

## Linkage and association studies of QTL for nuclear families by mixed models

RUZONG FAN\*

*Department of Statistics, Texas A&M University, 447 Blocker Building,  
College Station, TX 77843-3143, USA  
rfan@stat.tamu.edu*

MOMIAO XIONG

*Human Genetics Center, University of Texas - Houston,  
P.O. Box 20334, Houston, Texas 77225, USA*

### SUMMARY

The transmission disequilibrium test (TDT) has been utilized to test the linkage and association between a genetic trait locus and a marker. Spielman *et al.* (1993) introduced TDT to test linkage between a qualitative trait and a marker in the presence of association. In the presence of linkage, TDT can be applied to test for association for fine mapping (Martin *et al.*, 1997; Spielman and Ewens, 1996). In recent years, extensive research has been carried out on the TDT between a quantitative trait and a marker locus (Allison, 1997; Fan *et al.*, 2002; George *et al.*, 1999; Rabinowitz, 1997; Xiong *et al.*, 1998; Zhu and Elston, 2000, 2001). The original TDT for both qualitative and quantitative traits requires unrelated offspring of heterozygous parents for analysis, and much research has been carried out to extend it to fit for different settings. For nuclear families with multiple offspring, one approach is to treat each child independently for analysis. Obviously, this may not be a valid method since offspring of one family are related to each other. Another approach is to select one offspring randomly from each family for analysis. However, with this method much information may be lost. Martin *et al.* (1997, 2000) constructed useful statistical tests to analyse the data for qualitative traits. In this paper, we propose to use mixed models to analyse sample data of nuclear families with multiple offspring for quantitative traits according to the models in Amos (1994). The method uses data of all offspring by taking into account their trait mean and variance–covariance structures, which contain all the effects of major gene locus, polygenic loci and environment. A test statistic based on mixed models is shown to be more powerful than the test statistic proposed by George *et al.* (1999) under moderate disequilibrium for nuclear families. Moreover, it has higher power than the TDT statistic which is constructed by randomly choosing a single offspring from each nuclear family.

*Keywords:* Linkage disequilibrium mapping; Mixed models; Quantitative trait loci; TDT statistics.

### 1. INTRODUCTION

Using nuclear families to test linkage and association between a disease susceptibility locus and a marker is an interesting research area. In the presence of association between a qualitative trait locus and

\*To whom correspondence should be addressed

a marker, Spielman *et al.* (1993) introduced the transmission disequilibrium test (TDT) to test linkage. On the other hand, TDT can be applied to test for association in the presence of linkage for fine gene mapping (Spielman and Ewens, 1996; Martin *et al.*, 1997). If the data consist of nuclear families with a single affected offspring, TDT is a valid  $\chi^2$  test. However, if the data consist of nuclear families with multiple affected offspring, the affected offspring can not be treated as independent siblings since they are related to each other. One approach is to choose an affected offspring randomly from each multiple affected offspring family and construct a TDT based on the simplex nuclear families. However, much data may be discarded by this method. Martin *et al.* (1997) proposed two statistics that use data from all affected offspring for a qualitative trait. The statistics are based on the calculation of joint probability that a heterozygous parent transmits or does not transmit his/her marker alleles to the multiple affected offspring. Moreover, the statistics are shown to be more powerful than the TDT, which is constructed with a single randomly chosen affected offspring from each nuclear family. Martin *et al.* (2000) further extended the test by using data from both affected and unaffected offspring, and proposed a pedigree disequilibrium test (PDT) to analyse data.

In this paper, we investigate linkage and association studies between a quantitative trait locus (QTL) and a marker. First, we calculate conditional means and variance–covariances of the quantitative trait values of offspring of nuclear families, given the transmission status of marker alleles of heterozygous parents. With the mean and variance–covariance structures in hand, we propose to use mixed models to analyse the quantitative genetic data. Using the rich theory of mixed models (Harville, 1974, 1977; Jennrich and Schluchter, 1986; Laird and Ware, 1982; Miller, 1977; Pinheiro, 1994; Pinheiro and Bates, 2000; Robinson, 1991; Searle *et al.*, 1992), we discuss parameter estimations, and introduce a test statistic  $T$  to test linkage or association between the trait locus and the marker.

Mixed models have been utilized for data analysis of QTL of nuclear families in linkage and association studies (Allison *et al.*, 1999), which treats sibship as random factors. Abecasis *et al.* (2000); Fulker *et al.* (1999) and Sham *et al.* (2000) have explored linkage and association studies of quantitative traits by variance–component procedures allowing a simultaneous test of allelic association for family data. In this paper, we show how to construct appropriate models based on conditional mean and variance–covariance structures of the trait values of offspring of nuclear families. For data analysis, one may use the standard statistical packages such as SAS for parameter estimations (Littell *et al.*, 1996).

George *et al.* (1999) proposed interesting regression models to carry out TDT analysis for quantitative traits of nuclear family or pedigree data. Zhu and Elston (2000, 2001) further developed the method by using the distribution of offspring trait values conditional on parent trait values. Moreover, Zhu and Elston (2000, 2001) proposed three test statistics which have good powers in detecting linkage or association, and greatly improve the method of George *et al.* (1999). However, the variance–covariance matrices in George *et al.* (1999) and Zhu and Elston (2000, 2001) only contain the effects of polygenic loci and environment, ignoring the important part of the effect of major gene locus. Our method of this paper is to use the distribution of offspring trait values conditional on the marker allele transmission status of one heterozygous parent at the marker locus. Our variance–covariance matrix contains all effects of major gene locus, polygenic loci and environments.

## 2. MEAN AND VARIANCE–COVARIANCE STRUCTURES

Consider one major quantitative trait locus  $Q$  with two alleles  $Q_1, Q_2$  occurring with frequencies  $q_1, q_2$ . Assume that the expected phenotypic trait value of a person with genotype  $Q_r Q_s$  is  $\nu + \mu_{rs}$ ,  $r, s = 1, 2$ , respectively, where  $\nu$  is overall mean and  $\mu_{12} = \mu_{21}$ . Suppose that a marker locus  $M$  is linked to the trait locus  $Q$ . Denote the recombination fraction between the marker locus  $M$  and the trait locus  $Q$  by  $\theta$ . Assume that two alleles  $M_1, M_2$  are typed at the marker locus  $M$  occurring with frequencies  $p_1, p_2$ . The

haplotype frequency is denoted by  $h_{ri}$  for haplotype  $Q_r M_i$ ,  $r = 1, 2$ ,  $i = 1, 2$ . The linkage disequilibrium coefficient is defined by  $\Delta = h_{11} - p_1 q_1$ . Then  $\Delta \neq 0$  refers to the presence of association or linkage disequilibrium between the trait locus  $Q$  and the marker locus  $M$ . If  $\Delta = 0$ ,  $h_{ri} = q_r p_i$  for all  $r$  and  $i$ , i.e. no association between the marker  $M$  and the trait locus  $Q$ . Besides the major gene locus  $Q$ , assume there is a polygenic effect which affects the trait values.

Let  $Y$  be a phenotypic trait variable of an individual. Following Amos (1994), one may decompose  $Y = \nu + g + G + e$ , where  $\nu$  is the overall mean,  $g$  is the random major gene component such that  $E(g|Q_s Q_r) = \mu_{sr}$ ,  $G$  is the normal random polygenic effect such that  $E(G) = 0$  and  $\text{Var}(G) = \sigma_G^2$ , and  $e$  is an error or residual term which is distributed as normal with mean 0 and variance  $\sigma_e^2$ , i.e.  $e \sim N(0, \sigma_e^2)$ . Moreover, assume that  $g$ ,  $G$  and  $e$  are independent of each other. Let  $TQ$  denote the abbreviation of ‘transmitted quantitative trait allele’. Then we have the conditional mean  $\alpha_r = E[Y|TQ = Q_r] = \nu + \sum_{s=1}^2 \mu_{rs} q_s$ . Let  $\mu_r = \sum_{s=1}^2 \mu_{rs} q_s$ , and so  $\alpha_r = \nu + \mu_r$ . Assume that a parent has heterozygous genotype  $M_1 M_2$  at the marker locus  $M$ . Let  $P(M_1, M_2)$  be the probability of an offspring who receives marker allele  $M_1$  from his/her heterozygous parent but not alleles  $M_2$ . Then  $P(M_1, M_2) = [1/2][2p_1 p_2]$ , since the probability of a heterozygous parent possessing allele  $M_1$  at one copy of his/her chromosome and allele  $M_2$  at the other copy of his/her chromosome is  $2p_1 p_2$ , and the probability of giving one of his/her two alleles to an offspring is  $1/2$ . Similarly, we may show that  $P(M_2, M_1) = p_1 p_2$ . Let  $P(Q_r M_i, M_j)$  be the probability of a child who receives haplotype  $Q_r M_i$  from his/her heterozygous parent but not alleles  $M_j$ ,  $j \neq i$ . Then, we have relation  $P(Q_r M_i, M_j) = (1 - \theta)h_{ri} p_j + \theta h_{rj} p_i$ .

## 2.1 Nuclear families with a single child

For a family with two parents and a single offspring, we assume that at least one parent is heterozygous at the marker locus  $M$ . Moreover, assume we may infer clearly the transmission of parental marker alleles to the offspring, i.e. all offspring of homozygous  $\times$  heterozygous matings and all homozygous offspring of heterozygous  $\times$  heterozygous matings (George *et al.*, 1999; Zhu and Elston, 2000, 2001). In Figure 1, assume that a heterozygous mother  $M_1 M_2$  transmits allele  $M_1$  to her daughter in Graph I (i.e. the left-hand side one). On the other hand, the male offspring 2 in Graph II (i.e. the right-hand side one) receives allele  $M_2$  from his heterozygous mother  $M_1 M_2$ . Based on the information in the pedigree, we may calculate the conditional trait mean of the offspring provided that allele  $M_1$  or  $M_2$  is transmitted from the heterozygous parent. Let  $TM$  denote the abbreviation of ‘transmitted marker allele’,  $NM$  of ‘non-transmitted marker allele’, and  $TH$  of ‘transmitted haplotype’. Given that marker allele  $M_1$  is transmitted and allele  $M_2$  is not transmitted from the heterozygous mother, the conditional trait mean of the offspring 1 in graph I of Figure 1 is

$$\begin{aligned}
 \alpha_{1,2} &= E(y_1) = E[Y|TM = M_1, NM = M_2] \\
 &= \sum_{r=1}^2 E[Y|TH = Q_r M_1, NM = M_2] \frac{P(Q_r M_1, M_2)}{P(M_1, M_2)} \\
 &= \sum_{r=1}^2 \alpha_r [(1 - \theta)h_{r1} p_2 + \theta h_{r2} p_1] / [p_1 p_2] \\
 &= \nu + \sum_{r=1}^2 \mu_r [(1 - \theta)h_{r1} p_2 + \theta h_{r2} p_1] / [p_1 p_2].
 \end{aligned} \tag{1}$$

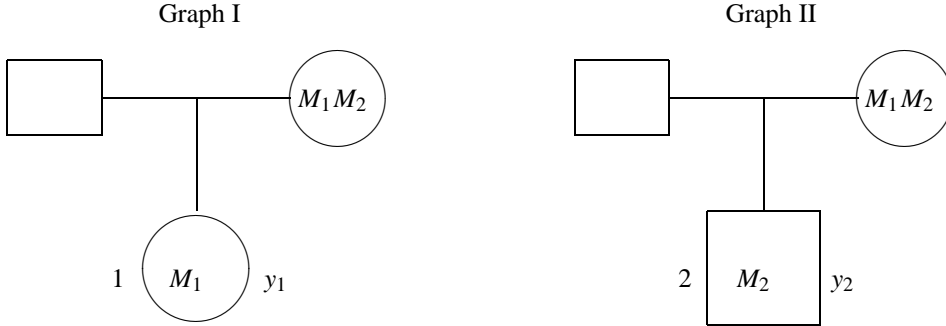


Fig. 1. Two nuclear families each has a single offspring, and a heterozygous  $M_1M_2$  mother at the marker locus  $M$ . For Graph I, the heterozygous mother transmits allele  $M_1$  to her daughter. In Graph II, the heterozygous mother transmits allele  $M_2$  to her son.

