i,j} \sum_{k=1}^{n_{ij}} (y_{ijk} - \mu) \cdot g'(\mu) D_j \\
&\quad - \sum_{i,j} \sum_{k=1}^{n_{ij}} \frac{\sum_{j'=0}^{2} \left[(y_{ijk} - \mu) / a(\phi) \right] g'(\mu) D_{j'} f(y_{ijk}; \theta_\bullet) p_{j'|i}}{f(y_{ijk}; \theta_\bullet)} \\
&= \frac{g'(\mu)}{a(\phi)} \sum_{i,j} \sum_{k=1}^{n_{ij}} (y_{ijk} - \mu) \cdot D_j \\
&\quad - \frac{g'(\mu)}{a(\phi)} \sum_{i,j} \sum_{k=1}^{n_{ij}} \sum_{j'=0}^{2} (y_{ijk} - \mu) \cdot D_{j'} p_{j'|i} \\
&= \frac{g'(\mu)}{a(\phi)} \sum_{i,j} \sum_{k=1}^{n_{ij}} (y_{ijk} - \mu) \cdot \{ D_j - E_i(D) \}, \qquad \text{(A3)}
\end{aligned}
$$

where $D_j = \partial u_j / \partial d$ and $E_i(D) = \sum_{j'=0}^{2} D_{j'} p_{j'|i}$ is the conditional expectation of D_j given the parental mating type i. Note that in the composition of U_{H_0}, the y_{ijk}-part is regarded as fixed and the D_j-part as random which varies according to the corresponding offspring genotype. The information of the parameter of interest (viz., parameter d) contains in the D_j-part. The y_{ijk}-part plays a role of weighing for D_j-part in U_{H_0}. Such an idea for the composition of U_{H_0} had been adopted by other researchers (Rabinowitz, 1997; Sun et al., 2000; Liu et al., 2002). Based on this idea and the assumption that the offspring genotypes in all triad families are independent, the variance of U_{H_0} is obtained as

$$
Var\left(U_{H_0}\right) = \left[\frac{g'(\mu)}{a(\phi)} \right]^2 \sum_{i,j} \sum_{k=1}^{n_{ij}} (y_{ijk} - \mu)^2 Var_i(D), \qquad \text{(A4)}
$$

where $Var_i(D)$ is the conditional variance of D given the parental mating type i. Values of $E_i(D)$ and $Var_i(D)$ for each parental mating type i under the recessive, additive and dominant models, additionally with the transmission probabilities $p_{j|i}$ are shown in Table A1.

Using the results in (A3) and (A4) the score statistic for test of $H_0 : d = 0$ is constructed as

$$
\begin{aligned}
T^* &= \hat{U}_{H_0} \left[Var\left(\hat{U}_{H_0}\right) \right]^{-1/2} \\
&= \frac{\sum_{i,j} \sum_{k=1}^{n_{ij}} (y_{ijk} - \bar{y}) \left\{ D_j - E_i(D) \right\}}{\sqrt{\sum_{i,j} \sum_{k=1}^{n_{ij}} (y_{ijk} - \bar{y})^2 Var_i(D)}}, \qquad \text{(A5)}
\end{aligned}
$$

in which the sample mean $\bar{y} = \frac{1}{n} \sum_{ij} \sum_{k}^{n_{ij}} y_{ijk}$ are taken as an estimate of the population trait mean μ. Since D_j takes a value of 0, t and 1 for each offspring, to implement the score test statistic, the parameter t needs to be determined according to the assigned genetic model. Consider the following three situations:

(1). When the genetic model is recessive, $t = 0$, $D_2 = 1$, $D_1 = D_0 = 0$; given each of the parental mating types, $E_1(D) = 1$, $E_2(D) = 1/2$, $E_4(D) = 1/4$ and $E_i(D) = 0$ for $i = 3, 5, 6$. Besides, the conditional variances of D_j are $Var_2(D) = 1/4$, $Var_4(D) = 3/16$ and $Var_i(D) = 0$ for $i = 1, 3, 5, 6$ (as shown in Table A1). Thus the score statistic in (A5) becomes

Table A1 The transmission probabilities and the conditional expectations and variances of D_j given each of the parental mating types under the recessive, additive and dominant models.

