Genetic Epidemiology 21:243–265 (2001)

Effect of Polygenes on Xiong's Transmission Disequilibrium Test of a QTL in Nuclear Families With Multiple Children

Hong-Wen Deng,[1–3]* Jing Li,[1,2] and Robert R. Recker[1]

[1]*Osteoporosis Research Center, Creighton University, Omaha, Nebraska*
[2]*Department of Biomedical Sciences, Creighton University, Omaha, Nebraska*
[3]*Laboratory of Molecular and Statistical Genetics, College of Life Sciences, Hunan Normal University, Hunan, P.R. China*

The transmission disequilibrium test (TDT), originally developed for mapping disease genes, has recently been extended to identify quantitative trait loci (QTL). For quantitative traits important for human health, generally multiple QTLs are involved. In the investigation of the statistical properties of the TDT, background polygenes (QTLs other than the QTL under test) generally have not been explicitly considered. The effects of background polygenes on the statistical properties of the TDT are thus largely unknown. Investigation of these effects will provide more realistic analyses of the statistical properties of the TDT under biologically plausible situations, and thus provide more accurate guidelines on the application of the TDT in practice. A general TDT (TDT_G) has been developed to test linkage of a QTL in nuclear families that may be composed of more than one heterozygous parent and multiple children. Using the TDT_G as an example, we develop an analytical method to investigate the effects of background polygenes on the power of the TDT. The accuracy of our analytical method is validated by computation simulations. We found that the power of the TDT_G is increased with background polygenes when more than one child is employed in nuclear families, and the effect is stronger with more children per family recruited for study. The power of the TDT_G increases dramatically when the number of children recruited from each nuclear family increases from one to two or from two to three. The type one error rate is not affected by the presence of background

Contract grant sponsor: NIH; Contract grant sponsor: Health Future Foundation; Contract grant sponsor: Hunan Normal University of P.R. China.

*Correspondence to: Hong-Wen Deng, Ph.D., Osteoporosis Research Center, Creighton University, 601 N. 30th St., Suite 6787, Omaha, NE 68131. E-mail: deng@creighton.edu

Received 21 August 2000; Accepted 6 November 2000

polygenes. The results of this study should be of theoretical significance in generalizing the investigation of the TDT to biologically plausible situations with background polygenes. They should also be of practical values in providing guidance on the recruitment of nuclear families with multiple children with the TDT_G.
Genet. Epidemiol. 21:243–265, 2001.

Key words: quantitative trait loci; transmission disequilibrium test; polygene; linkage; association

INTRODUCTION

Complex traits refer to those determined by multiple genetic and environmental factors (and potentially their interactions), whether they are discontinuously distributed complex diseases or continuously distributed quantitative traits. Mapping and identification of genes underlying complex traits, especially those of primary health importance, has been a challenge for geneticists. The challenge is largely due to the limited power of and the large samples required by many currently employed approaches, such as sib pair linkage studies [Risch and Merikangas, 1996]. A powerful approach, the transmission disequilibrium test (TDT), has been developed for identification of genes, originally for diseases [Spielman et al., 1993]. The TDT has been increasingly used to identify genes underlying complex diseases or to detect linkage and/or linkage disequilibrium (association) between markers and such genes [Spielman and Ewens, 1996; Schaid, 1998]. In testing candidate genes for association with complex diseases, the TDT is not plagued by the problem of population admixture or stratification [Ewens and Spielman, 1995; Spielman and Ewens, 1996]. In the presence of association between genotypes and phenotypes, the TDT can be employed to test linkage of candidate genes with complex traits [Spielman and Ewens, 1996; Schaid, 1998]. When markers are at or very close to the genes underlying complex traits, the TDT can be much more powerful than traditional sib pair linkage analyses [Risch and Merikangas, 1996].

In addition to complex diseases, many continuously distributed quantitative traits are of primary clinical and health significance. Examples of such quantitative traits are blood pressure, cholesterol level, obesity, and bone mineral density. Recently, the TDT has been extended to quantitative traits [e.g., Allison, 1997; Rabinowitz, 1997; Xiong et al., 1998; Allison et al., 1999; George et al., 1999; Schaid and Rowland, 1999; Monks and Kaplan, 2000; Abecasis et al., 2000] for identification of quantitative trait loci (QTL). Investigation of the statistical properties of these tests generally assumes absence of QTLs other than the QTL under test [but see Allison et al., 1999]. This is apparently not realistic, as we know that many or almost all those quantitative traits that are of clinical and health significance are polygenic [e.g., Chagnon et al., 1998; Deng et al., 2000b] in that multiple QTLs underlie the trait variation. Throughout this study, we will refer to the QTLs other than the QTL under test as background polygenes.

Ignoring background polygenes in the investigation of the TDT tests that employ only one child [such as the $TDT_{Q1–4}$ of Allison, 1997] is not a problem. However, potential problems may exist with the TDT that may employ multiple children from nuclear families (such as the TDT_{Q5} of Allison [1997] and the TDT_G of Xiong et al. [1998]. With the existence of background polygenes, there will be genetic cor-

relation among sibs due to these polygenes when multiple children are employed in nuclear families [Falconer, 1989]. In other words, the phenotypes of sibs are not independent. How this nonindependence would affect the statistical properties (such as the size, the type one error rate, and the statistical power) of the TDT has not been thoroughly investigated and is generally unknown. Allison et al. [1999] performed some preliminary investigation on the effects of residual variation (that may be due to background polygenes) on the statistical power of the sibling-based TDT test for a QTL. However, the detailed statistical properties (size and power) of the TDT for a QTL, especially those TDT tests for nuclear families with both parents and multiple children, are unknown. For the TDT to be robust (i.e., the size remains at the level specified) and in order to provide accurate guidelines for the application of the TDT in practice when multiple children are involved from nuclear families, the statistical properties under biologically realistic situations with background polygenes should be investigated thoroughly.

A general TDT (TDT_G) developed by Xiong et al. [1998] allows that both parents may be heterozygous and that multiple children from each family can be employed. Although the TDT_G is general in its practical application, its development and investigation implicitly assume that the QTL under test is the only QTL underlying the trait variation.

In this study, we investigate the statistical properties (size and power) of the TDT_G for identification of a QTL in the presence of background polygenes. The effects of background polygenes on the TDT_G test of a QTL are investigated for a range of parameter values. Other issues such as the effect of the number of children from each nuclear family on the power of the TDT_G are also investigated. The investigation is new in that 1) the effect of background polygenes is first explicitly investigated for a TDT test with nuclear families (parents and children, 2) the statistical properties (both the size and the power) are investigated, and 3) the detailed effects of the number of children on the statistical properties are investigated. The results will be discussed for practical applications of the TDT_G.

METHODS

In this section, we introduce the TDT_G test of Xiong et al. [1998]. Then we derive the noncentrality parameter of the TDT_G statistic in the presence of background polygenes. The noncentrality parameter is essential for the analytical computation of the statistical power of the TDT_G. Finally, we perform simulations to validate the accuracy of our analytical power computation. As shown by Xiong et al. [1998], even with multiple children and more than one heterozygous parent in nuclear families, the TDT_G is a valid test of linkage in the presence of population admixture. Therefore, to focus on studying the effects of background polygenes, we assume that the study population is randomly mating so that Hardy-Weinberg equilibrium holds. We also assume a two-allele model at the QTL and marker locus.