Similarly, the conditional mean of the offspring 2 in graph II of Figure 1 is

$$\begin{aligned}
 \alpha_{2,1} &= E(y_2) = E[Y|TM = M_2, NM = M_1] \\
 &= \sum_{r=1}^2 \alpha_r [(1 - \theta)h_{r2}p_1 + \theta h_{r1}p_2] / [p_1p_2] \\
 &= v + \sum_{r=1}^2 \mu_r [(1 - \theta)h_{r2}p_1 + \theta h_{r1}p_2] / [p_1p_2]. \quad (2)
 \end{aligned}$$

Equations (1) and (2) lead to

$$\alpha_{1,2} - \alpha_{2,1} = (1 - 2\theta) \sum_{r=1}^2 \mu_r (h_{r1}p_2 - h_{r2}p_1) / (p_1p_2) = (1 - 2\theta) \Delta(\mu_1 - \mu_2) / (p_1p_2). \quad (3)$$

Hence, one may construct statistics and models to test linkage in the presence of association (or association in the presence of linkage) between the trait locus and the marker by comparing  $\alpha_{1,2}$  and  $\alpha_{2,1}$ . To build valid test statistics and models, we need to calculate the conditional variance-covariances of the trait values of offspring in nuclear families. In Appendix A, we show that the conditional variance of trait value of the female offspring 1 in graph I is  $\sigma_{1,2}^2 = \Sigma_{12}^2 + \sigma_G^2 + \sigma_e^2$ , where  $\Sigma_{12}^2 = \sum_{r=1}^2 \sum_{s=1}^2 (v + \mu_{rs} - \alpha_{1,2})^2 q_s P(Q_r M_1, M_2) / P(M_1, M_2)$ . Similarly, the conditional variance of trait value of the male offspring 2 in graph II is  $\sigma_{2,1}^2 = \Sigma_{21}^2 + \sigma_G^2 + \sigma_e^2$ , where  $\Sigma_{21}^2 = \sum_{r=1}^2 \sum_{s=1}^2 (v + \mu_{rs} - \alpha_{2,1})^2 q_s P(Q_r M_2, M_1) / P(M_2, M_1)$ . The trait values of offspring from different families are independent, and so the covariance of the traits of offspring 1 and 2 in Figure 1 is 0.

## 2.2 General nuclear families

For pedigrees which have more than two offspring, we may get the conditional mean and variance-covariances in a similar way as in Section 2.1. For example, let us look at a pedigree depicted in Figure 2. Suppose we may clearly infer the transmission of parental marker alleles to the offspring. Assume that the genotype of the father at the marker locus is heterozygous  $M_1M_2$ . Moreover, the father transmits allele

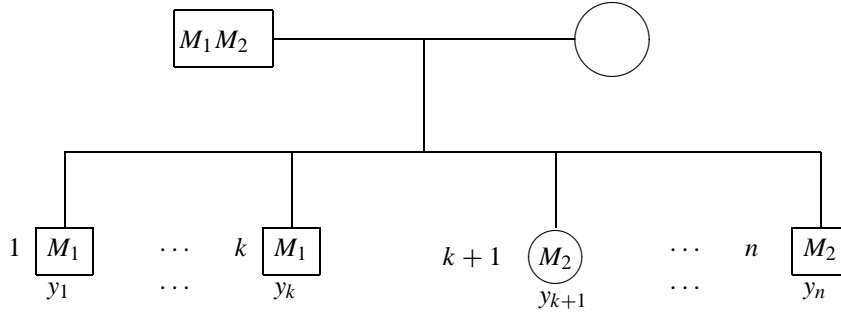


Fig. 2. A nuclear family with  $n$  offspring. Assume that the genotype of the father at the marker locus is heterozygous  $M_1M_2$ . Moreover, the father transmits allele  $M_1$  to kids  $1, \dots, k$ , and transmits allele  $M_2$  to kids  $k+1, \dots, n$ .

$M_1$  to children  $1, \dots, k$ , and transmits allele  $M_2$  to children  $k+1, \dots, n$ . For child  $i$ , the quantitative trait value is denoted by  $y_i, i = 1, \dots, n$ . As in equations (1) and (2), we may calculate the conditional means of the traits of offspring  $1, \dots, k$  as  $\alpha_{1,2}$ , and the conditional means of the traits of offspring  $k+1, \dots, n$  as  $\alpha_{2,1}$ . Moreover, we may obtain the conditional variances of  $y_1, \dots, y_k$  as  $\sigma_{1,2}^2 = \Sigma_{12}^2 + \sigma_G^2 + \sigma_e^2$ , and the conditional variances of  $y_{k+1}, \dots, y_n$  as  $\sigma_{2,1}^2 = \Sigma_{21}^2 + \sigma_G^2 + \sigma_e^2$ . In Appendix B, we calculate the conditional covariance between  $y_l (l = 1, \dots, k)$  and  $y_t (t \neq l, t = 1, \dots, k)$  as  $\Sigma_{12,12}$ , the conditional covariance between  $y_l (l = 1, \dots, k)$  and  $y_t (t = k+1, \dots, n)$  as  $\Sigma_{12,21} = \Sigma_{21,12}$ , and the conditional covariance between  $y_l (l = k+1, \dots, n)$  and  $y_t (t \neq l, t = k+1, \dots, n)$  as  $\Sigma_{21,21}$ .

### 3. MODELS FOR LINKAGE AND ASSOCIATION STUDIES

Based on the analysis in Section 2, we propose to use a mixed linear model to analyse the data from nuclear families. Suppose we have  $I$  nuclear families, each of them has at least one child. Moreover, assume that at least one of the parents is heterozygous, and that one can determine which marker allele is transmitted to the offspring from the heterozygous parent. For each child, a quantitative trait is observed. For the  $i$ th family, assume that there are  $n_i$  siblings, and the siblings' trait values are listed as  $y_{i1}, \dots, y_{in_i}$ .

Consider the  $n_i$  offspring in the  $i$ th family. Assume that the siblings consist of two parts: (1)  $k_i$  siblings correspond to that allele  $M_1$  is transmitted and allele  $M_2$  is not transmitted from their heterozygous parent, and their trait values are listed as  $y_{i1}, \dots, y_{ik_i}$ ; (2) the rest of the siblings correspond to that allele  $M_1$  is not transmitted and allele  $M_2$  is transmitted from their heterozygous parent, and their trait values are listed as  $y_{i,k_i+1}, \dots, y_{in_i}$ . Under the alternative hypothesis of linkage (or association) in the presence of association (or linkage) between the trait locus and the marker locus, one may use a full multivariate linear model

$$\begin{aligned} y_{ij} &= \nu + g_{ij}|_{(TM=M_1, NM=M_2)} + G_{ij} + e_{ij}, j = 1, 2, \dots, k_i, \\ y_{ij} &= \nu + g_{ij}|_{(TM=M_2, NM=M_1)} + G_{ij} + e_{ij}, j = k_i + 1, \dots, n_i. \end{aligned} \quad (4)$$

$y_{ij}$  are normal variables with means  $\alpha_{1,2}$  for  $j = 1, \dots, k_i$  and  $\alpha_{2,1}$  for  $j = k_i + 1, \dots, n_i$ , and an  $n_i \times n_i$

variance–covariance matrix

$$\Gamma_i = \begin{pmatrix} \sigma_{1,2}^2 & \Sigma_{12,12} & \cdots & \Sigma_{12,12} & \Sigma_{12,21} & \cdots & \Sigma_{12,21} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \Sigma_{12,12} & \Sigma_{12,12} & \cdots & \sigma_{1,2}^2 & \Sigma_{12,21} & \cdots & \Sigma_{12,21} \\ \Sigma_{21,12} & \Sigma_{21,12} & \cdots & \Sigma_{21,12} & \sigma_{2,1}^2 & \cdots & \Sigma_{21,21} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \Sigma_{21,12} & \Sigma_{21,12} & \cdots & \Sigma_{21,12} & \Sigma_{21,21} & \cdots & \sigma_{2,1}^2 \end{pmatrix}.$$

Let  $n = \sum_{i=1}^I n_i$  be the total number of offspring of all families,  $\vec{y}_i = (y_{i1}, \dots, y_{ini})^\tau$  and  $\vec{y} = (\vec{y}_1^\tau, \dots, \vec{y}_I^\tau)^\tau$ , and define  $\vec{g}_i, \vec{g}, \vec{G}_i, \vec{G}, \vec{e}_i$  and  $\vec{e}$ , accordingly. Define an  $n_i \times 2$  design matrix

$$X_i = \begin{pmatrix} 1 & \cdots & 1 & 0 & \cdots & 0 \\ 0 & \cdots & 0 & 1 & \cdots & 1 \end{pmatrix}^\tau,$$

and a total  $n \times 2$  design matrix  $X = (X_1^\tau, \dots, X_I^\tau)^\tau$ . Then,  $\vec{y}$  is normally distributed with the following mean and variance–covariance structure for model (4):

$$E(\vec{y}) = X \begin{pmatrix} \alpha_{1,2} \\ \alpha_{2,1} \end{pmatrix}, \quad \text{Cov}(\vec{y}) = \text{diag}(\Gamma_1, \dots, \Gamma_I).$$