	$p_{j\|i}$			*REC*		*ADD*		*DOM*	
$i^\dagger$	$j=2$	$j=1$	$j=0$	$E_i(D)$	$Var_i(D)$	$E_i(D)$	$Var_i(D)$	$E_i(D)$	$Var_i(D)$
1	1	0	0	1	0	1	0	1	0
2	1/2	1/2	0	1/2	1/4	3/4	1/16	1	0
3	0	1	0	0	0	1/2	0	1	0
4	1/4	1/2	1/4	1/4	3/16	1/2	1/8	3/4	3/16
5	0	1/2	1/2	0	0	1/4	1/16	1/2	1/4
6	0	0	1	0	0	0	0	0	0

†The six parental mating types are defined in Table 1.

$$
\begin{aligned}
&T^*_{REC}\\
&= \left(\frac{1}{4}\sum_j\sum_{k=1}^{n_{2j}}\left(y_{2jk}-\bar{y}\right)^2+\frac{3}{16}\sum_j\sum_{k=1}^{n_{4j}}\left(y_{4jk}-\bar{y}\right)^2\right)^{-1/2}\cdot\\
&\quad\left\{\sum_{k=1}^{n_{22}}(y_{22k}-\bar{y})\left(1-\frac{1}{2}\right)+\sum_{k=1}^{n_{21}}(y_{21k}-\bar{y})\left(0-\frac{1}{2}\right)\right.\\
&\quad+\sum_{k=1}^{n_{42}}(y_{42k}-\bar{y})\left(1-\frac{1}{4}\right)+\sum_{k=1}^{n_{41}}(y_{41k}-\bar{y})\left(0-\frac{1}{4}\right)\\
&\quad\left.+\sum_{k=1}^{n_{40}}(y_{40k}-\bar{y})\left(0-\frac{1}{4}\right)\right\},
\end{aligned}
\tag{A6}
$$

$$
\begin{aligned}
&+\sum_{k=1}^{n_{21}}(y_{21k}-\bar{y})\left(\frac{1}{2}-\frac{3}{4}\right)+\sum_{k=1}^{n_{42}}(y_{42k}-\bar{y})\left(1-\frac{1}{2}\right)\\
&+\sum_{k=1}^{n_{41}}(y_{41k}-\bar{y})\left(\frac{1}{2}-\frac{1}{2}\right)+\sum_{k=1}^{n_{40}}(y_{40k}-\bar{y})\left(0-\frac{1}{2}\right)\\
&+\sum_{k=1}^{n_{51}}(y_{51k}-\bar{y})\left(\frac{1}{2}-\frac{1}{4}\right)+\sum_{k=1}^{n_{50}}(y_{50k}-\bar{y})\left(0-\frac{1}{4}\right)\Bigg\},
\end{aligned}
$$

which can be approximated by

$$
\begin{aligned}
T_{ADD} &= \frac{\frac{1}{4}\sum_{k=1}^{n_{22}}(y_{22k}-\bar{y})-\frac{1}{4}\sum_{k=1}^{n_{21}}(y_{21k}-\bar{y})+\frac{1}{2}\sum_{k=1}^{n_{42}}(y_{42k}-\bar{y})-\frac{1}{2}\sum_{k=1}^{n_{40}}(y_{40k}-\bar{y})+\frac{1}{4}\sum_{k=1}^{n_{51}}(y_{51k}-\bar{y})-\frac{1}{4}\sum_{k=1}^{n_{50}}(y_{50k}-\bar{y})}{S\sqrt{\left(n_2/16+n_4/8+n_5/16\right)}}\\
&= \frac{\left\{\sum_{k=1}^{n_{22}}\frac{y_{22k}-\bar{y}}{S}-\sum_{k=1}^{n_{21}}\frac{y_{21k}-\bar{y}}{S}+2\sum_{k=1}^{n_{42}}\frac{y_{42k}-\bar{y}}{S}-2\sum_{k=1}^{n_{40}}\frac{y_{40k}-\bar{y}}{S}+\sum_{k=1}^{n_{51}}\frac{y_{51k}-\bar{y}}{S}-\sum_{k=1}^{n_{50}}\frac{y_{50k}-\bar{y}}{S}\right\}}{\sqrt{n_2+2n_4+n_5}}.
\end{aligned}
\tag{A7}
$$

which is asymptotically distributed as the standard normal distribution. Because under the null hypothesis $\frac{1}{n_{ij}}\sum_k^{n_{ij}}(y_{ijk}-\bar{y})^2$ converges to $Var(y)$ in probability, by the law of large number theorem, provided that n_{ij} is sufficiently large. Thus, the first term in (A6) will converge to $(\frac{1}{4}\sum_j n_{2j}\cdot Var(y)+\frac{3}{16}\sum_j n_{4j}\cdot Var(y))^{-1/2}$ when n is sufficiently large, where $Var(y)$ can be estimated by the sample variance $S^2=\frac{1}{n}\sum_{ij}\sum_k^{n_{ij}}(y_{ijk}-\bar{y})^2$. Consequently, the score test statistic for the recessive model is approximated by

The asymptotic distribution of T_{ADD} is also the standard normal distribution.