TDT_G Test

We assume that there are *n* nuclear families, each with at least one parent being heterozygous for the marker locus under test. Such families will be referred to as informative families. Assume that there are two alleles M and m at the marker locus

under test. For the *i*th ($i = 1,...,n$) nuclear family, we assume that the marker allele M is transmitted to n_{Mi} children from heterozygous parent(s). Let Y denote the phenotypic value of the quantitative trait under study. For the *j*-th child in the set of n_{Mi} children, let Y_{Mij} be his/her phenotypic value. We can denote n_{mi} and Y_{mij} similarly for the allele m. n_{Mi} and n_{mi} can be simply counted based on the genotypes of parents and children. The total number of children receiving M and m alleles from heterozygous parents are, respectively,

$$n_M = \sum_{i=1}^{n} n_{Mi} \text{ and } n_M = \sum_{i=1}^{n} n_{mi}.$$

Then the mean phenotypic values among children who receive M or m alleles from heterozygous parents are, respectively,

$$\bar{Y}_M = \frac{1}{n_m} \sum_{i=1}^{n} \sum_{j=1}^{n_{Mi}} Y_{Mij} \text{ and } \bar{Y}_m = \frac{1}{n_m} \sum_{i=1}^{n} \sum_{j=1}^{n_{mi}} Y_{mij}.$$

Define

$$S^2 = \frac{\sum_{i=1}^{n} \sum_{j=1}^{n_{Mi}} (Y_{mij} - \bar{Y}_M)^2 + \sum_{i=1}^{n} \sum_{j=1}^{n_{mi}} (Y_{mij} - \bar{Y}_m)^2}{n_M + n_m - 2},$$

then the TDT statistic can be computed as:

$$TDT_G = \frac{(\bar{Y}_M - \bar{Y}_m)^2}{\left(\frac{1}{n_M} + \frac{1}{n_m}\right) S^2}, \tag{1}$$

where

$$\left(\frac{1}{n_M} + \frac{1}{n_m}\right) S^2$$

is an unbiased estimator of the variance of $\bar{Y}_M - \bar{Y}_m$ [Xiong et al., 1998]. With large sample sizes, the TDT_G approximately follows a χ^2-distribution with 1 d.f.

To focus on the investigation of the effects of background polygenes on the power of the TDT_G, we consider the situation when the marker is a functional mutation of the QTL under study. The situation when the marker is not at a QTL but is linked to and is in linkage disequilibrium with a QTL is considered in the Appendix. The analytical results of both situations are validated later by our simulations.

Theory With Background Polygenes for the TDT_G

Assume that the QTL locus under study has two alleles, Q and q. Let p be the frequency of the allele Q and $p' = 1 - p$ be the frequency of the allele q. Let a (>0) be the mean (genotypic value) for individuals of the genotype QQ, d the genotypic value of Qq individuals, and -a the genotypic value of qq individuals. d is equal to 0, a, and $-a$, respectively with additive, dominant and recessive genetic effects. Under partial dominant or partial recessive genetic effects, $-a < d < a$ but $d \neq 0$. The additive genetic variance of this locus is $\sigma_A^2 = 2pp'[a + (p' - p)d]^2$, and the dominant genetic variance is $\sigma_D^2 = (2pp'd)^2$ [Falconer, 1989]. The total genetic variance due to this QTL is $\sigma_G^2 = \sigma_A^2 + \sigma_D^2$. We assume that the variance due to all other QTLs (background polygenes, σ_{pg}^2) and all random environmental effects (σ_e^2) is σ_E^2 ($\sigma_E^2 = \sigma_{pg}^2 + \sigma_e^2$). The heritability h^2 due to this QTL is

$$h^2 = \frac{\sigma_G^2}{\sigma_G^2 + \sigma_E^2}.$$

Under a genetic model (such as additive, dominant, or recessive), once three of the four parameters of the h^2, σ_E^2, and a and p at the QTL are given, the fourth parameter can be computed easily [Falconer, 1989]. The phenotypic value of an ith individual in the population is:

$$y_i = \mu + \mu_{g_j} + y_{pg} + e_i,$$

where μ is the mean baseline value of the quantitative trait, μ_{g_j} is the genotypic value at the QTL for the jth genotype, y_{pg} the genotypic value due to the background polygenes, e_i represents a random variable for all environmental effects. μ_{g_j} is equal to a, d, and $-a$, respectively, for the genotypes of g_2(QQ), g_1(Qq), and g_0(qq). Without loss of generality, we can assume that $\mu = 0$. e_i is assumed to follow a normal distribution with mean 0 and variance σ_e^2, i.e.,

$$e_i \sim N(0, \sigma_e^2),$$

where $N(\mu,\sigma^2)$ denotes the probability density function (p.d.f.) for a normal random variable x with mean μ and variance σ^2. We assume that the background polygenes and the QTL under study are in linkage equilibrium and unlinked. It is assumed that the effects (y_{pg}) of the background polygenes are additive and follow a normal distribution with mean μ_{pg} and variance σ_{pg}^2, i.e.,

$$y_{pg} \sim N(\mu_{pg}, \sigma_{pg}^2).$$

Let μ_Q and σ_Q^2 be the mean and variance, respectively, of phenotypic values of the children who receive the Q allele from heterozygous parents. μ_q and σ_q^2 are similarly defined for the q allele. Let n_Q and n_q, respectively, be the numbers of the

children who receive the Q and q alleles from heterozygous parents in informative nuclear families. Given a sample, the noncentrality parameter of the distribution of the statistic TDT_G is [Xiong et al., 1998]:

$$\lambda = \frac{[E(\bar{Y}_Q) - E(\bar{Y}_q)]^2}{Var(\bar{Y}_Q - \bar{Y}_q)} = \frac{(\mu_Q - \mu_q)^2}{Var(\bar{Y}_Q - \bar{Y}_q)}, \tag{2}$$

where $\bar{Y}_Q$ is the mean phenotypic value of the children who receive the Q allele from heterozygous parents in the sample. $\bar{Y}_q$ is similarly defined for the allele q. In the denominator, $Var(\bar{Y}_Q - \bar{Y}_q) = Var(\bar{Y}_Q) + Var(\bar{Y}_q) - 2Cov(\bar{Y}_Q, \bar{Y}_q)$.

$Var(\bar{Y}_Q)$, $Var(\bar{Y}_q)$ and $Cov(\bar{Y}_Q, \bar{Y}_q)$ are functions of σ_Q^2, σ_q^2, n_Q, and n_q, which will be derived in the following.

To compute analytically the statistical power of the TDT_G, λ and thus μ_Q, σ_Q^2, μ_q, σ_q^2, n_Q, and n_q should be derived in terms of the parameters such as p, p', genetic effects (such as a and d at the QTL under study) and σ_{pg}^2. Let g_o, g_f, and g_m, respectively, denote the genotypes of children, fathers, and mothers in informative families. Recalling that $\mu = 0$, within a nuclear family, conditional on the parental genotypes of g_f and g_m, the mean value of all children is

$$\begin{aligned} \mu_1 &= E(Y \mid g_f, g_m) \\ &= \sum_{g_o} E(Y \mid g_o, g_f, g_m) P(g_o \mid g_f, g_m), \\ &= \sum_{g_o} \mu_{g_o} P(g_o \mid g_f, g_m) + \mu_{pg} \end{aligned} \tag{3a}$$

where P denotes probability throughout and Y denotes the phenotypic value. Over all the informative nuclear families, the mean value of all the children is

$$\begin{aligned} \mu_1 &= E(Y) \\ &= \sum_{g_f} \sum_{g_m} P(g_f, g_m) \sum_{g_o} E(Y \mid g_o, g_f, g_m) P(g_o \mid g_f g_m) \\ &= \sum_{g_f} \sum_{g_m} \sum_{g_m} E(Y \mid g_o, g_f, g_m) P(g_o, g_f, g_m) \\ &= \sum_{g_f} \sum_{g_m} \sum_{g_o} \mu_{g_o} P(g_o, g_f, g_m) + \mu_{pg}. \end{aligned} \tag{3b}$$

The phenotypic variance of children within one nuclear family is:

$$\begin{aligned}\sigma_1^2 &= Var(Y \mid g_f, g_m) \\ &= \sum_{g_o=1}^{3} E(Y^2 \mid g_o, g_f g_m) P(g_o \mid g_f, g_m) - (E(Y \mid g_f, g_m))^2 \\ &= \sum_{g_o=1}^{3} (\sigma_{pg}^2 + \sigma_e^2 + \mu_{g_o}^2) P(g_o \mid g_f, g_m) - \mu_1^2 = \sum_{g_o=1}^{3} \mu_{g_o}^2 P(g_o \mid g_f, g_m) + \sigma_{pg}^2 + \sigma_e^2 - \mu_1^2.\end{aligned} \tag{3c}$$

The phenotypic variance of children over all informative nuclear families is:

$$\sigma_2^2 = Var(Y) = \sum_{g_m=1}^{3} \sum_{g_f=1}^{3} \left[\sum_{g_o=1}^{3} [(\mu_{g_o} + \mu_{pg})^2 P(g_o, g_f, g_m)] \right] + \sigma_{pg}^2 + \sigma_e^2 - \mu_2^2. \tag{3d}$$