Under the null hypothesis of no linkage (i.e.  $\theta = 1/2$ ) in the presence association (i.e.  $\Delta \neq 0$ ) between the trait locus and the marker locus, then  $\alpha_{1,2} |_{\theta=1/2, \Delta \neq 0} = \alpha_{2,1} |_{\theta=1/2, \Delta \neq 0}$  by relation (3). Moreover, the conditional variance  $\sigma_{1,2}^2 |_{\theta=1/2, \Delta \neq 0} = \sigma_{2,1}^2 |_{\theta=1/2, \Delta \neq 0}$  since  $P(Q_r M_1, M_2) = P(Q_r M_2, M_1)$ . Besides, we can show that  $\Sigma_{12,12} = \Sigma_{12,21} = \Sigma_{21,12} = \Sigma_{21,21} |_{\theta=1/2, \Delta \neq 0}$  by the formulae in Appendix B. Therefore, one may use a multivariate linear model to analyse the data

$$y_{ij} = \nu + g_{ij} + G_{ij} + e_{ij}, \quad j = 1, 2, \dots, n_i, \quad \text{reduced model I}, \quad (5)$$

where  $y_{ij}$  are normal variables with mean  $\alpha_{1,2} |_{\theta=1/2, \Delta \neq 0}$  and  $n_i \times n_i$  variance–covariance matrix

$$W_i = \begin{pmatrix} \sigma_{1,2}^2 & \Sigma_{12,12} & \cdots & \Sigma_{12,12} \\ \Sigma_{12,12} & \sigma_{1,2}^2 & \cdots & \Sigma_{12,12} \\ \vdots & \vdots & \vdots & \vdots \\ \Sigma_{12,12} & \Sigma_{12,12} & \cdots & \sigma_{1,2}^2 \end{pmatrix}_{\theta=1/2, \Delta \neq 0}.$$

Under the null hypothesis of no association in the presence of linkage between the trait locus and the marker locus,  $\alpha_{i,j} = \sum_{r=1}^2 \alpha_r q_r = \alpha, i \neq j, i, j = 1, 2$ , which is the population trait mean. Let us denote  $\sum_{r=1}^2 \sum_{s=1}^2 \mu_{rs} q_r q_s$  by  $\mu$ ,  $\gamma_{kl} = \mu_{kl} - \mu$ , and  $\gamma_k = \sum_l \gamma_{kl} q_l$ . The conditional variance  $\sigma_{i,j}^2 = \sum_{r=1}^2 \sum_{s=1}^2 (\mu_{rs} - \mu)^2 q_s q_r + \sigma_G^2 + \sigma_e^2, i \neq j$ , which is the total variance of a trait variable. Moreover, the conditional covariance  $\Sigma_{12,12} |_{\Delta=0, \theta \neq 1/2}$  is equal to the covariance  $\Sigma_{21,21} |_{\Delta=0, \theta \neq 1/2}$  of trait variables between siblings (Appendix B). Therefore, one may use a multivariate linear model to analyse the data

$$y_{ij} = \nu + g_{ij} + G_{ij} + e_{ij}, \quad j = 1, 2, \dots, n_i, \quad \text{reduced model II}, \quad (6)$$

where  $y_{ij}$  are normal variables with mean  $\alpha = v + \mu$  and  $n_i \times n_i$  variance–covariance matrix

$$V_i = \begin{pmatrix} \sigma_{1,2}^2 & \Sigma_{12,12} & \cdots & \Sigma_{12,12} & \Sigma_{12,21} & \cdots & \Sigma_{12,21} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \Sigma_{12,12} & \Sigma_{12,12} & \cdots & \sigma_{1,2}^2 & \Sigma_{12,21} & \cdots & \Sigma_{12,21} \\ \Sigma_{21,12} & \Sigma_{21,12} & \cdots & \Sigma_{21,12} & \sigma_{1,2}^2 & \cdots & \Sigma_{12,12} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \Sigma_{21,12} & \Sigma_{21,12} & \cdots & \Sigma_{21,12} & \Sigma_{12,12} & \cdots & \sigma_{1,2}^2 \end{pmatrix}_{\Delta=0, \theta \neq 1/2}.$$

#### 4. PARAMETER ESTIMATION AND TEST STATISTICS

In model (4), there are two mean coefficients  $\alpha_{1,2}$  and  $\alpha_{2,1}$ , five variance–covariance parameters  $\Sigma_{12}^2, \Sigma_{12,12}, \Sigma_{12,21}, \Sigma_{21}^2, \Sigma_{21,21}$ , polygenic variance  $\sigma_G^2$ , and error variance  $\sigma_e^2$ . Due to the redundancy, not all variance–covariance parameters are identifiable. However, five parameters  $(\sigma_{1,2}^2, \sigma_{2,1}^2, \Sigma_{12,12}, \Sigma_{12,21}, \Sigma_{21,21}) = \sigma^\tau$  are independent and identifiable. Similarly, in model (5), there are one population mean coefficient  $\alpha_{1,2|\theta=1/2, \Delta \neq 0}$  and two variance–covariance parameters  $\eta_I^\tau = (\sigma_{1,2}^2|_{\theta=1/2, \Delta \neq 0}, \Sigma_{12,12}|_{\theta=1/2, \Delta \neq 0})$  which are independent and identifiable. In model (6), there is one population mean coefficient  $\alpha$  and three variance–covariance parameters  $\eta_{II}^\tau = (\sigma_{1,2}^2|_{\Delta=0, \theta \neq 1/2}, \Sigma_{12,12}|_{\Delta=0, \theta \neq 1/2}, \Sigma_{12,21}|_{\Delta=0, \theta \neq 1/2})$  which are independent and identifiable. In the following, we are going to discuss the parameter estimations for model (4). The same method can be applied to models (5) and (6). Denote  $\beta^\tau = (\alpha_{1,2}, \alpha_{2,1})$ . The loglikelihood function of model (4) is

$$l = -\frac{n}{2} \log(2\pi) - \frac{1}{2} \sum_{i=1}^I \log |\Gamma_i| - \frac{1}{2} \sum_{i=1}^I (\bar{y}_i - X_i \beta)^\tau \Gamma_i^{-1} (\bar{y}_i - X_i \beta).$$

Based on the above loglikelihood function, one may estimate the parameters  $\beta$  and  $\sigma$  by Newton–Raphson or Fisher scoring algorithms (Jennrich and Schluchter, 1986).

Under the null hypothesis  $H_0 : \theta = 1/2$  of no linkage in the presence of association or  $H_0 : \Delta = 0$  of no association in the presence of linkage between the trait locus and the marker locus, one can get  $(1, -1)\beta = 0$  from relation (3). Hence, we may construct a  $\chi^2$ -statistic

$$T = \frac{(\hat{\alpha}_{1,2} - \hat{\alpha}_{2,1})^2}{(1, -1) \left[ \sum_{i=1}^I X_i^\tau \hat{\Gamma}_i^{-1} X_i \right]^{-1} \begin{pmatrix} 1 \\ -1 \end{pmatrix}}. \quad (7)$$

In Appendix C, we discuss the asymptotic distribution of statistic  $T$ . By assuming that the sample size is large, we may show that  $\hat{\beta}$  is asymptotic normal (see more details in Appendix C). Under the alternative hypothesis,  $T$  is distributed approximately as a non-central  $\chi^2$  random variable with non-central parameter

$$\lambda_T \approx \frac{(\alpha_{1,2} - \alpha_{2,1})^2}{(1, -1) \left[ \sum_{i=1}^I X_i^\tau \Gamma_i^{-1} X_i \right]^{-1} \begin{pmatrix} 1 \\ -1 \end{pmatrix}}.$$

The above relation is not strictly equal since the maximum likelihood estimates may not be unbiased.

The above formula can be simplified when we consider nuclear families with one single child or sib-pair. If  $n_i = 1$  for each family, then there is only one child in each family. Let  $y_i, i = 1, 2, \dots, m_1$

correspond to the trait values of offspring who receive allele  $M_1$  from their heterozygous parents, and  $y_i, i = m_1 + 1, 2, \dots, m_1 + m_2$  correspond to the trait values of offspring who receive allele  $M_2$  from their heterozygous parents. Then  $\hat{\alpha}_{1,2} = \sum_{i=1}^{m_1} y_i / m_1, \hat{\alpha}_{2,1} = \sum_{i=m_1+1}^{m_1+m_2} y_i / m_2, \hat{\sigma}_{1,2}^2 = \sum_{i=1}^{m_1} (y_i - \hat{\alpha}_{1,2})^2 / m_1,$  and  $\hat{\sigma}_{2,1}^2 = \sum_{i=m_1+1}^{m_1+m_2} (y_i - \hat{\alpha}_{2,1})^2 / m_2.$  The test statistic  $T$  is

$$T_{\text{singleton}} = \frac{(\hat{\alpha}_{1,2} - \hat{\alpha}_{2,1})^2}{\hat{\sigma}_{1,2}^2 / m_1 + \hat{\sigma}_{2,1}^2 / m_2}.$$

The non-centrality parameter of the singleton test statistic  $T_{\text{singleton}}$  is

$$\lambda_{\text{singleton}} \approx \frac{[(1 - 2\theta)\Delta(\mu_1 - \mu_2)/(p_1 p_2)]^2}{\sigma_{1,2}^2 / m_1 + \sigma_{2,1}^2 / m_2}.$$

Assume that the data consist of two parts: singleton and sib-pair families. Suppose there are  $k_1$  singleton offspring who receive allele  $M_1$  from their heterozygous parents,  $k_2$  singleton offspring who receive allele  $M_2$  from their heterozygous parents,  $k_{ii}$  ( $i = 1, 2$ ) sib pairs in each of them where both sibs receive allele  $M_i$  from their heterozygous parents, and  $k_{12}$  sib pairs in each of them where one sib receives allele  $M_1$  from his/her heterozygous parent and the other receives allele  $M_2$  from the same heterozygous parent. In Appendix D, we obtain the test statistic  $T$  as

$$T_{\text{singleton,sibs}} = \frac{(\hat{\alpha}_{1,2} - \hat{\alpha}_{2,1})^2 (\hat{c}\hat{f} - \hat{b}^2)}{\hat{c} + \hat{f} - 2\hat{b}}, \quad (8)$$

where

$$\begin{aligned} \hat{c} &= \frac{k_1}{\hat{\sigma}_{1,2}^2} + \frac{k_{12}\hat{\sigma}_{2,1}^2}{\hat{\sigma}_{1,2}^2\hat{\sigma}_{2,1}^2 - \hat{\Sigma}_{12,21}^2} + \frac{2k_{11}}{\hat{\sigma}_{1,2}^2 + \hat{\Sigma}_{12,12}} \\ \hat{b} &= \frac{k_{12}\hat{\Sigma}_{12,21}^2}{\hat{\sigma}_{1,2}^2\hat{\sigma}_{2,1}^2 - \hat{\Sigma}_{12,21}^2} \\ \hat{f} &= \frac{k_2}{\hat{\sigma}_{2,1}^2} + \frac{k_{12}\hat{\sigma}_{1,2}^2}{\hat{\sigma}_{1,2}^2\hat{\sigma}_{2,1}^2 - \hat{\Sigma}_{12,21}^2} + \frac{2k_{22}}{\hat{\sigma}_{2,1}^2 + \hat{\Sigma}_{21,21}}. \end{aligned} \quad (9)$$