(3) When the genetic model is dominant, $t=1$, $D_2=1$, $D_1=D_0=0$; according to the results in Table A1, the score test statistic is obtained as

$$
\begin{aligned}
T_{REC} &= \frac{\frac{1}{2}\sum_{k=1}^{n_{22}}(y_{22k}-\bar{y})-\frac{1}{2}\sum_{k=1}^{n_{21}}(y_{21k}-\bar{y})+\frac{3}{4}\sum_{k=1}^{n_{42}}(y_{42k}-\bar{y})-\frac{1}{4}\sum_{k=1}^{n_{41}}(y_{41k}-\bar{y})-\frac{1}{4}\sum_{k=1}^{n_{40}}(y_{40k}-\bar{y})}{S\sqrt{\left(n_2/4+3n_4/16\right)}}\\
&= \frac{\left\{2\sum_{k=1}^{n_{22}}\frac{y_{22k}-\bar{y}}{S}-2\sum_{k=1}^{n_{21}}\frac{y_{21k}-\bar{y}}{S}+3\sum_{k=1}^{n_{42}}\frac{y_{42k}-\bar{y}}{S}-\sum_{k=1}^{n_{41}}\frac{y_{41k}-\bar{y}}{S}-\sum_{k=1}^{n_{40}}\frac{y_{40k}-\bar{y}}{S}\right\}}{\sqrt{4n_2+3n_4}}.
\end{aligned}
$$

According to Slutsky's theorem, the asymptotic distribution of T_{REC} is the standard normal distribution.

(2) When the genetic model is additive, $t=1/2$, $D_2=1$, $D_1=1/2$, $D_0=0$; based on Table A1, the score test statistic is calculated by

$$
\begin{aligned}
T^*_{ADD} &= \left(\frac{1}{16}\sum_j\sum_{k=1}^{n_{2j}}\left(y_{2jk}-\bar{y}\right)^2+\frac{1}{8}\sum_j\sum_{k=1}^{n_{4j}}\left(y_{4jk}-\bar{y}\right)^2\right.\\
&\quad\left.+\frac{1}{16}\sum_j\sum_{k=1}^{n_{5j}}\left(y_{5jk}-\bar{y}\right)^2\right)^{-1/2}\cdot\left\{\sum_{k=1}^{n_{22}}(y_{22k}-\bar{y})\left(1-\frac{3}{4}\right)\right.
\end{aligned}
$$

$$
\begin{aligned}
&T^*_{DOM}\\
&= \left(\frac{3}{16}\sum_j\sum_{k=1}^{n_{4j}}\left(y_{4jk}-\bar{y}\right)^2+\frac{1}{4}\sum_j\sum_{k=1}^{n_{5j}}\left(y_{5jk}-\bar{y}\right)^2\right)^{-1/2}\cdot\\
&\quad\times\left\{\sum_{k=1}^{n_{42}}(y_{42k}-\bar{y})\left(1-\frac{3}{4}\right)+\sum_{k=1}^{n_{41}}(y_{41k}-\bar{y})\left(1-\frac{3}{4}\right)\right.\\
&\quad+\sum_{k=1}^{n_{40}}(y_{40k}-\bar{y})\left(0-\frac{3}{4}\right)+\sum_{k=1}^{n_{51}}(y_{51k}-\bar{y})\left(1-\frac{1}{2}\right)\\
&\quad\left.+\sum_{k=1}^{n_{50}}(y_{50k}-\bar{y})\left(0-\frac{1}{2}\right)\right\},
\end{aligned}
$$

and is approximated by

$$T_{DOM} = \frac{\frac{1}{4}\sum_{k=1}^{n_{42}}(y_{42k}-\bar{y}) + \frac{1}{4}\sum_{k=1}^{n_{41}}(y_{41k}-\bar{y}) - \frac{3}{4}\sum_{k=1}^{n_{40}}(y_{40k}-\bar{y}) + \frac{1}{2}\sum_{k=1}^{n_{51}}(y_{51k}-\bar{y}) - \frac{1}{2}\sum_{k=1}^{n_{50}}(y_{50k}-\bar{y})}{S\sqrt{\left(3n_4/16 + n_5/4\right)}}$$
$$= \frac{\left\{\sum_{k=1}^{n_{42}}\frac{y_{42k}-\bar{y}}{S} + \sum_{k=1}^{n_{41}}\frac{y_{41k}-\bar{y}}{S} - 3\sum_{k=1}^{n_{40}}\frac{y_{40k}-\bar{y}}{S} + 2\sum_{k=1}^{n_{51}}\frac{y_{51k}-\bar{y}}{S} - 2\sum_{k=1}^{n_{50}}\frac{y_{50k}-\bar{y}}{S}\right\}}{\sqrt{3n_4+4n_5}}.$$

which is asymptotically distributed as the standard normal distribution as well.