Let Qq_p denote the event that at least one parent (denoted by the subscript p) is heterozygous, and Q_o denote the event that a heterozygous parent transmits the allele Q to an offspring (denoted by the subscript o). Then, based on the derivative principle for equations 3a–d, the expected phenotypic value of a child who receives the alleles Q and q from heterozygous parents are, respectively:

$$\mu_Q = E(Y \mid Qq_p, Q_o) = \sum_{g_m=1}^{3} \sum_{g_f=1}^{3} \left[\sum_{g_o=1}^{3} (\mu_{g_o} P(g_o, g_f, g_m \mid Qq_p, Q_o)) \right] + \mu_{pg}, \tag{4a}$$

$$\mu_q = E(Y \mid Qq_p, q_o) = \sum_{g_m=1}^{3} \sum_{g_f=1}^{3} \left[\sum_{g_o=1}^{3} (\mu_{g_o} P(g_o, g_f, g_m \mid Qq_p, q_o)) \right] + \mu_{pg}. \tag{4b}$$

The phenotypic variances of a child who receives the alleles Q and q from heterozygous parents are, respectively:

$$\sigma_Q^2 = \sum_{g_m=1}^{3} \sum_{g_f=1}^{3} \left[\sum_{g_o=1}^{3} ((\mu_{g_o} + \mu_{pg})^2 P(g_o, g_f, g_m \mid Qq_p, Q_o)) \right] + \sigma_{pg}^2 + \sigma_e^2 - \mu_Q^2, \tag{4c}$$

$$\sigma_q^2 = \sum_{g_m=1}^{3} \sum_{g_f=1}^{3} \left[\sum_{g_o=1}^{3} ((\mu_{g_o} + \mu_{pg})^2 P(g_o, g_f, g_m \mid Qq_p, q_o)) \right] + \sigma_{pg}^2 + \sigma_e^2 - \mu_q^2, \tag{4d}$$

where $P(g_o, g_f, g_m \mid Qq_p, Q_o)$ is the probability of the genotypes of children and parents, conditional on that a nuclear family has at least one heterozygous parent and the heterozygous parent(s) transmit the Q allele to the child. It can be easily seen that

$$P(g_o, g_f, g_m \mid Qq_p, Q_o) = \frac{P(g_o, g_f, g_m, Qq_p, Q_o)}{\sum_{g_f}\sum_{g_m}\sum_{g_o} P(g_o, g_f, g_m, Qq_p, Q_o)},$$

where $P(g_o, g_f, g_m, Qq_p, Q_o)$ can be computed analytically, given a specific set of genotypes of parents and a child. For example, if the genotypes of father, mother, and their offspring are, respectively, Qq_f, Qq_m, and Qq_o, we have:

$$\begin{aligned} P(g_o, g_f, g_m, Qq_p, Q_o) &= P(Qq_o, Qq_f, Qq_m, Qq_p, Q_o) \\ &= P(Qq_o, Qq_f, Qq_m) \\ &= P(Qq_o \mid Qq_f, Qq_m) P(Qq_f, Qq_m) \\ &= 0.5(2pp')(2pp') = 2p^2 p'^2 \end{aligned}$$

$P(g_o, g_f, g_m \mid Qq_p, q_o)$ is similarly defined adn can be obtained similarly with the approach outlined above for the $P(g_o, g_f, g_m \mid Qq_p, Q_o)$.

Therefore, we can compute μ_Q, μ_q, σ_Q^2, and σ_q^2 analytically as outlined above. With σ_Q^2 and σ_q^2, we can compute $Var(\bar{Y}_Q)$ and $Var(\bar{Y}_q)$. Assume that there are n informative nuclear families in the sample and the ith ($i = 1,...,n$) nuclear family has n_{Qi} children who receive the Q allele from the heterozygous parent(s). Note that a child of the genotype QQ is counted twice if the parents are both heterozygotes (Qq). This is because, in this case, the QQ child receives a Q allele from the father and the other Q allele from the mother. More accurately, n_{Qi} is the number of the Q alleles transmitted to the children in the ith family from heterozygous parent(s). n_{qi} is similarly defined for the q allele in the ith family. Since only those children within one family are correlated due to the background polygenes and children of different randomly ascertained nuclear families are independent, we have:

$$\begin{aligned} Var(\bar{Y}_Q) &= Var\left(\sum_{i=1}^{n}\sum_{j=1}^{n_{Qi}} Y_{Qij} / n_Q\right) \\ &= \frac{\sigma_Q^2}{n_Q} + \frac{1}{n_Q^2}\sum_{i=1}^{n}\sum_{ij_1, ij_2} Cov(Y_{Qij_1}, Y_{Qij_2}), \end{aligned} \tag{5}$$

where ij_1 and ij_2 (ij_1, $ij_2 = 1,...,n_{Qi}$) index two children (the ij_1th and the ij_2th children) in the ith family. $Cov(Y_{Qij_1}, Y_{Qij_2})$ is the covariance of the phenotypic values of the ij_1th and ij_2th child in the ith family. n_Q is the total number of the Q allele passed

from heterozygous parents to children in all nuclear families. Under additive background polygenic effects, the covariance of two different children is:

$$Cov(Y_1, Y_2) = \begin{cases} 0 & \text{(when individuals 1 and 2 are from different families)} \\ \frac{1}{2}\sigma^2_{pg} & \text{(when individuals 1 and 2 are from the same family)} \end{cases}.$$

Note, it is not necessary to assume that the background polygenic effects are additive. When the background polygenic effects are not additive, the general relationship of $Cov(Y_1,Y_2)$ with σ^2_{pg} (and its components) can be found in Falconer [1989]. To derive

$$\sum_{ij_1 ij_2} Cov(Y_{Qij_1}, Y_{Qij_2}),$$

let us consider two mutually exclusive situations: One is that a nuclear family has only one heterozygous parent and the other is that a nuclear family has two heterozygous parents. In the first situation, it is relatively simple and we need only to consider the covariance of every two different children within each family. This is because a child may at most receive a Q allele from the one heterozygous parent in the family and the ij_1th and the ij_2th child ($ij_1 \neq ij_2$) cannot be the same individual. However, for the second situation it is different. The difference lies in that when both parents are hterozygous, a child can receive two Q alleles, one from each parent with a probability of 0.25. In this case, a QQ child is counted twice in the total number of the n_{Qi}, and then $Cov(Y_{Qij_1},Y_{Qij_2})$ is equal to $Var(Y_{QQ})$, because in this case the ij_1 and ij_2 are the same person. Therefore, the total number of covariance between children who receive a Q allele in the ith family is $0.5\, n_{Qi}(n_{Qi} - 1)$, in which there are $0.25n_{Qi}$ variances due to those children (QQ) who receive one Q allele from the mother and the other Q allele from the father and who is counted twice in the counting of n_{Qi}. Let the number of children in each nuclear family be J, the same for all families. J is also the expected number of n_{Qi} [This is simply because $n_{Qi} + n_{qi} = 2J$ and $E(n_{Qi}) = E(n_{qi})$].