The non-centrality parameter is

$$\lambda_{\text{singleton,sibs}} \approx \frac{[(1 - 2\theta)\Delta(\mu_1 - \mu_2)/(p_1 p_2)]^2 (cf - b^2)}{c + f - 2b},$$

where

$$\begin{aligned} c &= \frac{k_1}{\sigma_{1,2}^2} + \frac{k_{12}\sigma_{2,1}^2}{\sigma_{1,2}^2\sigma_{2,1}^2 - \Sigma_{12,21}^2} + \frac{2k_{11}}{\sigma_{1,2}^2 + \Sigma_{12,12}} \\ b &= \frac{k_{12}\Sigma_{12,21}^2}{\sigma_{1,2}^2\sigma_{2,1}^2 - \Sigma_{12,21}^2} \\ f &= \frac{k_2}{\sigma_{2,1}^2} + \frac{k_{12}\sigma_{1,2}^2}{\sigma_{1,2}^2\sigma_{2,1}^2 - \Sigma_{12,21}^2} + \frac{2k_{22}}{\sigma_{2,1}^2 + \Sigma_{21,21}}. \end{aligned} \quad (10)$$

## 5. POWER AND SAMPLE SIZE COMPARISON

To perform power and sample size calculation, we transform the phenotypic values for genotypes  $Q_1Q_1$ ,  $Q_1Q_2$ ,  $Q_2Q_2$  by deducting the average  $(\mu_{11} + \mu_{22})/2$  of expected homozygous trait values from each  $\mu_{ij}$

$$\begin{aligned}\mu'_{11} &= \mu_{11} - (\mu_{11} + \mu_{22})/2 = a, \\ \mu'_{12} &= \mu_{21} - (\mu_{11} + \mu_{22})/2 = d, \\ \mu'_{22} &= \mu_{22} - (\mu_{11} + \mu_{22})/2 = -a.\end{aligned}$$

The additive variance is  $\sigma_a^2 = 2q_1q_2(a + d(q_2 - q_1))^2$ , and the dominant variance is  $\sigma_d^2 = (2q_1q_2d)^2$ . The heritability is denoted by  $h^2$ , and  $h^2$  is defined by  $\sigma_a^2/(\sigma_a^2 + \sigma_d^2 + \sigma_e^2)$ .

## 5.1 Comparison with other tests

To compare powers of test statistic  $T$  with those in George *et al.* (1999) and Zhu and Elston (2000, 2001), we take the same parameter values  $q_1 = 0.5$ ,  $p_1 = 0.4$ ,  $\sigma_G^2 = 0.125$ ,  $\sigma_e = 0.50$ ,  $a = 0.7071$ ,  $d = 0.00$  for an additive mode of inheritance (George *et al.*, 1999, pp. 240–241; Zhu and Elston, 2000, p. 326). For a dominant mode of inheritance, we take  $\mu_{11} = \mu_{12} = 1.1547$ ,  $\mu_{22} = 0.00$ . For a recessive mode of inheritance, we take  $\mu_{11} = 1.1547$ ,  $\mu_{12} = \mu_{22} = 0.00$  (Zhu and Elston, 2000, p. 326). Then the trait locus variance  $\sigma_a^2 + \sigma_d^2 = 0.25$ .

For nuclear families each with two offspring, assume that there are  $k_{ii}$  ( $i = 1, 2$ ) sib pairs in each of them where both sibs receive allele  $M_i$  from their heterozygous parents, and  $k_{12}$  sib pairs in each of them where one sib receives allele  $M_1$  from his/her heterozygous parent and the other receives allele  $M_2$  from the same heterozygous parent. For nuclear families each with four offspring, assume that there are  $k_{iiii}$  ( $i = 1, 2$ ) families where in each of them every offspring receive allele  $M_i$  from their heterozygous parents,  $k_{1112}$  families where in each of them three offspring receives allele  $M_1$  from his/her heterozygous parent and the other one receives allele  $M_2$  from the same heterozygous parent,  $k_{1122}$  families in each of them two offspring receives allele  $M_1$  from their heterozygous parent and the other two receives allele  $M_2$  from the same heterozygous parent, and  $k_{1222}$  families in each of them one offspring receives allele  $M_1$  from their heterozygous parent and the other three receives allele  $M_2$  from the same heterozygous parent.

Table 1 shows the powers of test statistic  $T$  for sib-pair families and four offspring families. For sib-pair families, the rows with  $k_{11} = 50$ ,  $k_{12} = 100$ ,  $k_{22} = 50$  correspond to those in Table 1 of George *et al.* (1999) and Table I of Zhu and Elston (2000) that the total number of families is 200, and the rows with  $k_{11} = 25$ ,  $k_{12} = 50$ ,  $k_{22} = 25$  correspond to those in Table 1 of George *et al.* (1999) and Table I of Zhu and Elston (2000) that the total number of families is 100. For families with four offspring,  $k_{ijmn} = 40$  correspond to those in Table 1 of George *et al.* (1999) and Table I of Zhu and Elston (2000) that the total number of families is 200, and  $k_{ijmn} = 20$  correspond to those in Table 1 of George *et al.* (1999) and Table I of Zhu and Elston (2000) that the total number of families is 100. Therefore, the results in Table 1 are comparable to those in Table 1 of George *et al.* (1999) and Table I of Zhu and Elston (2000).

When there is moderate linkage disequilibrium, i.e.  $\Delta = 0.10$ , we notice that the powers are very much higher for small recombination fraction  $\theta$ . Under moderate disequilibrium ( $\Delta = 0.1$ ), the powers are very much higher than those in Table 1 of George *et al.* (1999), and are similar to those in Table I of Zhu and Elston (2000) by tests  $T_{ZE1}$  and  $T_{ZE2}$ . This shows the advantage of the test statistic  $T$ . Under the weak linkage disequilibrium, i.e.  $\Delta = 0.01$ , the powers are very low and similar to those in Table 1 of George *et al.* (1999).

Table 2 shows the powers of test statistic  $T$  for a dominant mode of inheritance and a recessive mode of inheritance under moderate disequilibrium  $\Delta = 0.10$ . At significant level 5%, the powers in Table 2

Table 1. Powers of test statistic  $T$  of an additive mode of inheritance when  $q_1 = 0.5$ ,  $p_1 = 0.4$ ,  $\sigma_G^2 = 0.125$ ,  $\sigma_e = 0.50$ ,  $a = 0.7071$ ,  $d = 0.00$ . Note:  $k_{ijmn} = 40$  implies that the total number of families is 200, and  $k_{ijmn} = 20$  implies that the total number of families is 100

No. of family	No. of offspring	$\Delta$	$\theta$	Power at 5% significance	Power at 1% significance	Power at 0.5 % significance
$k_{ijmn} = 40$	4	0.10	0.00	98.5	94.0	90.7
			0.01	98.2	93.0	89.3
			0.05	96.3	87.9	82.6
$k_{ijmn} = 40$	4	0.01	0.00	6.8	1.6	0.9
			0.01	6.8	1.6	0.9
			0.05	6.5	1.5	0.8
$k_{11} = 50$	2	0.10	0.00	89.0	72.9	64.7
$k_{12} = 100$			0.01	87.7	70.7	62.3
$k_{22} = 50$			0.05	81.9	61.6	52.5
$k_{11} = 50$	2	0.01	0.00	6.1	1.4	0.7
$k_{12} = 100$			0.01	6.1	1.4	0.7
$k_{22} = 50$			0.05	6.0	1.3	0.7
$k_{ijmn} = 20$	4	0.10	0.00	83.1	63.4	54.4
			0.01	81.7	61.4	52.3
			0.05	75.4	52.9	43.7
$k_{ijmn} = 20$	4	0.01	0.00	5.9	1.3	0.7
			0.01	5.9	1.3	0.7
			0.05	5.8	1.3	0.7
$k_{11} = 25$	2	0.10	0.00	61.5	37.3	28.9
$k_{12} = 50$			0.01	59.8	35.6	27.4
$k_{22} = 25$			0.05	52.8	29.2	21.8
$k_{11} = 25$	2	0.01	0.00	5.6	1.2	0.6
$k_{12} = 50$			0.01	5.5	1.2	0.6
$k_{22} = 25$			0.05	5.5	1.2	0.6

are similar to those of  $T_{ZE1}$  in Tables II and III of Zhu and Elston (2000). Generally,  $T$  has good powers to study quantitative traits.

## 5.2 Comparison of $T_{\text{singleton}}$ and $T_{\text{singleton,sibs}}$

In this section, we are going to compare the powers and sample sizes of test statistics  $T_{\text{singleton}}$  and  $T_{\text{singleton,sibs}}$ . To calculate the non-centrality parameters  $\lambda_{\text{singleton}}$  and  $\lambda_{\text{singleton,sibs}}$ , we need parameters such as the marker allele frequencies  $p_1$  and  $p_2$ , trait allele frequencies  $q_1$  and  $q_2$ , haplotype frequencies  $h_{ri}$ , recombination fraction  $\theta$ , trait values  $\mu_{rs}$ , polygenic variance  $\sigma_G^2$ , and error variance  $\sigma_e^2$ . Due to the evolutionary process of the population, the haplotype frequencies  $h_{ri}$  change from generation to generation. Under a Fisher–Wright model, Xiong and Guo (1997) showed that the haplotype frequencies can be modeled by a diffusion process. The expected haplotype frequencies can be calculated by  $E[h_{ri}] = h_{ri}(0)e^{-\theta A} + p_i q_r(1 - e^{-\theta A})$ , where  $A$  is the age of the most recent mutation at the trait locus and  $h_{ri}(0)$  is the initial haplotype frequency of haplotypes  $Q_r M_i$  at the generation of occurrence of the mutation at the trait locus. If there is only a single mutation in the population, one may assume that  $h_{11}(0) = q_1$ ,  $h_{12}(0) = 0$ , and  $h_{21}(0) = p_1 - q_1 \geq 0$ ,  $h_{22}(0) = p_2$ . Replacing  $h_{ri}$  in  $P(Q_r M_i, M_j)$  by  $E[h_{ri}]$ , we may calculate the approximations of the non-centrality parameters. If there is more than one

Table 2. Powers of test statistic  $T$  when  $q_1 = 0.5$ ,  $p_1 = 0.4$ ,  $\sigma_G^2 = 0.125$ ,  $\sigma_e = 0.50$ .  
 Note:  $k_{ijmn} = 40$  implies that the total number of families is 200, and  $k_{ijmn} = 20$  implies that the total number of families is 100