Appendix II

Derivation of the Simple form of the T_{ADD}

If the genetic model is additive, $t = 1/2$, $D_j = \partial u_j/\partial d = j/2$ for $j = 0, 1, 2$, and thus $[D_j - E_i(D)] = \frac{1}{2}\left[j - \sum j \cdot p_{j|i}\right]$, where $\sum j \cdot p_{j|i}$ is the conditional expectation of the offspring genotype index given the parental mating type i, i.e.,

$$\sum j \cdot p_{j|i} = \sum j \Pr\left(g_c = j | g_p = i\right)$$
$$= E\left(g_c = j | g_p = i\right).$$

Substituting the above result into (A5) and based on (A7), we obtain

$$T_{ADD} = \frac{\frac{1}{2}\left\{\sum_{i,j}^{n_{ij}}\sum_{k}\left(y_{ijk}-\bar{y}\right)\left[j - E\left(g_c|g_p = i\right)\right]\right\}}{S\sqrt{\left(n_2/16 + n_4/8 + n_5/16\right)}}$$
$$= \frac{2\left\{\sum_{i,j}^{n_{ij}}\sum_{k}\left(\frac{y_{ijk}-\bar{y}}{S}\right)\left[j - E\left(g_c|g_p = i\right)\right]\right\}}{\sqrt{n_2+2n_4+n_5}}.$$

Appendix III

Pairwise Correlations Among the Three Score Statistics

Recall that the three score test statistics derived in Appendix I are

$$T_{REC} = \left(\frac{n_2}{4} + \frac{3n_4}{16}\right)^{-1/2}\sum_{i,j}\sum_{k=1}^{n_{ij}}\left(\frac{y_{ijk}-\bar{y}}{S}\right) \times (D_j - E_i(D))_{REC},$$

$$T_{ADD} = \left(\frac{n_2}{16} + \frac{n_4}{8} + \frac{n_5}{16}\right)^{-1/2}\sum_{i,j}\sum_{k=1}^{n_{ij}}\left(\frac{y_{ijk}-\bar{y}}{S}\right) \times (D_j - E_i(D))_{ADD}$$

and

$$T_{DOM} = \left(\frac{3n_4}{16} + \frac{n_5}{4}\right)^{-1/2}\sum_{i,j}\sum_{k=1}^{n_{ij}}\left(\frac{y_{ijk}-\bar{y}}{S}\right) \times (D_j - E_i(D))_{DOM},$$

where the subscripts REC, ADD and DOM are used to distinguish the values of $(D_j - E_i(D))$ under the recessive, additive and dominant models. Because the offspring genotypes in all triad families are independent, correlation is absent between one offspring and the other for any two of the three statistics, but presents in D_j within the same offspring under different models. Let $c_{12} = (\frac{n_2}{4} + \frac{3n_4}{16})^{-1/2}\left(\frac{n_2}{16} + \frac{n_4}{8} + \frac{n_5}{16}\right)^{-1/2}$, $c_{13} = (\frac{n_2}{4} + \frac{3n_4}{16})^{-1/2}\left(\frac{3n_4}{16} + \frac{n_5}{4}\right)^{-1/2}$, $c_{23} = (\frac{n_2}{16} + \frac{n_4}{8} + \frac{n_5}{16})^{-1/2}\left(\frac{3n_4}{16} + \frac{n_5}{4}\right)^{-1/2}$ and $Z_{ijk} = (y_{ijk} - \bar{y})/S$. Under the null hypothesis, the correlation coefficient between T_{REC} and T_{ADD} is