Conditional on that at least one parent in the family is heterozygote, the probabilities that one and only one parent is heterozygote or that both parents are heterozygotes are, respectively:

$$P(\text{one parent heterozygous} \mid Qq_p = \frac{2(2pp')(p^2 + p')^2}{2(2pp')(p^2 + p'^2) + 2pp'(2pp')},$$

$$P(\text{two parents heterozygous} \mid Qq_p = \frac{2pp'(2pp')}{2(2pp')(p^2 + p'^2) + 2pp'(2pp')}.$$

Therefore, we have

$$\begin{aligned}
&\sum_{j_1,j_2} Cov(Y_{Qij_1}, Y_{Qij_2}) \\
&= P(\text{only one heterozygous parent} \mid \mathrm{Qq_p}) \sum_{j_1,j_2,one} Cov(Y_{Qij_1}, Y_{Qij_2}) \\
&+ P(\text{two heterozygous parents} \mid \mathrm{Qq_p}) \sum_{j_1,j_2,two} Cov(Y_{Qij_1}, Y_{Qij_2}) \\
&= \frac{2(2pp')(p^2+p'^2)}{1-(p^2+p'^2)^2} 0.5JP(Q_o \mid Qq_p(JP(Q_o \mid Qq_p)-1)Cov(Fullsibs) \\
&+ \frac{2pp'(2pp')}{1-(p^2+p'^2)^2}(0.25JVar(Y_{QQ}) + (0.5J(J-1)-0.25J)Cov(Fullsibs) \\
&= \frac{pp'}{1-(p^2+p'^2)^2}[0.5J(J-2)(p^2+p'^2)Cov(Fullsibs) \\
&+ Jp_Qp_qVar(Y_{QQ}) + p_Qp_q(2J^2-3J)Cov(Fullsibs)] \\
&= \frac{Jpp'}{1-(p^2+p'^2)^2}[(0.5(J-2)(p^2+p'^2)+(2J-3)pp')(0.25s^2_{pg}) + pp'(\sigma^2_e+\sigma^2_{pg})]
\end{aligned} \tag{6}$$

where by definition, $Var(Y_{QQ}) = \sigma^2_e + \sigma^2_{pg}$, and

$$\sum_{j_1,j_2,one} Cov(Y_{Qij}, Y_{Qij_2}) \text{ and } \sum_{j_1,j_2,two} Cov(Y_{Qij_1}, Y_{Qij_2})$$

account for the situations when nuclear families have one and two heterozygous parents, respectively.

Similarly, we can derive that:

$$Var\bar{Y}_q) = \frac{\sigma^2_q}{n_q} + \frac{n}{n^2_q}\frac{Jpp'}{[1-(p^2+p'^2)^2]}\left[(0.5(J-2)(p^2+p'^2)+(2J-3)pp')(0.25\sigma^2_{pg}) + pp'(\sigma^2_e+\sigma^2_{pg})\right] \tag{7}$$

where n_q is the total number of q alleles passed from heterozygous parents to children in the sampled nuclear families.

Finally, we need to derive the covariance of random variables $\bar{Y}_Q$ and $\bar{Y}_q$. Noting that children from different families are independent, we have:

$$\begin{aligned} Cov\bar{Y}_Q, \bar{Y}_q) &= Cov\left(\sum_{i=1}^{n} \sum_{j_1=1}^{n_{Qi}} Y_{Qij_1} / n_Q, \sum_{i=1}^{n} \sum_{j_1=1}^{n_{qi}} Y_{qij_2} / n_q \right) \\ &= \sum_{i=1}^{n} Cov\left(\sum_{j_1=1}^{n_{Qi}} Y_{Qij_1} / n_Q, \sum_{j_2=1}^{n_{qi}} Y_{qij_2} / n_q \right) \\ &= \frac{1}{n_Q n_q} \sum_{i=1}^{n} \sum_{j_1, j_2} Cov(Y_{Qij_1}, Y_{qij_2}) \end{aligned} \tag{8}$$

Similar to the derivation for the

$$\sum_{ij_1, ij_2} Cov(Y_{Qij_1}, Y_{Qij_2})$$

in equation 5, to count the number of covariances [$Cov(Y_{Qij_1}, Y_{Qij_2})$] between two children who receive the Q and q alleles, respectively, from heterozygous parents in one family, we again consider two situations. The first is that the nuclear families have one and only one parent being heterozygous. The second situation is that the nuclear families have two heterozygous parents. Under the null hypothesis, in the first situation the number of children who receive the Q allele is expected to be equal to the number of children who receive the q allele. No child can receive both the Q and the q alleles from the heterozygous parent at the same time. The ij_1th and the ij_2th children ($ij_1 \neq ij_2$) cannot be the same individual. However, for the second situation, it is more complex and will be considered further for several cases depending upon the genotypes of the two children under consideration. For the first case, consider the covariances of children of genotypes QQ and Qq, or QQ and qq, or Qq and qq, or Qq_1 and Qq_2. For the Qq_1 and Qq_2, the index 1 and 2 in the subscripts indicate that the two children are different individuals although they have the same genotype Qq. In this first case, the two children who receive the Q or q alleles are different individuals counted in the n_{Qi} and n_{qi}. For the second case, consider the variance of children of genotypes Qq. In this case, a child of the genotype Qq is counted twice, once in the n_{Qi} and once in the n_{qi}, hence $Cov(Y_{Qij_1}, Y_{qij_2}) = Var(Y_{Qq}) = \sigma_e^2 + \sigma_{pg}^2$. Let the number of children in each nuclear family be J. We have:

$$
\begin{aligned}
&\sum_{j_1,j_2} Cov(Y_{Qij_1}, Y_{qij_2}) = P(\text{only one parent heterozygous} \mid \mathrm{Q}q_P) \sum_{j_1 j_2, one} Cov(Y_{Qij_1}, Y_{qij_2}) \\
&+P(\text{both parents heterozygous} \mid \mathrm{Q}q_p) \sum_{j_1, j_2, two} Cov(Y_{Qij_1}, Y_{qij_2}) \\
&= \frac{2(pp')(p^2+p'^2)}{1-(p^2+p'^2)^2}(0.5J)^2 Cov(Fullsib) \\
&+\frac{2pp'(2pp')}{1-(p^2+p'^2)^2}\{[JP(QQ_o \mid Qq\times Qq)JP(Qq_o \mid Qq\times Qq) \\
&+JP(QQ_o \mid Qq\times Qq)JP(qq_o \mid Qq\times Qq) + JP(Qq_o \mid Qq\times Qq)JP(qq_o \mid Qq\times Qq) \\
&+0.5JO(Qq_o \mid Qq\times Qq)(JP(Qq_o \mid Qq\times Qq)-1)]Cov(Fullsib) + JP(Qq_o \mid Qq\times Qq)Var(Y_{Qq})\} \\
&= \frac{Jpp'}{1-(p^2+p'^2)^2}\left[0.25(p^2+p'^2)J\sigma_{pg}^2 + pp'\left(0.25(1.75J-1)\sigma_{pg}^2 + 0.5(\sigma_e^2+\sigma_{pg}^2)\right)\right]
\end{aligned}
\tag{9}
$$

Finally, we need to derive n_Q and n_q as functions of population parameters. Let us denote the total number of families needed to be screened to recruit n informative families as N_s, then it can be seen easily that

$$n = N_S[1-(p^2+p'^2)^2 = N_S[2(2pp')(p^2+p'^2)+2pp'(2pp')]. \tag{10a}$$

Therefore, the expected total number of heterozygous parents n_H in the n informative families is:

$$
\begin{aligned}
E(n_H) &= \frac{2n(2pp')(p^2+p'^2)}{2(2pp')(p^2+p'^2)+2pp'(2pp')} \\
&\quad + \frac{2n(2pp')(2pp')}{2(2pp')(p^2+p'^2)+2pp'(2pp')} \\
&= \frac{n}{pp'+p^2+p'^2} = \frac{n}{1\text{-}pp'}
\end{aligned}
\tag{10b}
$$

For a sample of nuclear families each having J children, the expected number of n_Q and n_q are:

$$E(n_Q) = E(n_q) = JE(n_H)/2 \tag{10c}$$

Specifying a significance level (α) and a statistical power (η), we can, by the aid of some statistical software package [e.g., Wolfram, 1996], obtain the value for the noncentrality parameter λ for the $\mathrm{TDT_G}$ statistic which approximately follows a χ^2-distribution with 1 d.f. With the λ value and the equations 2, 4–10, we can compute required sample sizes N_s, and n for specified α and η given parameter values (p, p',a, d, σ_{pg}^2, and J). A computer program for the analytical power computation is available from the authors upon request.