Dominant mode of inheritance: $\mu_{11} = \mu_{12} = \mu_{21} = 1.1547$ , $\mu_{22} = 0.00$						
No. of family	No. of offspring	$\Delta$	$\theta$	Power at 5% significance	Power at 1% significance	Power at 0.5 % significance
$k_{11} = 25$			0.00	40.5	19.6	13.9
$k_{12} = 50$	2	0.10	0.01	39.4	18.8	13.2
$k_{22} = 25$			0.05	34.8	15.7	10.8
			0.00	64.5	40.4	31.8
$k_{ijmn} = 20$	4	0.10	0.01	63.5	39.3	30.8
			0.05	58.9	34.8	26.7
			0.00	68.2	44.3	35.4
$k_{11} = 50$			0.00	68.2	44.3	35.4
$k_{12} = 100$	2	0.10	0.01	66.7	42.7	33.9
$k_{22} = 50$			0.05	60.2	36.1	27.8
			0.00	91.0	76.6	68.9
$k_{ijmn} = 40$	4	0.10	0.01	90.3	75.3	67.5
			0.05	87.1	69.6	61.1
Recessive mode of inheritance: $\mu_{11} = 1.1547$ , $\mu_{12} = \mu_{21} = \mu_{22} = 0.00$						
No. of family	No. of offspring	$\Delta$	$\theta$	Power at 5% significance	Power at 1% significance	Power at 0.5 % significance
$k_{11} = 25$			0.00	43.2	21.6	15.4
$k_{12} = 50$	2	0.10	0.01	41.9	20.6	14.6
$k_{22} = 25$			0.05	36.6	16.9	11.7
			0.00	62.4	38.2	29.7
$k_{ijmn} = 20$	4	0.10	0.01	60.9	36.7	28.4
			0.05	54.7	30.9	23.3
			0.00	71.6	48.2	39.1
$k_{11} = 50$			0.00	71.6	48.2	39.1
$k_{12} = 100$	2	0.10	0.01	69.9	46.3	37.2
$k_{22} = 50$			0.05	62.9	38.7	30.2
			0.00	89.6	73.9	65.9
$k_{ijmn} = 40$	4	0.10	0.01	88.5	72.1	63.9
			0.05	83.6	64.1	55.2

mutation in a population, one needs more care with the fluctuation of haplotype frequencies. However, assuming a single copy mutation would be a sound assumption for a rare disease.

Using the non-centrality parameters given in Section 4, we can carry out power and sample size calculations by setting different parameter values for the trait gene and the marker allele frequencies, heritability, recombination fraction, the age of mutation for the trait allele, etc. To make the comparison valid, we assume that only one offspring is picked up from the sib-pair families for analysis. Hence,  $k_{ii}$  offspring will be treated as singleton offspring who receive allele  $M_i$  from the heterozygous parents. Moreover, let us treat  $k_{12}/2$  offspring as singleton offspring who receive allele  $M_1$  from the heterozygous parents, and  $k_{12}/2$  offspring as singleton offspring who receive allele  $M_2$  from the heterozygous parents. Hence, we will have  $k_1 + k_{11} + k_{12}/2$  (or  $k_2 + k_{22} + k_{12}/2$ ) offspring who are treated as singleton offspring who receive allele  $M_1$  (or  $M_2$ ) from the heterozygous parents. In Figures 3–6, we assume that  $m_1 = m_2 = 100$  and  $k_i = k_{ij} = 40$ . Since  $k_1 + k_{11} + k_{12}/2 = k_2 + k_{22} + k_{12}/2 = 100$ ,  $T_{\text{singleton}}$  in these

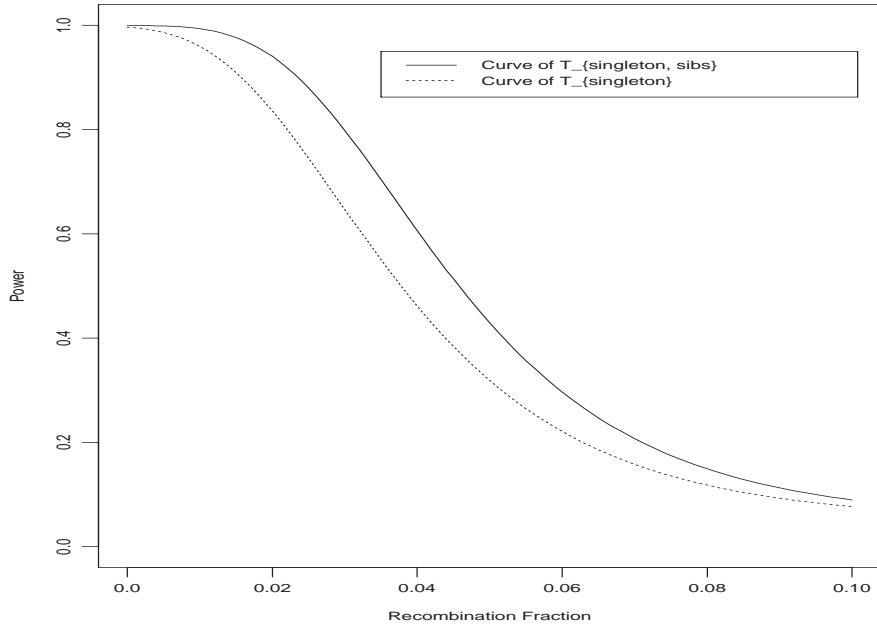


Fig. 3. Power curves of  $T_{\text{singleton}}$  and  $T_{\text{singleton,sibs}}$  when  $q_1 = 0.25$ ,  $h^2 = 0.60$ ,  $A = 20$ ,  $p_1 = 0.50$ ,  $m_1 = m_2 = 100$ ,  $k_1 = k_2 = 40$ ,  $k_{11} = k_{12} = k_{22} = 40$ ,  $\sigma_G^2 = 0.75$  for a dominant trait  $a = d = 1.0$ .

figures can be viewed as the singleton families constructed from the nuclear families for  $T_{\text{singleton,sibs}}$ .

Figures 3 and 4 plot the power curves of  $T_{\text{singleton}}$  and  $T_{\text{singleton,sibs}}$  against the recombination fraction for dominant and recessive traits. Figures 5 and 6 plot the power curves of  $T_{\text{singleton}}$  and  $T_{\text{singleton,sibs}}$  against the heritability for dominant and recessive traits. The four figures show that  $T_{\text{singleton,sibs}}$  has higher power than that of  $T_{\text{singleton}}$ .

## 6. AN EXAMPLE

As an example, we apply our method to Genetic Analysis Workshop 12 Oxford asthma data (Cookson and Abecasis, 2001). The data consist of 80 nuclear families with a total of 203 offspring. In these 80 families, 43 have two offspring, 31 have three offspring, and 6 have four offspring. In Daniel *et al.* (1996), linkage to bronchial responsiveness to methacholine (slope) and other quantitative traits was tested by the Haseman–Elston sib-pair technique (Haseman and Elston, 1972). Three regions of potential linkage to autosomal markers were detected with  $\log_e$ slope on chromosomes 4, 7 and 16 in Daniel *et al.* (1996).

We analyse the data using model (4), and test the linkage by using statistic  $T$  in (7). By using SAS mixed model procedures, we have got the following results:

Marker Locus	Estimate of parameters $\sigma^\tau$	P-value of $T$ in (7) to test $\alpha_{1,2} = \alpha_{2,1}$	P-value of Daniel <i>et al.</i> (1996)
D4S1540	(1.28, 2.70, 0.88, 1.23, 1.10)	0.02	<0.05
D7S484	(2.31, 1.94, 1.36, 1.07, 0.58)	0.02	<0.0005
D16S515	(3.02, 2.06, 0.95, 0.51, 0.80)	0.04	<0.05
D16S289	(2.25, 1.24, -1.10, 0.75, 0.31)	<0.0001	<0.05

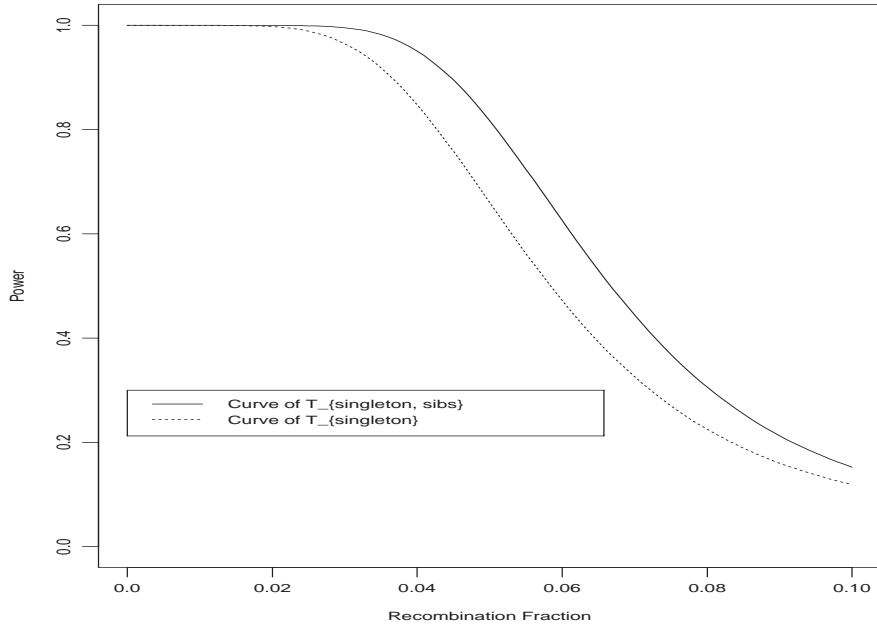


Fig. 4. Power curves of  $T_{\text{singleton}}$  and  $T_{\text{singleton,sibs}}$  when  $q_1 = 0.50$ ,  $h^2 = 0.50$ ,  $A = 20$ ,  $p_1 = 0.50$ ,  $m_1 = m_2 = 100$ ,  $k_1 = k_2 = 40$ ,  $k_{11} = k_{12} = k_{22} = 40$ ,  $\sigma_G^2 = 1.0$  for a recessive trait  $a = 1.0$  and  $d = -0.5$ .