$$\begin{aligned}
&\operatorname{corr}\left(T_{REC} \cdot T_{ADD}\right) = E\left(T_{REC} \cdot T_{ADD}\right) \\
&= c_{12}\sum_{i,j}\sum_{k=1}^{n_{ij}} Z_{ijk}^2 E\left\{(D_j - E_i(D))_{REC}(D_j - E_i(D))_{ADD}\right\} \\
&= c_{12}\left\{\sum_{j}\sum_{k=1}^{n_{2j}} Z_{2jk}^2\left[\left(1-\frac{1}{2}\right)\left(1-\frac{3}{4}\right)\cdot\frac{1}{2}\right.\right. \\
&\quad \left.+ \left(0-\frac{1}{2}\right)\left(\frac{1}{2}-\frac{3}{4}\right)\cdot\frac{1}{2}\right] \\
&\quad + \sum_{j}\sum_{k=1}^{n_{4j}} Z_{4jk}^2\left[\left(1-\frac{1}{4}\right)\left(1-\frac{1}{2}\right)\cdot\frac{1}{4}\right. \\
&\quad \left.\left.+ \left(0-\frac{1}{4}\right)\left(\frac{1}{2}-\frac{1}{2}\right)\cdot\frac{1}{2} + \left(0-\frac{1}{4}\right)\left(0-\frac{1}{2}\right)\cdot\frac{1}{4}\right]\right\} \\
&= c_{12}\left\{\frac{1}{8}\sum_{j}\sum_{k=1}^{n_{2j}} Z_{2jk}^2 + \frac{1}{8}\sum_{j}\sum_{k=1}^{n_{4j}} Z_{4jk}^2\right\}
\end{aligned}$$

Since $E_Y(Z_{ijk}^2) = 1$, by the law of large number theorem, when n is large, $\sum_{j=0}^{2}\sum_{k=1}^{n_{ij}} Z_{ijk}^2$ converges in probability to n_i for each i. Therefore, the correlation coefficient converges in probability to

$$c_{12}\left\{\frac{n_2}{8}+\frac{n_4}{8}\right\} = \frac{2n_2+2n_4}{\sqrt{(4n_2+3n_4)(n_2+2n_4+n_5)}}.$$

Similarly, the correlation coefficients between T_{REC} and T_{DOM} and between T_{ADD} and T_{DOM} are

$$\begin{aligned}
&\operatorname{corr}(T_{REC}\cdot T_{DOM}) = E(T_{REC}\cdot T_{DOM})\\
&= c_{13}\sum_{i,j}\sum_{k=1}^{n_{ij}} Z_{ijk}^2\\
&\quad\times E\left\{(D_j - E_i(D))_{REC}(D_j - E_i(D))_{DOM}\right\}\\
&= c_{13}\sum_{j}\sum_{k=1}^{n_{4j}} Z_{4jk}^2\left\{\left(1-\frac{1}{4}\right)\left(1-\frac{3}{4}\right)\cdot\frac{1}{4}\right.\\
&\quad\left.+\left(0-\frac{1}{4}\right)\left(1-\frac{3}{4}\right)\cdot\frac{1}{2}+\left(0-\frac{1}{4}\right)\left(0-\frac{3}{4}\right)\cdot\frac{1}{4}\right\}\\
&\to \frac{c_{13}}{16}\cdot n_4\\
&= \frac{n_4}{\sqrt{(4n_2+3n_4)(3n_4+4n_5)}}
\end{aligned}$$

and

$$\begin{aligned}
&\operatorname{corr}(T_{ADD}\cdot T_{DOM}) = E(T_{ADD}\cdot T_{DOM})\\
&= c_{23}\sum_{i,j}\sum_{k=1}^{n_{ij}} Z_{ijk}^2\\
&\quad\times E\left\{(D_j - E_i(D))_{ADD}(D_j - E_i(D))_{DOM}\right\}\\
&= c_{23}\left\{\sum_{j}\sum_{k=1}^{n_{4j}} Z_{4jk}^2\left[\left(1-\frac{1}{2}\right)\left(1-\frac{3}{4}\right)\cdot\frac{1}{4}\right.\right.\\
&\quad\left.+\left(\frac{1}{2}-\frac{1}{2}\right)\left(1-\frac{3}{4}\right)\cdot\frac{1}{2}+\left(0-\frac{1}{2}\right)\left(0-\frac{3}{4}\right)\cdot\frac{1}{4}\right]\\
&\quad+\sum_{j}\sum_{k=1}^{n_{5j}} Z_{5jk}^2\left[\left(\frac{1}{2}-\frac{1}{4}\right)\left(1-\frac{1}{2}\right)\cdot\frac{1}{2}\right.\\
&\quad\left.\left.+\left(0-\frac{1}{4}\right)\left(0-\frac{1}{2}\right)\cdot\frac{1}{2}\right]\right\}\\
&\to c_{23}\left\{\frac{n_4}{8}+\frac{n_5}{8}\right\}\\
&= \frac{2n_4+2n_5}{\sqrt{(3n_4+4n_5)(n_2+2n_4+n_5)}}.
\end{aligned}$$

Received: 7 July 2008
Accepted: 8 December 2008