Computer Simulations

To validate the above derivations and our analytical power computation for the TDT_G in the presence of background polygenes, we performed computer simulations for a range of parameter values. The comparison of simulation and analytical results can provide a mechanism to crosscheck and validate these results. In the absence of segregation distortion, random mating populations are simulated, in which p, p', a, d, and h^2 due to the QTL, σ^2_{pg}, and σ^2_e are specified. When the marker locus is not a QTL (the analytical investigation of this case is presented in the Appendix), the marker genotype frequency (f), the degree of linkage disequilibrium (δ), and the recombination rate (θ) between the marker and the QTL are also specified. With the parameters specified, genotypes of parents of nuclear families are first simulated. Only for those informative nuclear families with at least one parent heterozygous at the marker locus, J children′ genotypes and phenotypes are simulated. The genotypes of children are simulated according to random transmission from parents to children under the null hypothesis. Once the genotypes of children are simulated, their phenotypes are simulated as described earlier [Deng et al., 2000a]. Only informative nuclear families are employed for analyses (equation 1). The effect of background polygenes is indexed by their heritability (h^2_{pg}). The background polygenes are simulated by 10 QTLs, each with the same small additive effects so that at each of such QTLs, the heritability $h^2_1 = h^2_{pg}/10$, the frequency of the allele with higher trait values $p_1 = 0.7$, and the genotypic effect a_1 at these QTL can be easily determined by h^2_1 and p_1 [Falconer, 1989].

For a desired statistical power η and a specified significance level α, we first compute the sample size (n) of informative nuclear families (each with J children) needed by our analytical power computation method. Then nuclear families each with J children are simulated. The TDT_G is applied to the n nuclear families. When a QTL or a marker locus that is linked to and is in linkage disequilibrium with this QTL is simulated, the simulated statistical power is the proportion of times that the TDT_G analyses is significant in the simulations (10,000 times unless otherwise specified) performed. The simulated statistical power (η') can be compared with the specified level of η in the analytical power computation. The closer the η' to the η, the more accurate is our analytical power computation. Once our analytical power computation is validated by simulations, the investigation of the power of the TDT_G under various degrees of background polygenic effects is conducted by our analytical method. To validate the TDT_G in the presence of background polygenes, under a specified significance level α, we also examine the size (the type I error rate) in simulations (α') with a marker locus that is not linked to and/or is in linkage equilibrium with a QTL. The agreement of the simulated α' and the specified α would validate the TDT_G analyses in that they have correct levels of type I error rate.

RESULTS

Accuracy of Our Analytical Power Computation

Table I presents some representative data of our extensive simulation studies for a range of parameter values for the situations when the marker is a QTL and when the marker is not a QTL but is linked to and is in linkage disequilibrium with a QTL. It can be seen that, for all the three typical models of genetic effects (recessive, additive,

TABLE I. Accuracy of the Analytical Power Computation and the Validity of the TDT_G With Background Polygenes*

Genetic effects	h^2_{pg}	p	$n(\eta')$ (marker is a QTL)	$n(\eta')$ (marker is not a QTL)	α' ($\alpha = 0.05$)
Recessive	0.0	0.3	331 (0.79)	773 (0.81)	0.050
		0.5	244 (0.80)	482 (0.79)	0.053
		0.7	170 (0.80)	958 (0.82)	0.051
	0.3	0.3	328 (0.78)	757 (0.80)	0.050
		0.5	244 (0.80)	488 (0.80)	0.051
		0.7	168 (0.79)	923 (0.80)	0.052
	0.6	0.3	310 (0.76)	718 (0.76)	0.051
		0.5	231 (0.77)	462 (0.76)	0.048
		0.7	158 (0.76)	874 (0.76)	0.053
Additive	0.0	0.3	141 (0.80)	347 (0.81)	0.054
		0.5	160 (0.80)	315 (0.79)	0.057
		0.7	141 (0.80)	784 (0.83)	0.052
	0.3	0.3	139 (0.79)	335 (0.80)	0.052
		0.5	160 (0.80)	320 (0.80)	0.051
		0.7	139 (0.78)	758 (0.80)	0.050
	0.6	0.3	131 (0.76)	317 (0.76)	0.053
		0.5	151 (0.76)	303 (0.77)	0.052
		0.7	131 (0.75)	717 (0.77)	0.053
Dominant	0.0	0.3	170 (0.79)	420 (0.82)	0.055
		0.5	244 (0.79)	472 (0.79)	0.050
		0.7	331 (0.78)	1693 (0.82)	0.057
	0.3	0.3	168 (0.79)	404 (0.80)	0.050
		0.5	244 (0.79)	482 (0.80)	0.050
		0.7	328 (0.78)	1650 (0.80)	0.051
	0.6	0.3	158 (0.76)	382 (0.76)	0.051
		0.5	231 (0.76)	456 (0.76)	0.050
		0.7	310 (0.74)	1563 (0.77)	0.049

*n is the number of informative families (with at least one heterozygous parent) needed to achieve 80% power (η) with $\alpha = 10^{-4}$ computed by our analytical methods and η' is the power obtained by 10,000 repeated simulations with the sample size n. In the investigation for this table, two children are sampled for each nuclear family. At the QTL under test, p is specified, $h^2 = 0.1$, and a can be computed from the specified p and h^2 and the genetic effects (recessive, additive, and dominant). h^2_{pg} is the heritability due to background polygenes that are simulated by 10 unlinked QTLs each with $h^2_1 = h^2_{pg}/10$ and $p_1 = 0.7$. When the marker polymorphism is not the functional polymorphism of a QTL, $f_M = 0.4$, $c = 0.02$, $\delta = 0.9\delta_{\max} \cdot \delta_{\max}$ is the maximum linkage disequilibrium between the marker and the QTL in a population, $\delta_{\max}$ is the minimum of p/m and p'/f_M, where M and m are the two alleles at the marker locus [Deng et al., 2000a]. α' is the empirical size (type I error rate) for the TDT_G test obtained from 10,000 repeated simulations when the marker is not a QTL and/or is not linked to a QTL, and/or is not in linkage disequilibrium with a QTL. It is the proportion of the times that the TDT_G test is significant under the specified significance level of α (= 0.05). The significance level of $\alpha = 0.05$ is chosen to avoid unnecessary excessive simulations for the α at much lower levels such as $\alpha = 10^{-4}$.

and dominant), the sample sizes (n) computed from our analytical method under a specified statistical power (η), if employed in computer simulations, can yield the simulated statistical power (η') that is very close to η. This is true when the marker locus is a QTL and when the marker locus is not a QTL but is linked to and is in linkage disequilibrium with a QTL. Therefore, the accuracy of our analytical derivation and the power computation for the TDT_G is validated by our computer simulations.

Validity of the TDT_G in the Presence of Background Polygenes

The last column of Table I presents the results of the simulated significance level α' in the presence of background polygenes under the null hypothesis that the marker locus is not linked to and/or not in linkage disequilibrium with a QTL. It can be seen that, for a range of parameters simulated and under all the models of genetic effects investigated, the simulated significance level is very close to the specified significance level of $\alpha = 0.05$, even when the background polygenes account for as large as 60% of phenotypic variation (i.e., $h^2_{pg} = 0.6$). Therefore, the TDT_G is valid and robust in the presence of background polygenes in that it can ensure the significance level achieved in practice is the same as that specified in the testing.

Effects of Background Polygenes on the TDT_G

The investigation of the effect of background polygenes is conducted by comparing the power of the TDT_G under the situations with and without background polygenes (Tables I and II), and under the situations with different effects of background polygenes as reflected by h^2_{pg} (Table I). The investigation in Tables I and II assumes two children from each nuclear family. The effect of the number of children per nuclear family on the power of the TDT_G is investigated later in Figs. 1 and 2. The power is reflected by the number (n) of informative families required for the TDT_G analyses (Table I) or the number (N_s) of families needed to be screened (Table II) in order to recruit n informative families to achieve a certain statistical power by the TDT_G. It can be seen (Tables I and II, Fig. 1) that, with background polygenes,

TABLE II. Sample Sized (Ns) Needed to Be Screened by the TDT_G in the Presence and Absence of Background Polygenes*

	Recessive		Additive		Dominant	
p	Absent	Present	Absent	Present	Absent	Present
$h^2 = 0.05$						
0.3	984	941	437	417	527	504
0.5	660	642	437	424	660	642
0.7	527	504	437	417	984	941
$h^2 = 0.1$						
0.3	499	476	213	201	255	242
0.5	325	314	213	205	325	314
0.7	255	242	213	201	499	476
$h^2 = 0.2$						
0.3	256	242	101	94	120	113
0.5	157	150	101	96	157	150
0.7	120	113	101	94	256	242
$h^2 = 0.3$						
0.3	175	165	63	58	75	69
0.5	101	96	63	60	101	96
0.7	75	69	63	58	175	165

*In the first column, p and h^2 are, respectively, the frequency of the allele Q and the heritability for the QTL under test. The numbers under the column Absent are the number of families that is needed to be screened to achieve 80% power with $\alpha = 10^{-4}$ when the background polygenes are absent. The numbers under the column Present are the number of families that is needed to be screened to achieve 80% power with $\alpha = 10^{-4}$ when the background polygenes are present ($h^2_{pg} = 0.5$) and accounted for. The marker locus is located at a QTL.