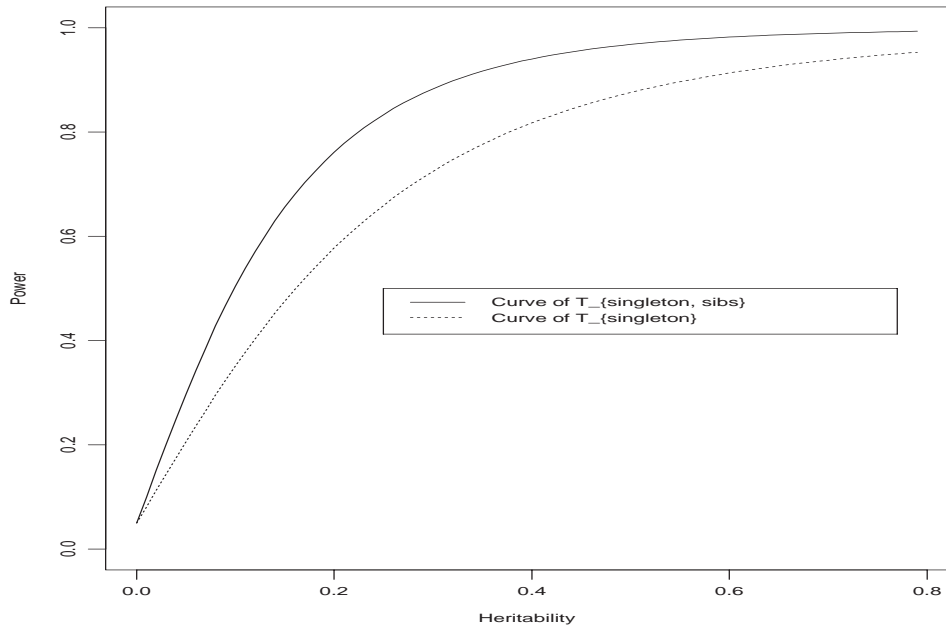


Fig. 5. Power curves of  $T_{\text{singleton}}$  and  $T_{\text{singleton,sibs}}$  when  $q_1 = 0.25$ ,  $\theta = 0.005$ ,  $A = 20$ ,  $p_1 = 0.50$ ,  $m_1 = m_2 = 100$ ,  $k_1 = k_2 = 40$ ,  $k_{11} = k_{12} = k_{22} = 40$ ,  $\sigma_G^2 = 0.75$  for a dominant trait  $a = d = 1.0$ .

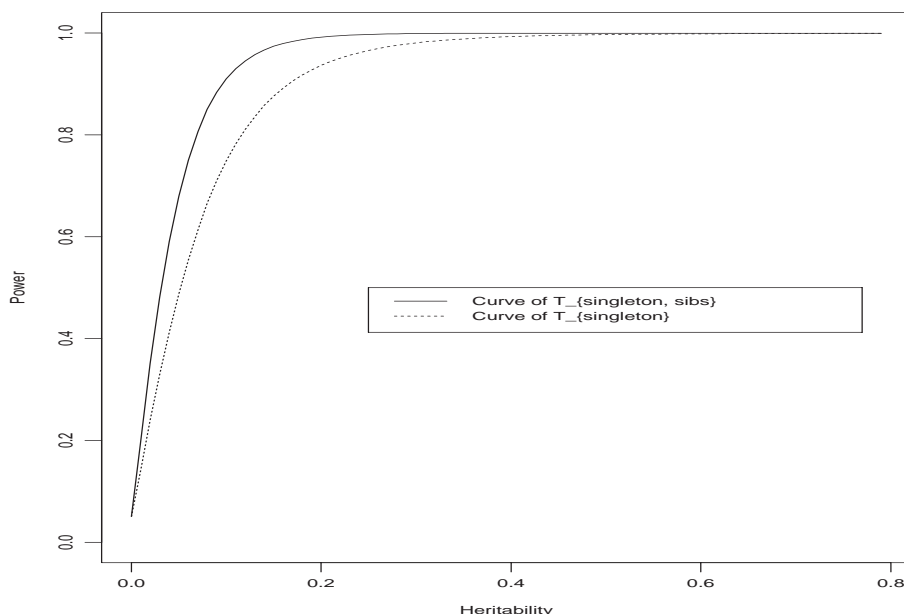


Fig. 6. Power curves of  $T_{\text{singleton}}$  and  $T_{\text{singleton,sibs}}$  when  $q_1 = 0.50$ ,  $\theta = 0.005$ ,  $A = 20$ ,  $p_1 = 0.50$ ,  $m_1 = m_2 = 100$ ,  $k_1 = k_2 = 40$ ,  $k_{11} = k_{12} = k_{22} = 40$ ,  $\sigma_G^2 = 1.0$  for a recessive trait  $a = 1.0$  and  $d = -0.5$ .

The  $p$ -values of testing the equality of the coefficients  $\alpha_{1,2}$  and  $\alpha_{2,1}$  are smaller than 0.05 significant level for three markers D4S1540, D7S484 and D16S515, and smaller than 0.0001 for marker D16S289. Hence, our results confirm the findings in Daniel *et al.* (1996) that the three regions are potentially linked to asthma phenotype  $\log_e$  slope.

## 7. DISCUSSION

The objective of this paper is to explore models and test statistics of linkage and association studies between a marker and a quantitative trait locus for sample data of nuclear families. Although the trait values of offspring from unrelated nuclear families can be treated as independent variables, the trait values of multiple offspring of a family are dependent on each other. Hence, the TDT statistics for singleton offspring nuclear families may not be used directly to analyse sample data of multiple offspring nuclear families. One way to handle this problem is to randomly choose an offspring from each nuclear family. However, this method, because of considerable information loss, is less powerful than the mixed model method proposed in this paper.

George *et al.* (1999) and Zhu and Elston (2000, 2001) have proposed very interesting TDT regression models to analyse quantitative family data. Their variance-covariance matrices take into account of the polygenic and environmental effects, ignoring the effects on major gene locus. In this paper, we propose a mixed model whose variance-covariance matrix contains all effects of major gene locus, polygenic loci and environment. The powers of the test statistic  $T$  proposed in this paper is higher than those of George *et al.* (1999), and similar to those of  $T_{ZE1}$  in Zhu and Elston (2000) under moderate disequilibrium.

After comparing the sample sizes and powers between singleton offspring families and sib-pair families, we conclude that the statistic is more powerful than the TDT constructed by randomly choosing

an offspring from each nuclear family. This conclusion is similar to that of Martin *et al.* (1997) for qualitative traits. Moreover, one may use likelihood ratio tests for the linkage and association studies.

The method proposed in this paper can use all data of all offspring of homozygous  $\times$  heterozygous matings, and homozygous offspring of heterozygous  $\times$  heterozygous matings. For the nuclear families that both parents and each offspring have the same heterozygous genotypes  $M_1M_2$ , it is impossible to tell if a parent transmits allele  $M_1$  or  $M_2$  to each of his/her offspring. Hence, the mean and variance–covariance structure is problematic. For this kind of family, one may use the Bayesian method to analyse the data by considering each possible transmission and treating it via a prior probability.

In the case where the sample data have general pedigrees (e.g. more than two generation pedigrees) in addition to the nuclear families, one may use mixed models to analyse the data according to the principles described in this paper. However, one needs to calculate the mean and variance–covariance structure for each family, and then construct appropriate mixed models. Moreover, one may include interesting covariates in the models such as gender and age of the offspring. Using standard statistical packages such as SAS, the analysis should be readily carried out.

In real data analysis, we often encounter multi-allelic markers. One may collapse the multiple marker alleles to form a bi-allelic marker, and apply the theory of the bi-allelic marker for analysis. In the meantime, one may still want to use the multi-allelic marker for a composite analysis since it may capture more information and has higher power (Fan *et al.*, 2002; Sham and Curtis, 1995). One may generalize the mixed model method of this paper for multi-allelic markers. However, it is more involved in terms of notation and theoretical inferences.

#### APPENDIX A

Let us utilize the notation introduced in Section 2 such as  $TQ$ ,  $TH$ ,  $TM$  and  $NM$ . Consider a heterozygous parent with a genotype  $M_1M_2$  at the marker locus  $M$ . First of all, we may calculate the conditional variance of the trait value for the offspring given the transmitted allele  $Q_r$

$$\begin{aligned}\sigma_r^2 &= \text{Var}(Y|TQ = Q_r) = \sum_{s=1}^2 \text{E}[(v + g + G + e - \alpha_r)^2 | Q_r Q_s] q_s \\ &= \sum_{s=1}^2 (\mu_{rs} - \mu_r)^2 q_s + \sigma_G^2 + \sigma_e^2.\end{aligned}$$

Then, we have conditional variance of the trait values for  $i \neq j$ , given that the marker allele  $M_i$  is transmitted and  $M_j$  is not transmitted

$$\begin{aligned}\sigma_{i,j}^2 &= \text{Var}[Y|TM = M_i, NM = M_j] = \text{E}[(Y - \alpha_{i,j})^2 | TM = M_i, NM = M_j] \\ &= \sum_{r=1}^2 \text{E}[(Y - \alpha_{i,j})^2 | TH = Q_r M_i, NM = M_j] \frac{P(Q_r M_i, M_j)}{P(M_i, M_j)} = \Sigma_{ij}^2 + \sigma_G^2 + \sigma_e^2.\end{aligned}$$

#### APPENDIX B

In this Appendix, we calculate the covariance  $\Sigma_{12,12}$  of  $y_l$  ( $l = 1, \dots, k$ ) and  $y_t$  ( $t \neq l, t = 1, \dots, k$ ), and covariance  $\Sigma_{12,21}$  of  $y_l$  ( $l = 1, \dots, k$ ) and  $y_t$  ( $t = k+1, \dots, n$ ) in Figure 2. Without loss of generality, assume that  $k = 2$  and  $n = 3$ . Let  $TM_1$  be the abbreviation of the ‘transmitted marker allele for child 1’, and  $NM_1$  be the abbreviation of the ‘non-transmitted marker allele for child 1’, from the heterozygous father  $M_1M_2$  in Figure 2. Similarly, we define  $TM_i, NM_i, i = 2, 3$ .