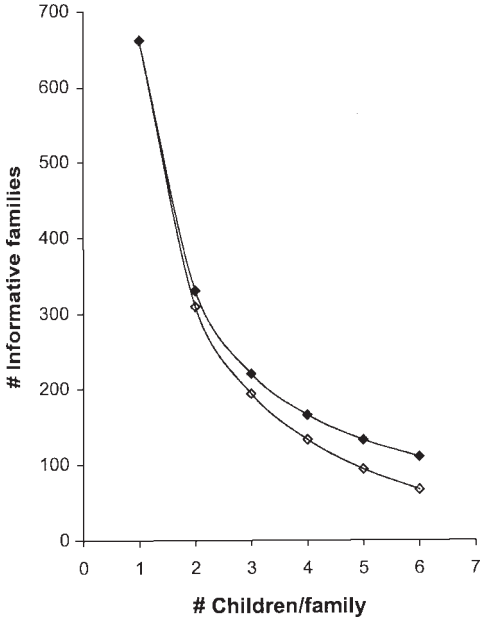

Fig. 1. The effect of the number of children per nuclear family on the power of the TDT_G with and without background polygenes. In Figs. 1 and 2, the power of the TDT_G is reflected by the number n of informative nuclear families required in order to achieve the statistical power of 0.8 with $\alpha = 0.0001$. The marker is at the QTL under study, $a = 1$, $p = 0.7$, and $h^2 = 0.1$ for the TDT under test, and $h^2_{pg} = 0.6$. The filled diamonds represent the n required without background polygenes, and the open diamonds represent the n required with background polygenes. Dominant genetic effect at the QTL is assumed. In Figs. 1 and 2, the marker locus is at a QTL.

the power of the TDT_G is increased. The power increase is slight when only two children are recruited per family. The power increases with increasing effects of background polygenes as reflected by a smaller sample size required under a larger h^2_{pg} (Table I).

Let TDT_2 denote the TDT_G test when each nuclear family has two children. It is noted that, in the absence of background polygenes, the sample sizes needed to be screened in order to achieve a specified power by the TDT_2 are substantially less than those given in tables 1–3 of Xiong et al. [1998] for the same corresponding parameters. The analytical derivation for the power computation of Xiong et al. [1998] when the marker is a QTL is correct and is the same as our own for the case when there is no background polygenes. However, the numerical numbers given in tables 1–3 in Xiong et al. [1998] seem to be incorrect. There are no simulations to confirm the numerical results derived from the analytical computations in Xiong et al. [1998]. Our results are all confirmed by cross-validating the results of simulations and analytical computations and should be correct. In addition, it is actually straightforward to re-compute the numbers in tables 1–3 in Xiong et al. [1998] using their analytical method and it can be shown that the numerical results in their tables 1–3 are not correct. For example, for an additive model, the noncentrality parameter for the TDT_1 (with only one child per family) given by Equation 10 in Xiong et al. [1998] can be easily shown to be the same as our result when only one child per family is recruited and $\lambda = N_s h^2(2 - h^2)$. Therefore, given the heritability of 0.1 and a significant level 0.0001, to achieve 0.8 power, the theoretical noncentrality parameter value is $\lambda = 22.4$. Then the number of nuclear families (with one child) needed to be screened N_s is 426. For a family with two children, N_s is 213 in the absence of background polygenes and is the same as that given in our Table II, whereas in Table I of Xiong et al. [1998], the number is 1,194. Their error is probably due to incorrect critical values (corresponding to the specified power and significance levels) employed.

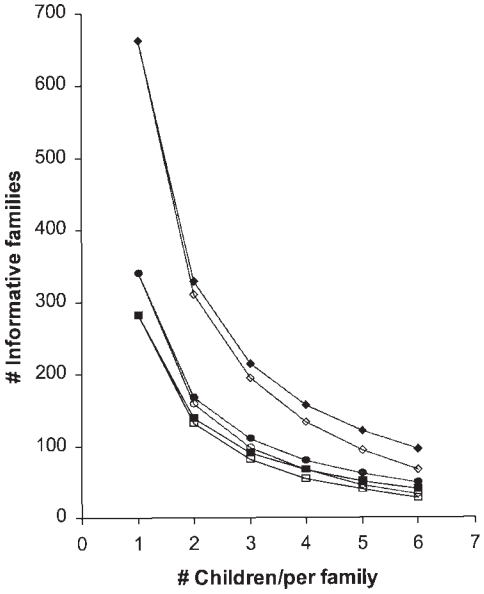

Fig. 2. The effect of the number of children per nuclear family on the power of the TDT_G under various models of genetic effect at the QTL under test and under various h^2_{pg}. The marker is at the QTL under study, $a = 1$, $p = 0.7$, and $h^2 = 0.1$ for the TDT under test. The filled dots represent $h^2_{pg} = 0.3$, and the open dots represent $h^2_{pg} = 0.6$. Diamonds, dominant genetic effects; squares, additive effects; circles, recessive effects.

In the absence of background polygenes, the trends of our numerical results (TDT_2 in Table II) are qualitatively consistent with those given in tables 1–3 of Xiong et al. [1998] as reflected by the following. N_s does not change with the allele frequency p of the allele Q at the QTL under additive genetic effects. N_s decreases with p under recessive genetic effects and increases with p under dominant genetic effects. In the presence of background polygenes (Tables I and II), the trend is the same qualitatively as the case without background polygenes under recessive and dominant genetic effects (Table II). However, different from the case without background polygenes, under additive genetic effects, n decreases with an increasing h^2_{pg} (Table I) and N_s is influenced by p with background polygenes (Table II).

Effect of the Number of Children Within Nuclear Families on the TDT_G Power in the Presence of Background Polygenes

Although the TDT_G was proposed as a general TDT for QTL identification that may allow for multiple children from each nuclear family, how the number of children from each nuclear family would affect the TDT_G power is unclear. This is especially true in the presence of background polygenes. We compare the relationship of the TDT_G power with the number of children from each nuclear family in the presence and absence of background polygenes (Fig. 1). Although Fig. 1 only presents results for the dominant genetic model, the results not shown for additive and recessive effects are the same. In addition, we also investigate the relationship of the TDT_G power with the number of children from each nuclear family for biologically plausible situations with background polygenes under various genetic models and different effects of background polygenes (Fig. 2).

In Figures 1 and 2, the TDT_G power is reflected by the number (n) of informative familes required to achieve a certain statistical power. The smaller the n, the better and the more powerful is the situation under consideration. It can be seen (Fig. 1)

that the difference of the power of the TDT_G in the presence and absence of background polygenes increases with an increasing number of children recruited from each family. As expected, when only one child is recruited per family, the TDT_G power is not influenced by the presence of background polygenes. However, for example, for the parameter values investigated when six children are recruited per family, to achieve 80% power for $\alpha = 0.0001$, it would require 110 families without the background polygenes and only 67 nuclear families with the background polygenes. Therefore, although the effect of background polygenes on the TDT is only slight when only two children are recruited per family as revealed in Tables I and II, the difference can be large and significant when more children are recruited per family. Figure 2 illustrates the increase of the power of the TDT_G with more children per family recruited and with large h^2_{pg} under various genetic models. It is apparent from Figs. 1 and 2 that when the number of children from each nuclear family increases, the TDT_G power increases as reflected by the decrease in the required number of informative families (n). n decreases in a decelerated fashion when the number of children per family increases and n decreases most dramatically when the number of children per nuclear family changes from one to two, or from two to three. The rate of decrease of n diminishes with an increasing number of children per nuclear family. Therefore, it seems that the effort to recruit nuclear families with 2 or 3 children would be most fruitful in terms of the power increase per additional child recruited.

DISCUSSION

In this study, we develop an analytical method to compute the power of the TDT_G of Xiong et al. [1998] under various degrees of the contribution of background polygenes to phenotypic variation. The accuracy of the analytical method is validated by computation simulations. A computer program (in Visual C++) for implementing the power computation of the TDT_G in the presence or absence of background polygenes is available from the authors upon request. In addition, this computer program contains a module for obtaining simulated power of the TDT_G. We found that the power of the TDT_G is affected by background polygenes when multiple children are employed in nuclear families. The effect is more significant with more children recruited from each nuclear family. The power of the TDT_G increases dramatically when the number of children recruited from each nuclear family increases from one to two or from two to three.

The results of this study should be of theoretical significance in generalizing the investigation of the TDT to biologically plausible situations with background polygenes. Although we employ the TDT_G as an example of the TDT for investigation, the principle and the approach may apply to other appropriate TDT tests, such as the TDT_{Q5} of Allison [1997]. As shown by our work, the consideration of background polygenes introduces much complexity in the analytical investigation of the statistical power of the TDT_G under various parameters so that we need to resort to computer programming to implement our analytical computation approach. However, since almost all quantitative traits that are of primary clinical and health significance in humans are likely polygenic, incorporation of background polygenes into the investigation of the statistical properties of the TDT_G is necessary. In addition to the theoretical and simulation investigation of the statistical power of the TDT_G, we also investigated the

validity of the TDT_G as a test of linkage with multiple children in the presence of background polygenes by simulations. The validity of the TDT_G as a test of linkage with multiple children in the presence of background polygenes is investigated by comparing its simulated type I error with the pre-specified type I error for testing in simulations for a range of parameter values. The close agreement between the simulated and pre-specified significance levels validates the TDT_G as a test of linkage of QTL with multiple children in the presence of background polygenes. The results given are accurate and valid as they have been cross-checked by both our analytical computation method and computer simulations. By investigating some common sets of parameter values with Xiong et al. [1998], we found that the numerical values given in their tables 1–3 for the TDT_2 are not correct and are overestimates of the true values. However, their analytical derivations for the case when the marker is a QTL are largely correct, although they did not perform simulations for confirmation. Their analytical derivation is only approximate for the case when the marker is not a QTL but is linked to and is in linkage disequilibrium with a QTL.

Our results should also be of practical value in providing guidance on the recruitment of nuclear families with multiple children. Nuclear families with multiple children are valuable and can increase the power of QTL identification and reducing the total number of individuals that need to be genotyped and phenotyped. For example (Fig. 2), to achieve 80% power at a significance level of 0.0001 under dominant genetic model with $h^2_{pg} = 0.3$, to detect a QTL with $h^2 = 0.1$, it would require 662 informative nuclear families with one child, 328 nuclear families with two children, 215 families with three children, 157 families with four children, and 121 families with five children. The total individuals that need to be phenotyped for these types of nuclear families are, respectively, 662, 656, 645, 628, and 605. The total number of individuals that need to be genotyped for these types of informative nuclear families are, respectively, 1,986, 1,312, 1,075, 942, and 847. Therefore, recruitment of nuclear families with five children is more efficient than nuclear families with fewe children with regard to the genotyping and phenotyping amount involved. However, the efficiency with regard to genotyping and phenotyping may need to be considered jointly with the availability of nuclear families with more children and the difficulty of recruiting more children from nuclear families in order to maximize the overall efficiency of the project as a whole. This kind of consideration may be tailored to specific situations of individual investigators for the availability and costs of recruitment of families with multiple children with the aid of our computer program that is available from authors upon request.

ACKNOWLEDGMENTS

This study was partly supported by a graduate student tuition waiver to J.L. from Creighton University. Helpful comments from two anonymous reviewers and especially those from Professor D.J. Schaid greatly improved the manuscript.

REFERENCES

Abecasis GR, Cardon LR, Cookson WO. 2000. A general test of association for quantitative traits in nuclear families. Am J Hum Genet 66:279–92.

Allison DB. 1997. Transmission-disequilibrium tests for quantitative traits. Am J Hum Genet 60:676–90.

Allison DB, Heo M, Kaplan N, Martin ER. 1999. Sibling-based tests of linkage and association for quantitative traits. Am J Hum Genet 64:1754–63.

Chagnon YC, Perusse L, Bouchard C. 1998. The human obesity gene map: the 1997 update. Obes Res 1998;5:76–92.

Deng HW, Chen WM, Recker RR. 2000a. QTL fine mapping by measuring and testing for Hardy-Weinberg and linkage disequilibrium at a series of linked marker loci in extreme samples of populations. Am J Hum Genet 66:1027–45.

Deng H-W, Chen W-M, Convey T, Davies M, Deng HY, Recker RR. 2000b. Determination of BMD at hip and spine by genetic and life-style factors. Genetics Epidemiology 19:160–77.

Ewens WJ, Spielman RS. 1995. The transmission/disequilibrium test: history, subdivision, and admixture. Am J Hum Genet 57:455–64.

Falconer DS. 1989. Introduction to Quantitative Genetics. Longman, England.

George V, Tiwari HK, Zhu X, Elston RC. 1999. A test of transmission/disequilibrium for quantitative traits in pedigree data, by multiple regression. Am J Hum Genet 65:236–45.

Monks SA, Kaplan NL. 2000. Removing the sampling restrictions from family-based tests of association for a quantitative-trait locus. Am J Hum Genet 66:576–92.

Risch N, Merikangas K. 1996. The future of genetic studies of complex human diseases. Science 273:1516–7.

Rabinowitz D. 1997. A transmission disequilibrium test for quantitative trait loci. Hum Hered 47:342–50.

Schaid DJ. 1998. Transmission disequilibrium, family controls, and great expectations. Am J Hum Genet 63:935–41.

Schaid DJ, Rowland CM. 1999. Quantitative trait transmission disequilibrium test: allowance for missing parents. Genet Epidemiol 17(suppl 1):S307–12.

Spielman RS, Ewens WJ. 1996. The TDT and other family-based tests for linkage disequilibrium and association. Am J Hum Genet 59:983–9.

Spielman RS, McGinnis RE, Ewens WJ. 1993. Transmission test for linkage disequilibrium: the insulin gene region and insulin-dependent diabetes mellitus (IDDM). Am J Hum Genet 52:506–16.

Wolfram S. 1996. The mathematica. Champaign, IL: Wolfram Research, Inc.

Xiong MM, Krushkal J, Boerwinkle E. 1998. TDT statistics for mapping quantitative trait loci. Ann Hum Genet 62:431–52.

APPENDIX

We outline here the analytical power computation for the case when the marker is not at a QTL but is linked to and is in linkage disequilibrium with a QTL in the presence of background polygenes. More detailed derivations are available upon request from the authors. Let the marker locus have two alleles M and m (with frequencies of f and f', respectively), and the QTL under test have two alleles Q and q. The means and variances among children who receive the allele M and m from heterozygous parents are, respectively,

$$\mu_M = E(Y \mid Mm_p, M_o) = \sum_{g_m=1}^{3}\sum_{g_f=1}^{3}\left[\sum_{g_o=1}^{3}(\mu_{g_o}P(g_o, g_f, g_m \mid Mm_p, M_o))\right] + \mu_{pg}, \quad \text{(A1a)}$$

$$\mu_M = E(Y \mid Mm_p, M_o) = \sum_{g_m=1}^{3}\sum_{g_f=1}^{3}\left[\sum_{g_o=1}^{3}(\mu_{g_o}P(g_o, g_f, g_m \mid Mm_p, M_o))\right] + \mu_{pg}, \quad \text{(A1b)}$$

$$\sigma_M^2 = \sum_{g_m=1}^{3}\sum_{g_f=1}^{3}\left[\sum_{g_o=1}^{3}((\mu_{g_o}+\mu_{pg})^2 P(g_o,g_f,g_m \mid Mm_p, M_o))\right] - \mu_M^2 + \sigma_e^2 + \sigma_{pg}^2, \quad \text{(A1c)}$$