Denote  $A = (TM_1 = M_1, NM_1 = M_2, TM_2 = M_1, NM_2 = M_2)$ . Then we have  $P(A) = [2p_1p_2][1/2][1/2] = p_1p_2/2$ . Using the notation in Lange (1997, Chapter 5), let  $S_7, S_8$  and  $S_9$  be three condensed identity states (Jacquard, 1974). Specifically,  $S_7$  is the state where two offspring share two identical major trait locus genes by descent,  $S_8$  is the state where two offspring share one identical major trait locus gene by descent, and  $S_9$  is the state where two offspring share no identical major trait locus gene by descent. Then

$$\begin{aligned}\Sigma_{12,12} &= \text{Cov}(y_1, y_2) = \text{E}[(Y_1 - \alpha_{1,2})(Y_2 - \alpha_{1,2})|A] \\ &= [\text{E}(g_1g_2\mathbf{1}_{(A \cap S_7)}) + \text{E}(g_1g_2\mathbf{1}_{(A \cap S_8)}) + \text{E}(g_1g_2\mathbf{1}_{(A \cap S_9)})]/(p_1p_2/2) - (v - \alpha_{1,2})^2 + \text{E}(G_1G_2),\end{aligned}$$

where  $\text{E}(G_1G_2) = \sigma_G^2/2$  (Amos, 1994; Zhu and Elston, 2001). Let  $S_{7kl}$  be the state where two offspring share two identical trait alleles  $Q_k$  and  $Q_l$  by descent, and  $Q_l$  is from the heterozygous father and  $Q_k$  is from the mother;  $S_{8klr}$  be the state where two offspring share one identical trait allele  $Q_k$  by descent, and the other two trait alleles  $Q_l$  and  $Q_r$  are not identical by descent; and  $S_{9krls}$  be the state where two offspring share no identical trait alleles by descent, and two trait alleles  $Q_l, Q_s$  of the two offspring are from the heterozygous father, and the other two trait alleles  $Q_k, Q_r$  are from the mother. Then

$$\begin{aligned}\Sigma_{12,12} &= \left[ \sum_k \sum_l \mu_{kl}^2 P(A \cap S_{7kl}) + \sum_k \sum_l \sum_r \mu_{kl}\mu_{kr} P(A \cap S_{8klr}) \right. \\ &\quad \left. + \sum_k \sum_l \sum_r \sum_s \mu_{kl}\mu_{rs} P(A \cap S_{9krls}) \right] / (p_1p_2/2) - (v - \alpha_{1,2})^2 + \sigma_G^2/2, \quad (11)\end{aligned}$$

where

$$\begin{aligned}P(A \cap S_{7kl}) &= \frac{q_k}{2} \cdot \left( 2h_{l1}p_2 \frac{1-\theta}{2} \frac{1-\theta}{2} + 2h_{l2}p_1 \frac{\theta}{2} \frac{\theta}{2} \right) \\ &= q_k(h_{l1}p_2(1-\theta)^2 + h_{l2}p_1\theta^2)/4 \\ P(A \cap S_{8klr}) &= \frac{q_lq_r}{2} \cdot \left( 2h_{k1}p_2 \frac{1-\theta}{2} \frac{1-\theta}{2} + 2h_{k2}p_1 \frac{\theta}{2} \frac{\theta}{2} \right) + \frac{q_k}{2} \cdot (h_{l1}h_{r2} + h_{r1}h_{l2})2\theta(1-\theta)/4 \\ &= q_lq_r(h_{k1}p_2(1-\theta)^2 + h_{k2}p_1\theta^2)/4 + q_k(h_{l1}h_{r2} + h_{r1}h_{l2})\theta(1-\theta)/4 \\ P(A \cap S_{9krls}) &= \frac{q_kq_r}{2} \cdot (h_{l1}h_{s2} + h_{s1}h_{l2})2\theta(1-\theta)/4 \\ &= q_kq_r(h_{l1}h_{s2} + h_{s1}h_{l2})\theta(1-\theta)/4.\end{aligned}$$

Similarly, denote  $B = (TM_1 = M_1, NM_1 = M_2, TM_3 = M_2, NM_3 = M_1)$ . We can calculate the conditional covariance of offspring 1 and 3 in Figure 2

$$\begin{aligned}\Sigma_{12,21} &= \Sigma_{21,12} = \text{Cov}(y_1, y_3) \\ &= \left[ \sum_k \sum_l \mu_{kl}^2 P(B \cap S_{7kl}) + \sum_k \sum_l \sum_r \mu_{kl}\mu_{kr} P(B \cap S_{8klr}) \right. \\ &\quad \left. + \sum_k \sum_l \sum_r \sum_s \mu_{kl}\mu_{rs} P(B \cap S_{9krls}) \right] / (p_1p_2/2) - (v - \alpha_{1,2})(v - \alpha_{2,1}) + \sigma_G^2/2, \quad (12)\end{aligned}$$

where

$$\begin{aligned}
P(B \cap S_{7kl}) &= \frac{q_k}{2} \cdot (2h_{l1}p_2 + 2h_{l2}p_1)\theta(1-\theta)/4 \\
&= q_k(h_{l1}p_2 + h_{l2}p_1)\theta(1-\theta)/4 \\
P(B \cap S_{8klr}) &= \frac{q_l q_r}{2} \cdot (2h_{k1}p_2 + 2h_{k2}p_1)\theta(1-\theta)/4 \\
&\quad + \frac{q_k}{2} \cdot (h_{l1}h_{r2} + h_{r1}h_{l2}) \frac{\theta^2 + (1-\theta)^2}{4} \\
&= q_l q_r \cdot (h_{k1}p_2 + h_{k2}p_1)\theta(1-\theta)/4 + q_k \cdot (h_{l1}h_{r2} + h_{r1}h_{l2}) \frac{\theta^2 + (1-\theta)^2}{8} \\
P(B \cap S_{9krls}) &= \frac{q_l q_r}{2} \cdot (h_{l1}h_{s2} + h_{s1}h_{l2}) \frac{\theta^2 + (1-\theta)^2}{4} \\
&= q_k q_r \cdot (h_{l1}h_{s2} + h_{s1}h_{l2}) \frac{\theta^2 + (1-\theta)^2}{8}.
\end{aligned}$$

Assuming that  $\theta = 1/2$  and  $\Delta \neq 0$ , then relations (11) and (12) lead to  $\Sigma_{12,12} = \Sigma_{12,21} = \Sigma_{21,12} = \Sigma_{21,21} |_{\theta=1/2, \Delta \neq 0}$ . Similarly, the conditional covariance  $\Sigma_{12,12} |_{\Delta=0, \theta \neq 1/2}$  is equal to the covariance  $\Sigma_{21,21} |_{\Delta=0, \theta \neq 1/2}$ . Let  $\gamma_{kl} = \mu_{kl} - \mu$ ,  $\mu = \sum_k \sum_l \mu_{kl} q_k q_l$ , and  $\gamma_k = \sum_l \gamma_{kl} q_l$ . If there is no linkage (i.e.  $\theta = 1/2$ ) and no association (i.e.  $h_{ri} = q_r p_i$  for all  $r$  and  $i$ ) between the marker and the trait locus, then one may obtain

$$\begin{aligned}
\Sigma_{12,12} &= \Sigma_{21,12} = \sum_k \sum_l (\gamma_{kl} - \gamma_k - \gamma_l)^2 q_k q_l / 4 + \sum_k \gamma_k^2 q_k + \sigma_G^2 / 2 \\
&= \sigma_d^2 / 4 + \sigma_a^2 / 2 + \sigma_G^2 / 2,
\end{aligned}$$

which is the covariance of trait values between siblings (Almasy and Blangero, 1998; Amos, 1994; Blangero and Almasy, 1997; Lange, 1997).

#### APPENDIX C

In this Appendix, we are going to discuss the asymptotic property of the test statistic  $T$  in Section 4. To simplify the notation, we assume that the data consist of two parts: singleton and sib-pair families. Suppose there are  $k_1$  singleton offspring who receive allele  $M_1$  from their heterozygous parents,  $k_2$  singleton offspring who receive allele  $M_2$  from their heterozygous parents,  $k_{ii}$  ( $i = 1, 2$ ) sib pairs in each of them where both sibs receive allele  $M_i$  from their heterozygous parents, and  $k_{12}$  sib pairs in each of them, one sib receives allele  $M_1$  from his/her heterozygous parent and the other receives allele  $M_2$  from the same heterozygous parent.

Let us denote  $\sigma^\tau = (\sigma_1 = \sigma_{1,2}^2, \sigma_2 = \sigma_{2,1}^2, \sigma_3 = \Sigma_{12,12}, \sigma_4 = \Sigma_{12,21}, \sigma_5 = \Sigma_{21,21})$ . Let  $c, b, f$  be the constants defined in (10). Under the alternative hypothesis, we may get the following expected second

partial derivatives:

$$\begin{aligned} \frac{\partial^2 l_F}{\partial \beta \partial \beta^\tau} &= - \begin{pmatrix} c & -b \\ -b & f \end{pmatrix}, \mathbb{E} \left( \frac{\partial^2 l_F}{\partial \beta \partial \sigma^\tau} \right) = 0, \\ \mathbb{E} \left( \frac{\partial^2 l_F}{\partial \sigma_1^2} \right) &= -\frac{k_1}{2\sigma_1^2} - \frac{k_{11}(\sigma_1^2 + \sigma_3^2)}{(\sigma_1^2 - \sigma_3^2)^2} - \frac{k_{12}\sigma_2^2}{2(\sigma_1\sigma_2 - \sigma_4^2)^2}, \\ \mathbb{E} \left( \frac{\partial^2 l_F}{\partial \sigma_2^2} \right) &= -\frac{k_2}{2\sigma_2^2} - \frac{k_{22}(\sigma_2^2 + \sigma_5^2)}{(\sigma_2^2 - \sigma_5^2)^2} - \frac{k_{12}\sigma_1^2}{2(\sigma_1\sigma_2 - \sigma_4^2)^2}, \\ \mathbb{E} \left( \frac{\partial^2 l_F}{\partial \sigma_3^2} \right) &= -\frac{k_{11}(\sigma_1^2 + \sigma_3^2)}{(\sigma_1^2 - \sigma_3^2)^2}, \mathbb{E} \left( \frac{\partial^2 l_F}{\partial \sigma_4^2} \right) = -\frac{k_{12}(\sigma_1\sigma_2 + \sigma_4^2)}{(\sigma_1\sigma_2 - \sigma_4^2)^2}, \\ \mathbb{E} \left( \frac{\partial^2 l_F}{\partial \sigma_5^2} \right) &= -\frac{k_{22}(\sigma_2^2 + \sigma_5^2)}{(\sigma_2^2 - \sigma_5^2)^2}, \mathbb{E} \left( \frac{\partial^2 l_F}{\partial \sigma_1 \partial \sigma_2} \right) = -\frac{k_{12}\sigma_4^2}{2(\sigma_1\sigma_2 - \sigma_4^2)^2}, \\ \mathbb{E} \left( \frac{\partial^2 l_F}{\partial \sigma_1 \partial \sigma_3} \right) &= \frac{2k_{11}\sigma_1\sigma_3}{(\sigma_1^2 - \sigma_3^2)^2}, \mathbb{E} \left( \frac{\partial^2 l_F}{\partial \sigma_1 \partial \sigma_4} \right) = \frac{k_{12}\sigma_2\sigma_4}{(\sigma_1\sigma_2 - \sigma_4^2)^2}, \\ \mathbb{E} \left( \frac{\partial^2 l_F}{\partial \sigma_1 \partial \sigma_5} \right) &= \mathbb{E} \left( \frac{\partial^2 l_F}{\partial \sigma_2 \partial \sigma_3} \right) = 0, \\ \mathbb{E} \left( \frac{\partial^2 l_F}{\partial \sigma_2 \partial \sigma_4} \right) &= \frac{k_{12}\sigma_1\sigma_4}{(\sigma_1\sigma_2 - \sigma_4^2)^2}, \mathbb{E} \left( \frac{\partial^2 l_F}{\partial \sigma_2 \partial \sigma_5} \right) = \frac{2k_{22}\sigma_2\sigma_5}{(\sigma_2^2 - \sigma_5^2)^2}, \\ \mathbb{E} \left( \frac{\partial^2 l_F}{\partial \sigma_3 \partial \sigma_4} \right) &= \mathbb{E} \left( \frac{\partial^2 l_F}{\partial \sigma_3 \partial \sigma_5} \right) = \mathbb{E} \left( \frac{\partial^2 l_F}{\partial \sigma_4 \partial \sigma_5} \right) = 0. \end{aligned}$$

Assume that  $k_1, k_2, k_{ij} \rightarrow \infty, i, j = 1, 2$ . To make it simple, assume  $k_1 \leq k_2, k_1 \leq k_{ij}$ . Furthermore, assume  $k_2/k_1 \rightarrow k'_2, k_{ij}/k_1 \rightarrow k'_{ij}$  with some positive constants  $k'_2$  and  $k'_{ij}$ . Then we can show that  $-\frac{1}{k_1} \frac{\partial^2 l_F}{\partial \beta \partial \beta^\tau}$  and  $-\frac{1}{k_1} \mathbb{E} \left( \frac{\partial^2 l_F}{\partial \sigma \partial \sigma^\tau} \right)$  are positive definite. Now we are in a position to use the method in Miller (1977) and Pinheiro (1994) according to the theory of Weiss (1971, 1973). Actually, taking  $k_1$  to replace  $v_j$  we can see that the key condition, i.e. Assumption 3.1.7 of Pinheiro (1994, p. 28), holds. Then by the same arguments in Pinheiro (1994, Chapter 3), we can show that  $\sqrt{k_1} \hat{\beta}$  converges to normal in distribution. This implies that the test statistic  $T$  is asymptotically  $\chi^2$  by considering the denominator of  $T$  as the estimate of the variance of the difference  $\hat{\alpha}_{1,2} - \hat{\alpha}_{2,1}$ .

## APPENDIX D

Note that

$$\begin{aligned}
& \sum_{i=1}^I X_i^{\tau} \hat{\Gamma}_i^{-1} X_i \\
&= \begin{pmatrix} \frac{k_1}{\hat{\sigma}_{1,2}^2} & 0 \\ 0 & \frac{k_2}{\hat{\sigma}_{2,1}^2} \end{pmatrix} + k_{11} \begin{pmatrix} 1 & 1 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} \hat{\sigma}_{1,2}^2 & \hat{\Sigma}_{12,12} \\ \hat{\Sigma}_{12,12} & \hat{\sigma}_{1,2}^2 \end{pmatrix}^{-1} \begin{pmatrix} 1 & 0 \\ 1 & 0 \end{pmatrix} \\
&\quad + k_{12} \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} \hat{\sigma}_{1,2}^2 & \hat{\Sigma}_{12,21} \\ \hat{\Sigma}_{12,21} & \hat{\sigma}_{2,1}^2 \end{pmatrix}^{-1} \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \\
&\quad + k_{22} \begin{pmatrix} 0 & 0 \\ 1 & 1 \end{pmatrix} \begin{pmatrix} \hat{\sigma}_{2,1}^2 & \hat{\Sigma}_{21,21} \\ \hat{\Sigma}_{21,21} & \hat{\sigma}_{2,1}^2 \end{pmatrix}^{-1} \begin{pmatrix} 0 & 1 \\ 0 & 1 \end{pmatrix} \\
&= \begin{pmatrix} \hat{c} & -\hat{b} \\ -\hat{b} & \hat{f} \end{pmatrix},
\end{aligned}$$

where  $\hat{c}$ ,  $\hat{b}$  and  $\hat{f}$  are quantities given in relations (9). Plugging the above relation into (7), we may get the test statistic in (8).

## ACKNOWLEDGEMENTS

We thank Dr Joanna Floros for helpful suggestions in writing the paper, and Dr Yuedong Wang for discussions on mixed models. Dr Zeger, one Associate Editor and one referee have raised very good questions and suggestions which led to much improvements of the paper. Dr Pinheiro kindly supplied us his PhD thesis for the investigation of asymptotic property of the test statistics. Ms Jeesun Jung helped us in data analysis. R. Fan was supported partially by a pilot project at Texas A&M University. M. Xiong was supported by NIH grant R01-GM56515, and MH59518.

## REFERENCES

- ABECASIS, G. R., CARDON, L. R. AND COOKSON, W. O. C. (2000). A general test of association for quantitative traits in nuclear families. *American Journal of Human Genetics* **66**, 279–292.
- ALLISON, D. B. (1997). Transmission-disequilibrium tests for quantitative traits. *American Journal of Human Genetics* **60**, 676–690.
- ALLISON, D. B., HEO, M., KAPLAN, N. AND MARTIN, E. R. (1999). Sibling-based tests of linkage and association for quantitative traits. *American Journal of Human Genetics* **64**, 1754–1764.
- ALMASY, L. AND BLANGERO, J. (1998). Multipoint quantitative-trait linkage analysis in general pedigrees. *American Journal of Human Genetics* **62**, 1198–1211.
- AMOS, C. (1994). Robust variance-components approach for assessing genetic linkage in pedigrees. *American Journal of Human Genetics* **54**, 535–543.
- BLANGERO, J. AND ALMASY, L. (1997). Multipoint oligogenic linkage analysis of quantitative traits. *Genetic Epidemiology* **14**, 959–964.
- COOKSON, W. AND ABECASIS, G. (2001). Oxford genome screen for asthma-associated traits. *Genetic Epidemiology* **21** (Suppl 1), S1–S3.

- DANIEL, S. E., BHATTACHARYA, S. AND JAMES, A. *et al.* (1996). A genome-wide search for quantitative trait loci underlying asthma. *Nature* **383**, 247–250.
- FAN, R. Z., FLOROS, J. AND XIONG, M. M. (2002). Models and tests of linkage and association studies of QTL for multi-allele marker loci. *Human Heredity* (in press).
- FULKER, D. W., CHERNY, S. S., SHAM, P. C. AND HEWITT, (1999). Combined linkage and association sib-pair analysis for quantitative traits. *American Journal of Human Genetics* **64**, 259–267.
- GEORGE, V., TIWARI, H. K., ZHU, X. F. AND ELSTON, R. C. (1999). A test of transmission/disequilibrium for quantitative traits in pedigree data, by multiple regression. *American Journal of Human Genetics* **65**, 236–245.
- HARVILLE, D. A. (1974). Bayesian inference for variance components using only error contrasts. *Biometrika* **61**, 383–385.
- HARVILLE, D. A. (1977). Maximum likelihood approaches to variance component estimation and to related problems. *Journal of the American Statistical Association* **72**, 320–338.
- HASEMAN, J. K. AND ELSTON, R. C. (1972). The investigation of linkage between a quantitative trait and a marker locus. *Behaviour Genetics* **2**, 3–19.
- JACQUARD, A. (1974). *The Genetic Structure of Populations*. New York: Springer.
- JENNRICH, R. I. AND SCHLUCHTER, M. D. (1986). Unbalanced repeated-measures models with structured covariance matrices. *Biometrics* **42**, 805–820.
- LAIRD, N. M. AND WARE, J. H. (1982). Random-effects models for longitudinal data. *Biometrics* **38**, 963–974.
- LANGE, K. (1997). *Mathematical and Statistical Methods for Genetic Analysis*. New York: Springer.
- LITTELL, R. C., MILLIKEN, G. A., STROUP, W. W. AND WOLFINGER, R. D. (1996). *SAS System for Mixed Models*. North Carolina: SAS Institute Inc.
- MARTIN, E. R., KAPLAN, N. L. AND WEIR, B. S. (1997). Tests for linkage and association in nuclear families. *American Journal of Human Genetics* **61**, 439–448.
- MARTIN, E. R., MONKS, S. A., WARREN, L. L. AND KAPLAN, N. L. (2000). A test for linkage and association in general pedigrees: the pedigree disequilibrium test. *American Journal of Human Genetics* **67**, 146–154.
- MILLER, J. J. (1977). Asymptotic properties of maximum likelihood estimates in the mixed model of the analysis of variance. *The Annals of Statistics* **5**, 746–762.
- PINHEIRO, J. C. (1994). Topics in Mixed-effects models, PhD thesis, University of Wisconsin-Madison.
- PINHEIRO, J. C. AND BATES, D. M. (2000). *Mixed-Effects Models in S and S-plus*. New York: Springer.
- RABINOWITZ, D. (1997). A transmission disequilibrium test for quantitative trait loci. *Human Heredity* **47**, 342–350.
- ROBINSON, G. K. (1991). That BLUP is a good thing: the estimation of random effects. *Statistical Science* **6**, 15–51.
- SAS/STAT USER'S GUIDE, (1999). *Chapter 41 The Mixed Procedure*, Version 8. SAS Institute Inc.
- SEARLE, S. R., CASELLA, G. AND MCCULLOCH, C. E. (1992). *Variance Components*. New York: Wiley.
- SHAM, P. C., CHERNY, S. S., PURCELL, S. AND HEWITT, J. K. (2000). Power of linkage versus association analysis of quantitative traits, by use of variance-components models, for sibship data. *American Journal of Human Genetics* **66**, 1616–1630.
- SHAM, P. C. AND CURTIS, D. (1995). An extended transmission/disequilibrium test (TDT) for multi-allele marker loci. *Annals of Human Genetics* **59**, 323–336.
- SPIELMAN, R. S. AND EWENS, W. J. (1996). The TDT and other family-based tests for linkage disequilibrium and association. *American Journal of Human Genetics* **59**, 983–989.
- SPIELMAN, R. S., MCGINNIS, R. E. AND EWENS, W. J. (1993). Transmission test for linkage disequilibrium: the

- insulin gene region and insulin-dependent diabetes mellitus (IDDM). *American Journal of Human Genetics* **52**, 506–516.
- WEISS, L. (1971). Asymptotic properties of maximum likelihood estimators in some nonstandard cases I. *Journal of the American Statistical Association* **66**, 345–350.
- WEISS, L. (1973). Asymptotic properties of maximum likelihood estimators in some nonstandard cases II. *Journal of the American Statistical Association* **68**, 428–430.
- XIONG, M. M. AND GUO, S. W. (1997). Fine-Scale genetic mapping based on linkage disequilibrium: Theory and Application. *American Journal of Human Genetics* **60**, 1513–1531.
- XIONG, M. M., KRUSHKAL, J. AND BOERWINKLE, E. (1998). TDT statistics for mapping quantitative loci. *Annals of Human Genetics* **62**, 431–452.
- ZHU, X. F. AND ELSTON, R. C. (2000). Power comparison of regression methods to test quantitative traits for association and linkage. *Genetic Epidemiology* **18**, 322–330.
- ZHU, X. F. AND ELSTON, R. C. (2001). Transmission/disequilibrium tests for quantitative traits. *Genetic Epidemiology* **20**, 57–74.

[Received March 15, 2001; first revision June 28, 2001; second revision August 29, 2001; third revision November 7, 2001; fourth revision December 14, 2001; accepted for publication January 10, 2002]