$$\sigma_m^2 = \sum_{g_m=1}^{3}\sum_{g_f=1}^{3}\left[\sum_{g_o=1}^{3}((\mu_{g_o}+\mu_{pg})^2 P(g_o,g_f,g_m \mid Mm_p, m_o))\right] - \mu_m^2 + \sigma_e^2 + \sigma_{pg}^2. \quad \text{(A1d)}$$

$$P(g_o,g_f,g_m \mid Mm_p, M_o) = \frac{P(g_o,g_f,g_m,Mm_p,M_o)}{\sum_{g_f}\sum_{g_m}\sum_{g_o} P(g_o,g_f,g_m,Mm_p,M_o)},$$

where $P(g_o, g_f, Mm_p, M_o)$ is the joint probability that father, mother, and an offspring have genotypes of g_f, g_m, and g_o, respectively, and there is at least one parent heterozygous at the marker locus (the event denoted by Mm_p) and a heterozygous parent transmits the marker allele M to the child. For a specific combination of g_f, g_m, and g_o, $P(g_o, g_f, g_m, Mm_p, M_o)$ can be computed which is a function of c (the recombination rate between the marker locus and the QTL), p (the frequency of Q), f, and the coefficient (δ) of linkage disequilibrium and the recombination rate c between the marker and QTL. With this and a and d, we can compute the means and variances in equations A1a–d.

Let $\bar{Y}_M$ and $\bar{Y}_m$ be the mean phenotypic values of the children who receive the alleles M and m, respectively, in the sample. The derivations for $Var(\bar{Y}_M)$, $Var(\bar{Y}_m)$, and $Cov(\bar{Y}_M, \bar{Y}_m)$ are almost the same as the derivations for $Var(\bar{Y}_Q)$, $Var(\bar{Y}_q)$, and $Cov(\bar{Y}_Q, \bar{Y}_q)$ as shown in detail in the text by replacing the allele Q with M, and q with m. Briefly,

$$Var(\bar{Y}_M - \bar{Y}_m) = Var(\bar{Y}_M) + Var(\bar{Y}_m) - 2Cov(\bar{Y}_M, \bar{Y}_m),$$
$$Var(\bar{Y}_M) = Var\left(\sum_{i=1}^{n}\sum_{j=1}^{n_{Mi}} Y_{Mij}/n_M\right) = \frac{\sigma_M^2}{n_M} + \frac{1}{n_M^2}\sum_{i=1}^{n}\sum_{j_1,j_2} Cov(Y_{Mij_1}, Y_{Mij_2}).$$

In the above equation,

$$\sum_{j_1,j_2} Cov(Y_{Mij_1}, Y_{Mij_2}) = \frac{ff'}{1-(f^2+f'^2)^2}[0.5J(J-2)(f^2+f'^2)Cov(Fulsib) + Jff'Var(Y_{MM}) + ff'(2J^2-3J)Cov(Fullsib)].$$

$Var(Y_{MM})$ is the phenotypic variance of the individuals with the marker genotype MM. Let $E(Y_{MM})$ be the expected phenotypic value of individuals with the marker genotype MM, which is,

$$E(Y_{MM}) = \frac{P_{MQ}^2\mu_{QQ} + 2P_{MQ}P_{Mq}\mu_{Qq} + P_{Mq}^2\mu_{qq}}{f^2}.$$

Therefore, $Var(Y_{MM})$ can be obtained as follows,

$$\begin{aligned} Var(Y_{MM}) &= E(Y_{MM}^2) = E^2(Y_{MM}) \\ &= \frac{[P_{MQ}^2(\mu_{QQ}^2 + \sigma_{pg}^2 + \sigma_e^2) + 2P_{MQ}P_{Mq}(\mu_{Qq}^2 + \sigma_{pg}^2 + \sigma_e^2) + P_{Mq}^2(\mu_{qq}^2 + \sigma_{pg}^2 + \sigma_e^2)]}{f^2} \\ &\quad - \left(\frac{P_{MQ}^2\mu_{QQ} + 2P_{MQ}P_{Mq}\mu_{Qq} + P_{Mq}^2\mu_{qq}}{f'^2}\right)^2. \end{aligned}$$

Similarly,

$$\begin{aligned} Var(\bar{Y}_m) = \frac{\sigma_m^2}{n_m} + \frac{n}{n_m^2}\frac{ff'}{[1-(f^2+f'^2)^2]}&[0.5J(J-2)(f^2+f'^2)Cov(Fullsib) \\ &+ Jff'Var(Y_{mm}) + ff'(2J^2-3J)Cov(Fullsib)] \end{aligned}$$

where

$$\begin{aligned} Var(Y_{mm}) &= \frac{[P_{mQ}^2(\mu_{QQ}^2 + \sigma_{pg}^2 + \sigma_e^2) + 2P_{mQ}P_{mq}(\mu_{Qq}^2 + \sigma_{pg}^2 + \sigma_e^2) + P_{mq}^2(\mu_{qq}^2 + \sigma_{pg}^2 + \sigma_e^2)]}{f^2} \\ &\quad - \left(\frac{P_{mQ}^2\mu_{QQ} + 2P_{mQ}P_{mq}\mu_{Qq} + P_{mq}^2\mu_{qq}}{f'^2}\right)^2. \end{aligned}$$

$$Cov(\bar{Y}_M, \bar{Y}_m) = \frac{1}{n_M n_m}\sum_{i=1}^{n}\sum_{j_1,j_2} Cov(Y_{Mij_1}, Y_{mij_2}),$$

where

$$\begin{aligned}
&\sum_{j_1,j_2} Cov(Y_{Mij_1}, Y_{mij_2}) \\
&= \frac{2(2ff')(f^2 + f'^2)}{1-(f^2 + f'^2)^2}\frac{J}{2}\frac{J}{2}Cov(Fullsib) \\
&\quad + \frac{2ff'(2ff')}{1-(f^2 + f'^2)^2}\{[JP(MM_o \mid Mm \times Mm)JP(Mm_o \mid Mm \times Mm) \\
&\quad + JP(MM_o \mid Mm \times Mm)JP(mm_o \mid Mm \times Mm) \\
&\quad + JP(Mm_o \mid Mm \times Mm)JP(mm_o \mid Mm \times Mm) \\
&\quad + 0.5JP(Mm_o \mid Mm \times Mm)(JP(Mm_o \mid Mm \times Mm) - 1)]Cov(Fullsib) \\
&\quad + JP(Mm_o \mid Mm \times Mm)Var(Y_{Mm})\}.
\end{aligned}$$

In the above equation

$$\begin{aligned}
Var(Y_{Mm}) = &[P_{MQ}P_{mQ}(\mu_{QQ}^2 + \sigma_{pg}^2 + \sigma_e^2) + (P_{MQ}P_{mq} + P_{Mq}P_{mQ})(\mu_{Qq}^2 + \sigma_{pg}^2 + \sigma_e^2) \\
&+ P_{Mq}P_{mq}(\mu_{qq}^2 + \sigma_{pg}^2 + \sigma_e^2)] \\
&\div (ff') - E^2(Y_{Mm})
\end{aligned}$$

and

$$E(Y_{Mm}) = \frac{P_{MQ}P_{mQ}\mu_{QQ} + (P_{MQ}P_{mq} + P_{Mq}P_{mQ})\mu_{Qq} + P_{Mq}P_{mq}\mu_{qq}}{ff'}.$$

The number (n) of informative nuclear families in the total of N_s screened ones is

$$n = N_s[1-(f^2 + f'^2)^2] = N_s[2(2ff')(f^2 + f'^2) + 2ff'(2ff')].$$

The total heterozygous parents in the n informative nuclear families is

$$\begin{aligned}
E(n_H) &= n\frac{2(2ff')(f^2 + f'^2)}{2(2ff')(f^2 + f'^2) + 2ff'(2ff')} + 2n\frac{2ff'(2ff')}{2(2ff')(f^2 + f'^2) + 2ff'(2ff')} \\
&= \frac{n}{ff' + f^2 + f'^2} = \frac{n}{1-ff'}
\end{aligned}$$

and $E(n_M) = E(n_m) = JE(n_H)/2$.

With all the above derivations and by the same procedures as detailed in the text, the analytical power of the TDT_G when the marker locus is not a QTL can be computed.